Genomic evidence reveals three sequential W-autosome fusions in *Heliconius* butterflies

Nicol Rueda - M[1,7], Carolina Pardo-Diaz[1], Gabriela Montejo-Kovacevich[2], W. Owen McMillan[3], Krzysztof M. Kozak[3,4], Carlos F. Arias[3,5], Jonathan Ready[5,6], Shane McCarthy[7], Richard Durbin[2,7], Chris D. Jiggins[2], Joana I. Meier[2,7*], 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]Museum of Vertebrate Zoology, UC Berkeley, USA

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

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

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

[7]Tree of Life Programme, Wellcome Sanger Institute, United Kingdom

[+]camilo.salazar@urosario.edu.co (CS)

Abstract

Sex chromosomes are evolutionarily labile in many animals and sometimes fuse with autosomes, creating so-called neo-sex chromosomes. Fusions between sex chromosomes and autosomes have been proposed to reduce sexual conflict and to promote adaptation and reproductive isolation among species. Recently, advances in genomics have fuelled the discovery of such fusions across the tree of life. Here, we discovered multiple fusions leading to neo-sex chromosomes in the *sara/sapho* clade of the classical adaptive radiation of *Heliconius* butterflies. *Heliconius* butterflies generally have 21 chromosomes with very high synteny. However, the five species in the *sapho* subclade show large variation in chromosome number ranging from 21 to 60. We found that the W chromosome is fused with chromosome 4 in all of them. Two sister species pairs showed subsequent fusions between the W and chromosomes 9 or 14, respectively. These sequential fusions between autosomes and sex chromosomes make *Heliconius* butterflies an ideal system for studying the role of neo-sex chromosomes in adaptive radiations and the degeneration of sex chromosomes over time. Our findings emphasize the capability of WGS technology to detect neo-sex chromosomes even when sex chromosomes are not explicitly assembled.

Author Summary

Fusions between sex chromosomes and autosomes are thought to have the potential to resolve sexual conflict and enhance local adaptation or reproductive isolation between species. Here, we discovered such fusions in *Heliconius* butterflies. *Heliconius* butterflies mostly have a very stable karyotype with 21 chromosomes and high synteny across species. However, the five species in the *sapho* subclade have up to 60 chromosomes, suggesting that they have undergone many chromosomal fissions. We document that in addition to the fissions, the *sapho* subclade also showed multiple fusions between the female-specific sex chromosome (W) and autosomes. We found a W-4 fusion shared by all five species and additional W fusions with chromosomes 9 and 14 shared by two species each. Even though in many *sapho* subclade species the autosomes have undergone fissions, the chromosomes fused with the W represent the full-length autosomes, as seen in other *Heliconius* species, suggesting that fusing to the W protected them from fissions. Our study reveals the power of short-read sequencing to discover neo-sex chromosomes and showcases *Heliconius* butterflies as a promising system for studying the causes and consequences of sex chromosome evolution.

Introduction

Sex chromosome-autosome (Sex-A) fusions contribute to the evolution of neo-sex chromosomes (1,2), but it remains unclear what promotes them. Sexually antagonistic selection, direct selection, genetic drift, meiotic drive, and sheltering of deleterious mutations have all been suggested as possible drivers of Sex-A fusions (3–6). Sexually antagonistic selection is thought to favour the fusion of sex chromosomes with autosomes harbouring genes under sexually antagonistic selection (7). There is limited evidence for this hypothesis e.g. in sticklebacks (8), *Drosophila* flies (9), warblers (10) and butterflies (11). Sex-A fusions can also become fixed due to meiotic drive (including holocentric drive in holocentric organisms) (12), such as female meiotic drive elements on W/X-A fusions that preferentially end up in the egg instead of the polar bodies (13). An alternative hypothesis is deleterious mutation sheltering, which Sex-A fusions are favoured because they prevent the expression of recessive deleterious alleles in the heterogametic sex (3). As with other chromosomal rearrangements, Sex-A fusions can reduce recombination and potentially strengthen reproductive isolation (14,15). For instance, in the Japanese threespine stickleback *Gasterosteus aculeatus,* a Sex-A fusion resulted in a neo-X chromosome that linked loci underlying behavioural isolation traits and hybrid sterility (4). Sex-A fusions may also facilitate adaptation, such as the Sex-A fusion in *Cydia pomonella* (Tortricidae), which apparently linked two insecticide-resistance genes and genes involved in detoxifying plant metabolites (16).

Cytogenetic and genomic studies revealed that Sex-A fusions have occurred many times across vertebrates (6,17,18) and invertebrates such as spiders (19,20), *Drosophila* flies (21,22) or true bugs of the genus *Dysdercus* (23). In Lepidoptera (butterflies and moths), examples of Sex-A fusions include *Danaus* (24,25) and *Leptidea* butterflies (26,27), among others (16,28–31). 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 (32). Nonetheless, butterflies and moths have remarkably constrained chromosome evolution (33), with most species having a ZW or Z0 sex determination system and a haploid chromosome number ranging between 28 and 32, except for few groups that have experienced extensive fission and fusion events (31,34).

Here, we focused on *Heliconius* butterflies, which have undergone 10 ancestral fusions and thus display 20 autosomes, along with Z and W sex chromosomes with high collinearity across species (35). Only a few species in the genus differ in this ancestral chromosome number, especially species in the *sara/sapho* clade, with some having up to 60 chromosomes (36). The *sara/sapho* clade comprises 12 species (31) that are different from other *Heliconius* due to their inability to synthesize cyanogens (compounds that deter predators) thus forcing them to rely on sequestered plant toxins (37). A subclade of seven species within the *sara/sapho* clade shows particularly high diversification rates (38) and high number of chromosomes (36).

We assembled a reference genome of *Heliconius sara*, a species that shows the normal karyotype of 21 chromosomes. We found that it was fully collinear with the other two species with chromosome-level assemblies. Next, we generated whole-genome resequencing data from 114 individuals of all seven species of the *sara/sapho* clade with high diversification rates to completely resolve their phylogenomic relationships and study genomic differences between the species. We discovered that the *sapho* subclade, which shows the high number of chromosomes in some species, exhibits fusions between the W chromosome and autosomes (W-A fusions). One W-A fusion is shared by all five species, whereas two additional W-A fusions are shared by two species each. The autosomes fused to the W represent the full-length chromosomes, indicating that they have not undergone the same high rates of fissions as the unfused autosomes. These sequential W-A fusions make the *Heliconius sapho* subclade a prime study system for the evolution of neo-sex chromosomes.

Results

High-quality reference genome assembly for *Heliconius sara*

We assembled the genome of *Heliconius sara* using two laboratory-reared females from a stock population originating from Panama (S1 Table). Using the first individual (BioSample SAMEA8947140), we obtained 24 Gbp PacBio CLR data with a coverage of 50x and an N50 subread length of 16.5 kb 122 Gbp of 10X linked-read Illumina data with a coverage of 294x. From the second individual (BioSample SAMEA8947139), we

obtained 137 Gbp Hi-C Illumina data with 111x coverage. The final assembled genome consists of 348.8 Mbp in 384 scaffolds. The contig N50 was 8.2 Mbp and that of scaffold N50 was 17.8 Mbp. Our genome showed the highest contiguity (S2 Table) and BUSCO statistics (S3 Table) of some currently published *Heliconius* genomes. The BUSCO completeness, using the Lepidoptera gene set, achieved 98.2% single-copy BUSCOs and less duplicated, fragmented, and missing BUSCOs than in the genomes of *H. erato* (39), *H. melpomene* (40) and *H. charithonia* (41) (S3 Table). We assigned the largest twenty-two scaffolds to 20 autosomes and one scaffold to the Z chromosome based on synteny with the *Heliconius melpomene* genome. *H. sara* chromosomes are collinear with this genome, as well as with *H. erato* and *H. charithonia* (S1 Fig). For more information on the genome see https://tolqc.cog.sanger.ac.uk/durbin/jiggins/Heliconius_sara/ and https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_917862395.2/.

The W chromosome did not assemble well, as is commonly seen in lepidopteran genomes (31). We thus used whole genome resequencing data from 114 individuals collected during this study (see below “Whole-genome resequencing dataset section”) to assign scaffolds to the W chromosome based on sequencing depth differences between males and females. Among the 360 scaffolds not assigned to a specific chromosome, 32 exhibited a higher mean depth in females than in males in *Heliconius sara*. This pattern suggests that these 32 scaffolds likely constitute a part of the W chromosome (S2 Fig). Interestingly, the reads from the *sapho* subclade species (*H. antiochus*, *H. sapho*, *H. hewitsoni*, *H. eleuchia*, and *H. congener*) did not align to these 32 scaffolds,

suggesting that W chromosome of these species is either too divergent from the *H. sara* W chromosome (S2 Fig) or not present. The W scaffolds in *H. sara* correspond to a single homolog in *H. charithonia* (S1C Fig).

Whole-genome resequencing dataset

A total of 114 individuals were successfully whole-genome resequenced. Our dataset exhibits high taxonomic completeness covering all 7 species within the *sara/sapho* subclade (*H. sara*, *H. leucadia*, *H. antiochus*, *H. sapho*, *H. hewitsoni*, *H. eleuchia*, and *H. congener*) and 19 out of the 28 described subspecies (42) (S1 Table). The average mapping percentage to the *H. sara* genome was 95.56% (range: 77.38% - 99.14%) (S1 Table). We observed a strong phylogenetic signal in the mapping proportion and, consequently, in the proportion of missing data per individual (S1 Table and S3 Fig). The average coverage was 97.7%, 97.6%, and 96.4% for *H. sara* specimens, its sister species *H. leucadia*, and the *sapho* group, respectively. One *H. congener* and one *H. antiochus* individual exhibited a particularly high proportion of missing data (19.5% and 39.5%, respectively) and low mean coverage (7.4X and 9.1X, respectively) (S1 Table and S3 Fig), and were thus excluded from further analyses.

Phylogenetic analysis reveals two main subclades and uncovered incongruence across the genome

We reconstructed a Maximum Likelihood (ML) phylogenetic tree using 183,282,470 concatenated sites. This phylogeny separated individuals into two main subclades, consistent with the PCA analyses (S4 Fig): (i) *sara* and (ii) *sapho* (Fig 1). The *sara* subclade 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* subclade was split into two well-resolved lineages (*H. antiochus* and a subclade composed of two monophyletic groups: *H. eleuchia/H. congener* and *H. sapho/H. hewitsoni*). *H. antiochus* appeared as a monophyletic group split into an Andean group (*H. a. aranea* and *H. a. araneides*), and an Amazonian group (*H. a. antiochus* and *H. a. salvini*) (Fig 1). *H. antiochus* nested into the *sapho* subclade, whereas *H. hewitsoni* was found to be sister to *H. sapho,* thus resolving the previously undetermined position of these species (38).

To complement the concatenated phylogeny, we also reconstructed a species tree with Astral. The species phylogeny closely mirrored the genome wide phylogeny (S5A Fig). Although *H. hewitsoni* and *H. sapho* were recovered as sister species, their branch lengths (measured in coalescence units) (S5A Fig) were short. This was also the case for the subspecies of *H. eleuchia,* where *H. e. eleusinus* and *H. e. primularis* sometimes group with *H. congener* (S5A Fig). These phylogenetic discordances were also evident in the DensiTree analysis visualising 271 phylogenies together (S5 Fig).

We found a strong phylogenetic incongruence across chromosomes. The whole-genome topology was only recovered on eight chromosomes (S6-S27 Fig), while nine chromosomes showed *H. congener* appearing as sister either to *H. e. eleuchia* or to a clade composed of *H. e. eleusinus* + *H. e. primularis* (S6-S27 Fig). Similarly, the position of *H. hewitsoni* relative to *H. sapho* varied on eight chromosomes (S6–S27 Fig). Interestingly, we observed sex-specific clustering on three chromosomes (S27B Fig). On chromosome 4 (Chr4), all species in the *sapho* subclade showed females and males forming separate clades within each species, whereas the males of *H. congener* and *H. eleuchia* formed a shared clade and their females formed a shared clade (S27B and S9 Fig). Sex-specific clades were also observed on chromosome 9 in *H. sapho* and *H. hewitsoni* (S27B and S14 Fig) and on chromosome 14 in *H. eleuchia* and *H. congener* (S27B and S19 Fig).

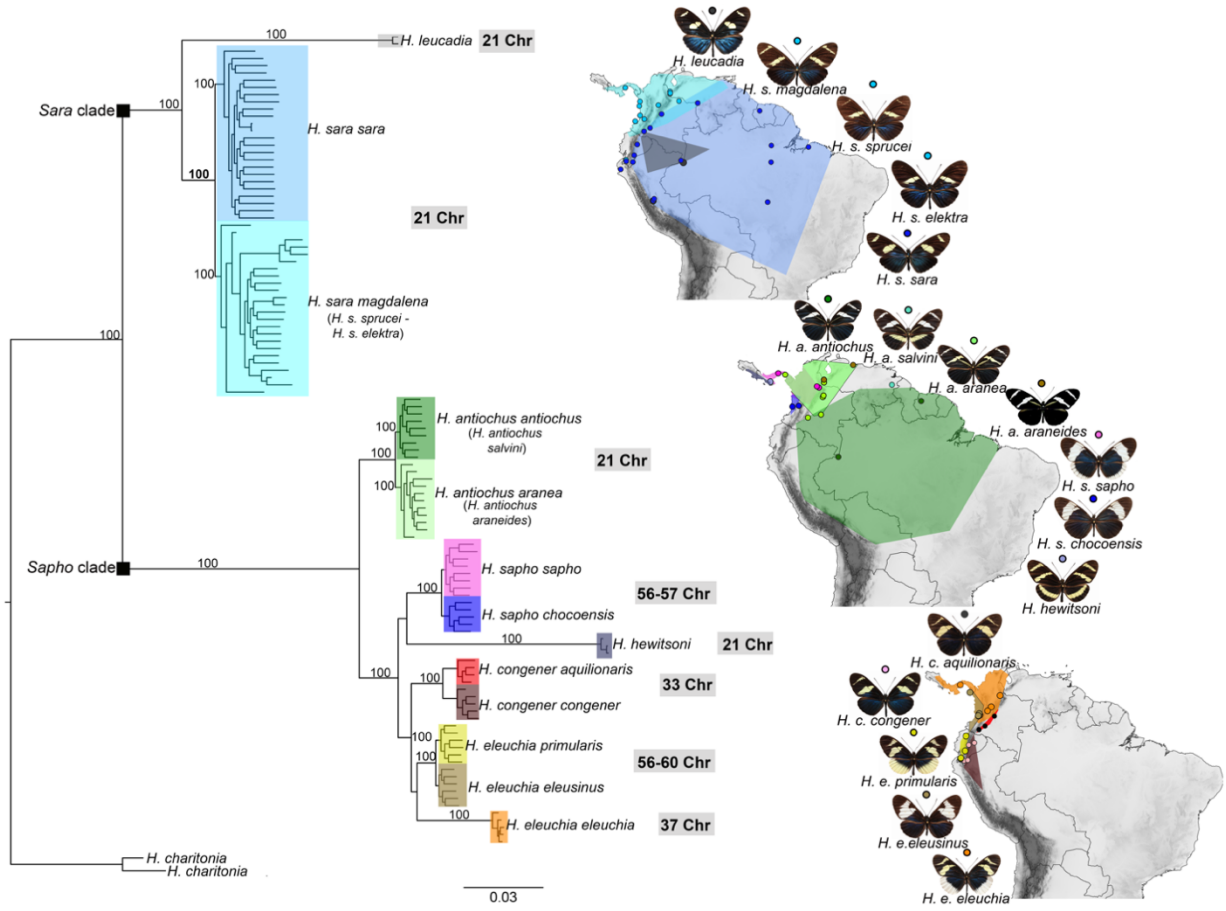


Fig 1. Phylogeny and distribution of the *sara/sapho* clade. Genome-wide ML phylogeny, highlighting the position of the *Sara* and the *Sapho* subclades, with bootstrap support values on the branches. The distribution of all 17 subspecies in the clade from (43) is shown, where dots represent sampling locations in the present study. Each of the 17 subspecies is represented by a single colour, that is the same in the phylogeny and the distribution. The chromosome number of each species is shown in grey rectangles (43). The species' distribution map was created using the free and open source QGIS (44), with a base shapefile obtained from https://www.naturalearthdata.com/downloads/10m-cultural-vectors/ and an elevation raster from https://srtm.csi.cgiar.org/srtmdata/.

Haplotype-based phylogenetic analysis on chromosomes 4, 9 and 14

The grouping by sex we observed in the phylogenetic trees of Chr4, Chr9 and Chr14 suggest possible fusions between these autosomes and either the Z or W chromosome, or possibly both (females of *Heliconius* are ZW and males are ZZ) (42). As females of Lepidoptera lack crossing over and their meiosis is achiasmatic, they do not recombine (45). This means that if the W chromosome is involved in the fusion (Fig 2A), the Sex-A fusion would be restricted to females and the fused chromosome would tend to accumulate mutations and/or structural variants leading to divergence from its unfused homologous. The unfused chr4 would become a neo-Z2 chromosome in all species of the *sapho* subclade, chr9 would become a neo-Z3 in *H. sapho,* and *H. hewitsoni* and chr14 would become a neo-Z3 in *H. congener* and *H. eleuchia*. Alternatively, if the Z chromosome is involved in the fusion (Fig 2B), females would initially still have the unfused homologue (neo-W) that would start to accumulate mutations and/or structural variants, leading to the divergence from the Z-fused homologue. If the sex-autosome fusion was with the Z in the *sapho* subclade, chr4 would become neo-W2 in all five species, chr9 would become neo-W3 in *H. sapho* and *H. hewitsoni* and chr14 would become neo-W3 in *H. congener* and *H. eleuchia*. Lastly, it is possible that both the Z and the W fused to the autosomes, leading to neo-W and neo-Z, reducing the number of chromosomes.

Under either fusion to the Z or the W, we predict i) site-specific genealogies where each female has one haplotype that forms part of the same clade as the male haplotypes, while the other haplotype forms a female-specific clade (hereafter called female-specific haplotype clustering), ii) low genetic differentiation (F_{ST}) on Chr4, Chr9, and Chr14, due to higher divergence between males and females within populations and lower variation between populations if they share the same sex-autosome fusions (see "Patterns of genetic differentiation"), iii) high proportion of sites where all females are heterozygous due to the presence of two different haplotypes (see "Sex-specific differences in heterozygosity and mean depth"), and iv) low sequencing depth in females due to poor mapping of the female-specific haplotypes that have accumulated mutations and structural variants (see "Sex-specific differences in heterozygosity and mean depth") (Fig 2).

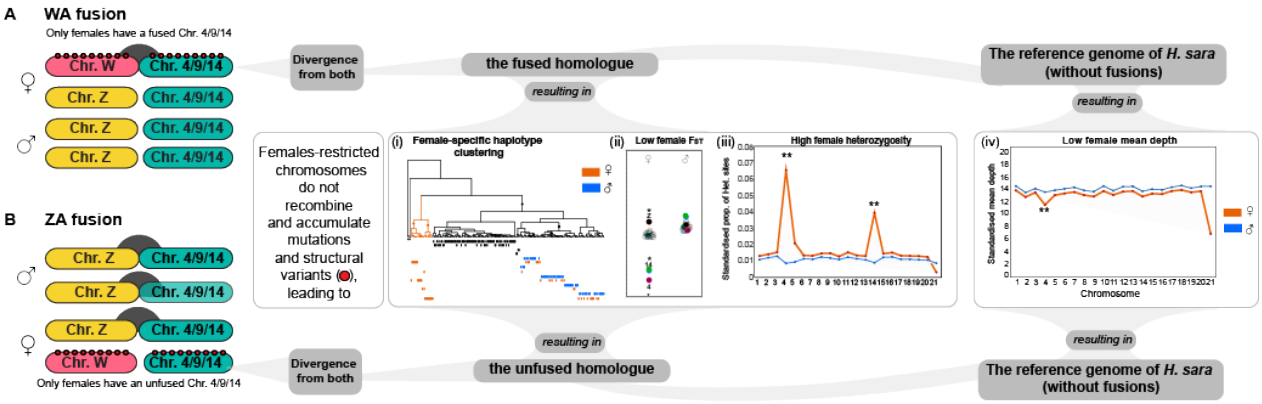


Fig 2. Scenarios of Sex-A fusions involving either the W or Z chromosomes. Scenario of (A) WA fusion and (B) ZA fusion for Chr4, Chr9 and Chr14 differentiating the expected pattern by sex: (i) Female-specific haplotype clustering, (ii) low female Fst, (iii) high female heterozygosity and (iv) low female mean depth. For more details see main text. Yellow: Z chromosome. Pink: W chromosome. Green: autosome. Dark-grey semicircles: fusions. Red dots: mutations and/or structural variants. Lastly, it is possible that both the W and Z chromosomes may have fused with these autosomes, resulting in a neo-W and neo-Z sex chromosome constitution.

Consistent with our hypotheses, we identified 218,839 SNPs on Chr4, where all males within the *sapho* subclade were homozygous, and most females were heterozygous (<1 female per species was homozygous). To study the phylogenetic relationships among male and female haplotypes at these sites, we phased our dataset and inferred marginal phylogenies from ancestral recombination graphs constructed using Relate. We subsampled the dataset to every 1000th SNP with high female heterozygosity (214 SNPs). At 31% of these sites, we recovered the expected phylogeny with a female-specific haplotype clade (Fig 3A). The sites displaying this pattern were distributed across the entire chromosome and were not concentrated in a specific region (Fig 3B). Another set of SNPs (53%) exhibited genealogies where female haplotypes from at least two species within the *sapho* subclade clustered as expected for a Sex-A fusion (S28 Fig). The remaining sites (16%) exhibited a mixed signal, similar to the unphased phylogenetic tree obtained for the entire chromosome (S27B Fig). This is likely due to phasing errors where the fused and unfused homologues of the females are mixed. On Chr9, we found 66,344 sites where *H. sapho* and *H. hewitsoni* exhibited the expected pattern, given a Sex-A fusion. In this case, we found that the 23% of the 66 sampled sites displayed the expected genealogy (S29A Fig). The remaining 77% of the SNPs showed a mixed signal as in Chr4 (Fig S27B). Finally, on Chr14 for *H. eleuchia* and *H. congener*, we identified 57,953 sites with the described pattern. Among the 57 subsampled sites, 44% showed the expected genealogical clustering (S30A Fig), while the remaining 56% displayed the mixed signal described before for Chr4 and Chr9 (Fig S27B).

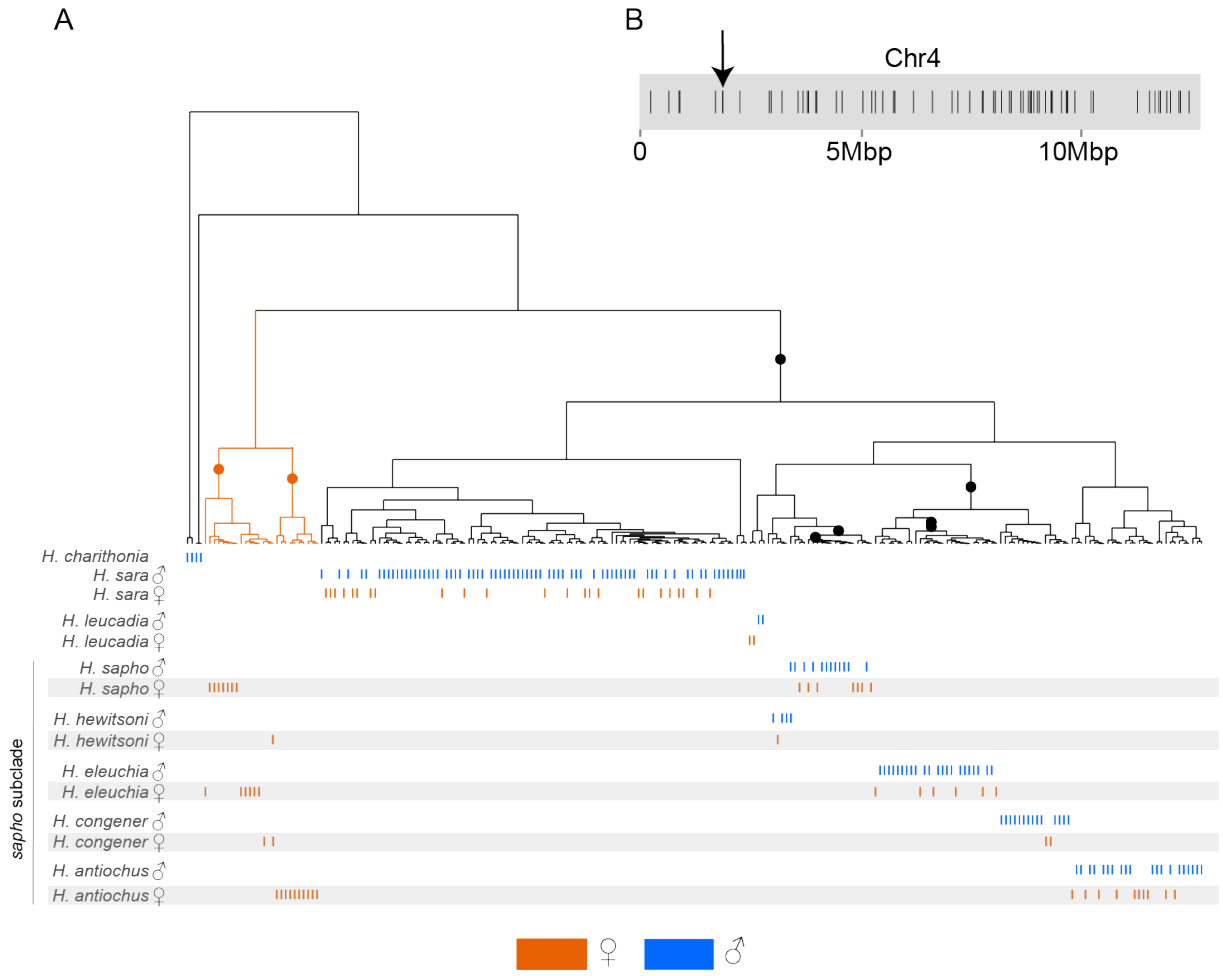


Fig 3. Marginal tree for one of the SNPs on Chr4. Genealogy of the SNP 1,900,983 in Chr4 showing a pattern fully consistent with a Sex-A fusion; a total of 66 subsampled SNPs across this chromosome showed the same pattern. Each vertical line represents an individual haplotype, and the haplotypes of all individuals are shown differentiating those of females (orange) from those of males (blue). Note that one allele of the females in the *sapho* subclade clustered with the alleles of males, while the other female allele formed a separate group (highlighted in orange). (B) Position of each of the 66 SNPs in Chr4 that showed genealogies consistent with a Sex-A fusion; they were not clustered in a specific region but rather distributed along the entire chromosome. The position of the SNP 1,900,983, whose genealogy shown in A, is indicated by an arrow.

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) (S31 Fig). *H. sara* vs. *H. leucadia* were the least differentiated pair (average F_{ST} = 0.05) (S31 Fig). We also observed elevated F_{ST} values on the Z chromosome compared to autosomes in all but one comparison (*H. eleuchia* vs. *H. congener*) (S31 Fig). In line with predictions from sex-autosome fusions in the *sapho* subclade, Chr4 showed lower than average F_{ST} values in the *sapho* subclade, but not in *H. sara* vs. *H. leucadia* (Fig 4A and S31 Fig). This pattern is expected if the females have a haplotype that is more similar to the females of the other population or species compared than to their second haplotype or the males of their own population, thus leading to high within-population diversity and low between-population differences. Chr9 showed lower F_{ST} in *H. sapho* vs. *H. hewitsoni* Chr9 (Fig 4A and S31 Fig), and Chr14 in *H. eleuchia* and *H. congener* (Fig 4A and S31 Fig). The same pattern was observed when we compared F_{ST} between subspecies (S32 – S34 Fig). In line with expectations (Fig. 2), the observed pattern of lower F_{ST} on Chr4, Chr9, and Chr14 was exclusive to females and absent in males across all comparisons mentioned above within the *sapho* subclade (Fig 4B). The observed differences in F_{ST} between males and females were statistically significant for these three chromosomes (Wilcoxon test $p < 0.01$) (S35 Fig). The F_{ST} values of these chromosomes within females were also significantly lower compared to those of the other chromosomes (Wilcoxon test $p < 0.01$) (S35 Fig).

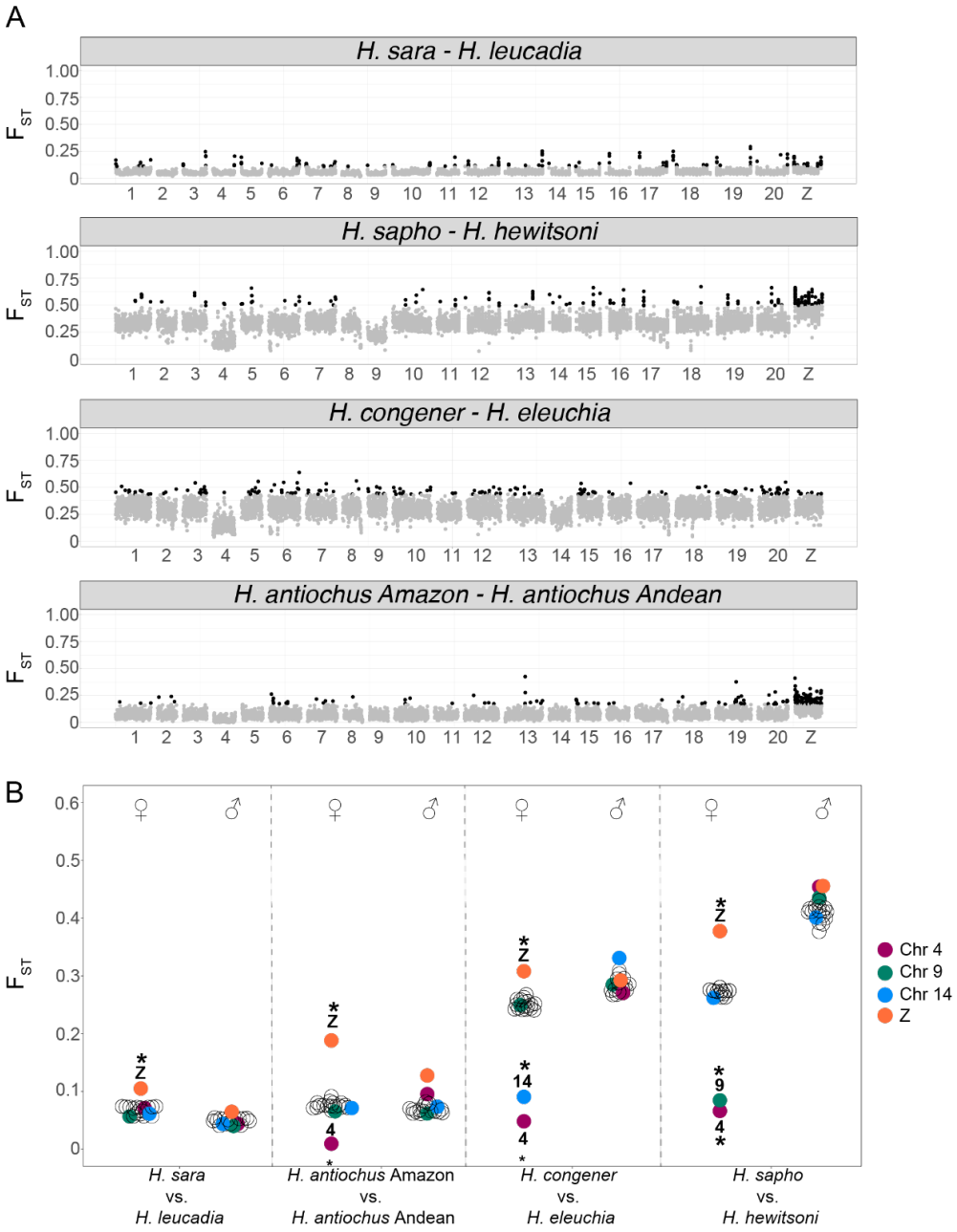


Fig 4. Genome-wide divergence (F_{ST}) in the *sara/sapho* clade. (A) F_{ST} between pairs of species. Each point represents a 50Kb window, whereby the. the top 5% windows are shown in black. The numbers below correspond to *H. sara* chromosomes. Subspecies comparisons are shown in S32-S34 Figs. (B) F_{ST} between pairs of species by sex. Each circle represents a chromosome, and chromosomes with evidence of Sex-A fusions are colour coded (* indicates outlier chromosomes, $p<0.01$). The observed reduction in F_{ST} in females is due to higher genetic diversity within species due to the divergence between the sex chromosome-fused and unfused haplotypes in females.

Sex-specific differences in heterozygosity and mean depth

Consistent with our hypotheses (Fig 2), Chr4, Chr9, and Chr14 showed striking sex-specific differences in the proportion of heterozygous sites and mean depth only in the species of the *sapho* subclade, supporting three sequential fusions of these chromosomes with the Z or W chromosomes or both (Fig 5). The strongest difference in the proportion of heterozygous sites was observed on Chr4, where females of *H. eleuchia*, *H. congener*, *H. sapho*, *H. hewitsoni,* and *H. antiochus* showed a higher proportion of heterozygous sites than males and the other autosomal chromosomes in females (Fig 5B). Females of *H. eleuchia* and *H. congener* also showed a high proportion of heterozygous sites on Chr14 and in *H. hewitsoni* and *H. sapho* on Chr9 (Fig 5B). Differences between males and females were significant on these three chromosomes for all species of the *sapho* subclade (Wilcoxon test, $p < 0.01$) (S36 Fig), except for *H. hewitsoni* where differences could not be tested due to low sample size. The proportion of heterozygous sites of females was also significantly higher for Chr4 than for the other chromosomes in *H. eleuchia*, *H. sapho*, *H. antiochus* and *H. congener* (Wilcoxon test, $p < 0.01$) (S36 Fig). The same was true for Chr14 in *H. eleuchia* and *H. congener,* and Chr9 in *H. sapho* (Wilcoxon test, $p < 0.01$) (S36 Fig). A high proportion of heterozygous sites was not observed in the females of the species *H. sara* and *H. leucadia* on any chromosome (Fig 5 and S36 Fig). Chromosome Z exhibited lower proportion of heterozygous sites than autosomal chromosomes in females across all species (Fig 5 and S36 Fig).

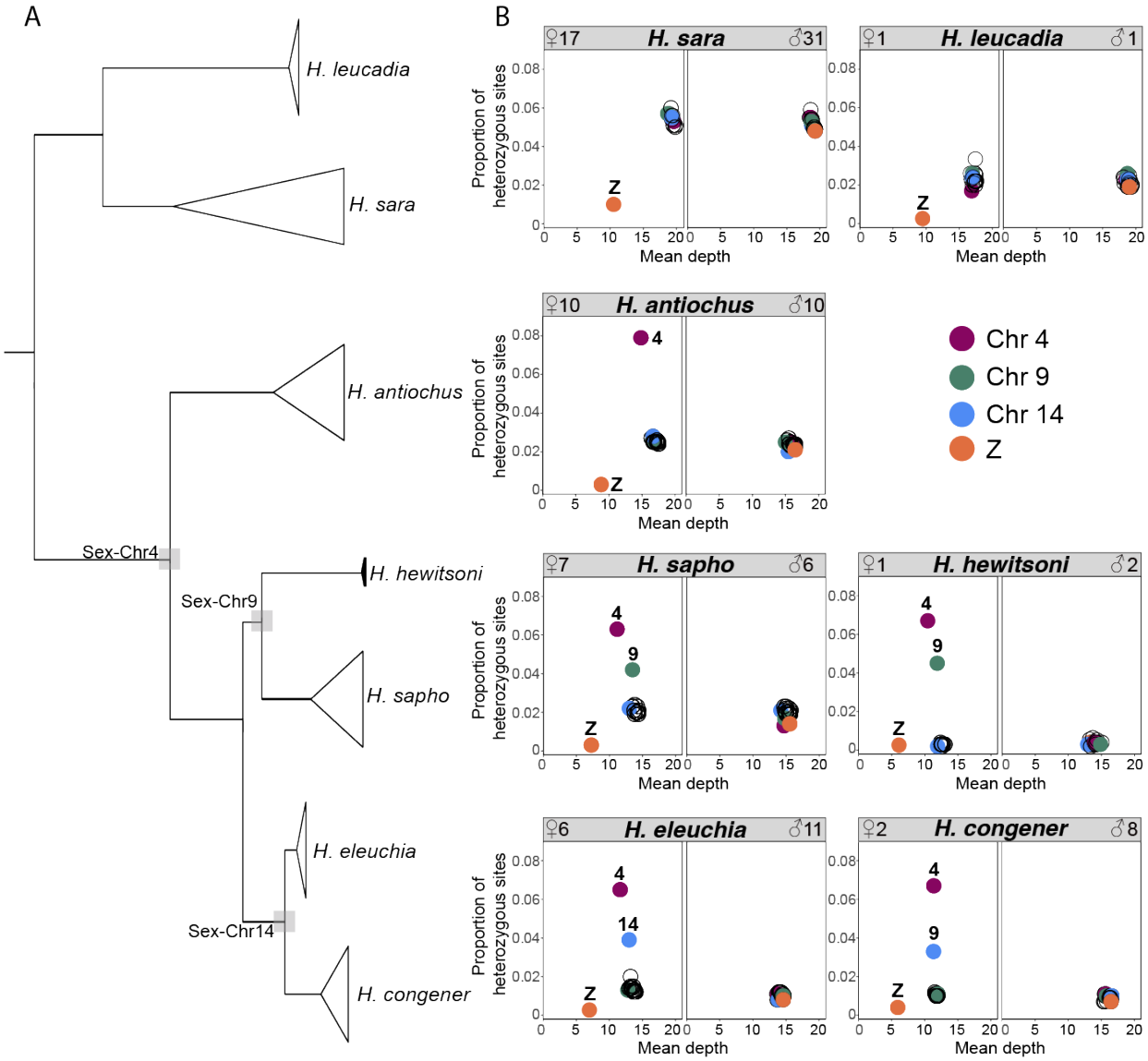
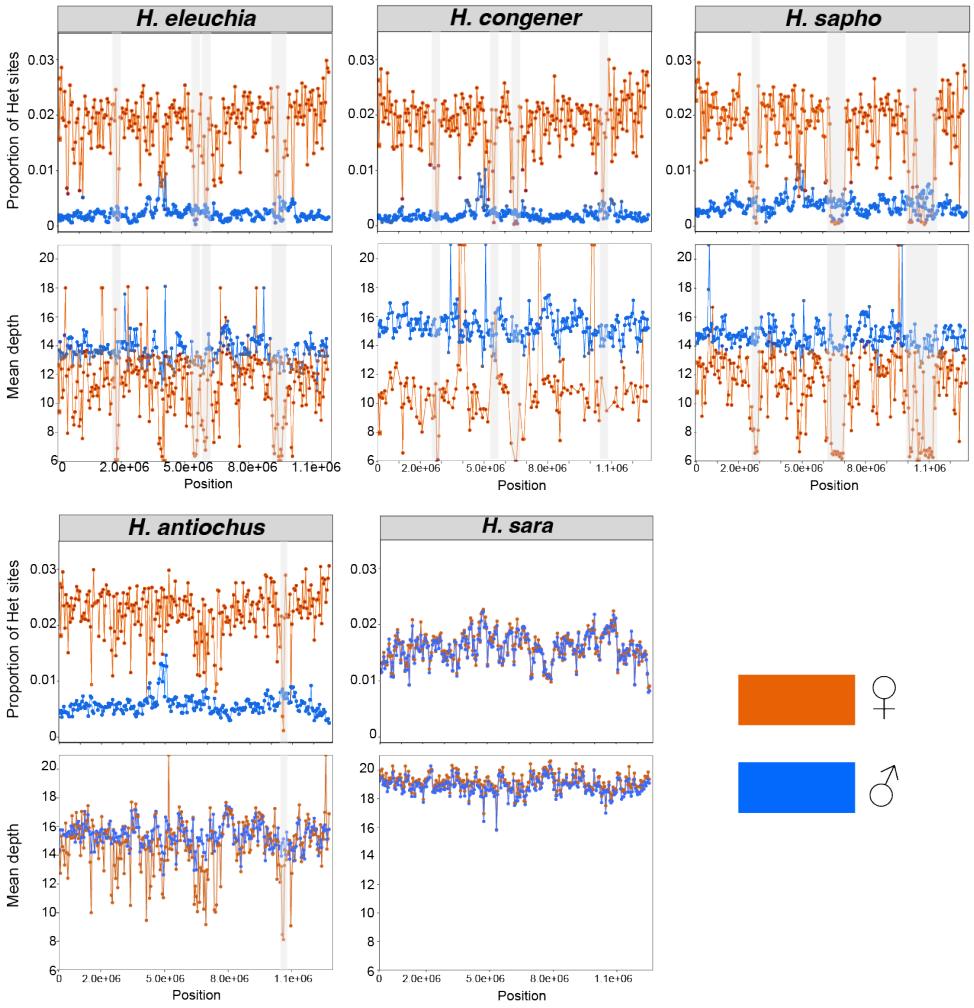


Fig 5. Genome-wide topology and patterns of heterozygosity and sequencing depth across the genome. (A) Genome-wide topology with grey squares highlighting nodes with putative Sex-A fusions. (B) The proportion of heterozygous sites vs. mean sequencing depth per chromosome, by sex and in each species. Each circle represents a chromosome, and chromosomes with evidence of Sex-A fusions are colour coded. The high heterozygosity in females is due to the presence of the fusion in only one of the haplotypes, which becomes divergent from its counterpart. The low mean depth is because the haplotype limited to females diverged enough to be difficult to map onto the reference genome of *H. sara.*

Females of *H. eleuchia*, *H. sapho*, *H. hewitsoni,* and *H. antiochus* also showed a reduced mean depth on chromosome 4, whereas the mean depth on that chromosome in males was normal (Fig 5B). However, these differences were only significant in *H. sapho* (Wilcoxon test, $p \leq 0.01$) (S37D Fig). This pattern was not true for Chr14 in *H. eleuchia* and *H. congener*, nor Chr9 for *H. sapho* and *H. hewitsoni* (Figs 5). The mean depth of Chr4 was also lower than that of all other autosomes in females of *H. sapho* (Wilcoxon test, $p \leq 0.01$) (S37D Fig). However, this was not true for Chr14 and Chr9 (S37 Fig).

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 specific region on the chromosomes; Wilcoxon test, $p < 0.01$) (Fig 6 and S38-S40 Fig). Also, mean depth values were lower for females than males in most windows on Chr4 for *H. eleuchia, H. congener, H. sapho* and *H. antiochus*, Chr9 for *H. sapho* and Chr14 for *H. eleuchia* and *H. congener* (Wilcoxon test, $p < 0.01$) (Fig 6 and S38-S40 Fig). As expected, some peaks of high sequencing depth, likely due to expansions of repeats or duplications were also visible in both females and males (Fig 6). Interestingly, one female of *H. congener* has a region of eight windows on Chr4 with very high sequencing depth (Fig 6), indicating that repeat expansions may still be ongoing. There were also few windows on Chr4, Chr9, and Chr14 where females exhibited both lower mean sequencing depth than males and (almost) no heterozygous sites (highlighted with a grey rectangle) (Fig 6). In these regions, the female-specific haplotype likely diverged too much from the reference genome to map well and thus the heterozygosity is low in females because only one haplotype is represented in the data.

337 *H. sara* and *H. leucadia* were the only species in the clade that did not show sex-specific
338 patterns in heterozygosity and mean depth in Chr4, Chr9 and Chr14 (Fig 6 and S38-S40
339 Fig). Interestingly, the heterozygosity in females of *H. congener* and *H. sapho* dropped
340 at the end of Chr14 and Chr9, respectively, to values similar to those of the males (S38
341 Fig), indicating that the last part of these chromosomes may not be fused to a sex
342 chromosome in this species. However, the mean depth pattern did not change in these
343 regions (S38 Fig).



344

Fig 6. Patterns of heterozygosity and depth across Chr4. Proportion of heterozygous (Het) sites and mean sequencing depth in 50 kb sliding windows for each species. Males are shown in blue and females in orange. Grey rectangles highlight regions 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. The mean depth figures were trimmed to a maximum value of 20 in order to better visualize the differences between sexes.

Evidence for three W-autosome fusions

In order to elucidate if the W or Z chromosome or both are involved in the sex-autosome fusions, we produced Illumina Hi-C data for a female *H. congener* (0.71 Gbp) and a male *H. sapho* (0.73 Gbp). We mapped the Hi-C data of these two *sapho* subclade individuals and the *H. sara* Hi-C data produced for the reference genome to our *H. sara* genome and to a previously published *H. charithonia* genome which has the W assembled (41). In line with phylogenetic distances, the proportion of reads mapping to the *H. sara* reference genome were much higher (96%, 93% and 92%) than to the *H. charithonia* reference genome (69%, 70%, 71% for *H. sara*, *H. congener* and *H. sapho*, respectively. The mean mapping quality for all three species against both genomes exceeded a Phred quality score of 35.

We did not observe an excess of Hi-C contacts either between autosomes or between autosomes and sex chromosomes in the *H. sara* female (Fig 7A), as expected if the sex-autosome fusions were only in the *sapho* subclade. The Hi-C signal of the *H. sapho* male and the *H.* congener female showed that their genomes are split into 56 and 33 chromosomes, respectively (Fig 7B-C), consistent with findings by Brown et al (36)

and suggesting a high number of chromosome fissions. However, no fusion was observed in the *H. sapho* male, suggesting that the Chr4 and Chr9 fusions are likely not with the Z, but with the W (Fig 7B). In contrast, the female *H. congener* showed an excess of contacts between Chr4 and Chr14, in line with a fusion (Fig 7C). There was no excess of Hi-C contacts between Chr4 and Chr14 with the Z chromosome (Fig 7C and 7D), indicating that the Chr4 and Chr14 are likely fused with the W instead. Even though there is signal of Hi-C contact between these chromosomes and the W, the low mapping rates of Illumina reads of the *sapho* subclade to the W of the *Heliconius sara* genome (S2 Fig) likely explains the absence of a stronger signal.

As the Hi-C signal of chromosome fusions in the *H. congener* female represents a mix of signals from the fused and unfused haplotypes (Fig 7C-D), we phased the Hi-C data across the chromosomes of interest (Chr4, Chr14, W, Z) and used *chomper* (46) to split the Hi-C data into two subsets of read pairs representing the two haplotypes (47.5% and 52.2% of the read pairs in each subset). The excess of Hi-C contacts between Chr4 and Chr14 was completely absent on one haplotype (Fig 7E) and very strong on the other haplotype (Fig 7F). The Hi-C contacts between Chr4 and Chr14 are unevenly distributed (Fig 7E), suggesting not only a fusion between these chromosomes but also the presence of other chromosomal rearrangements such as inversions or translocations.

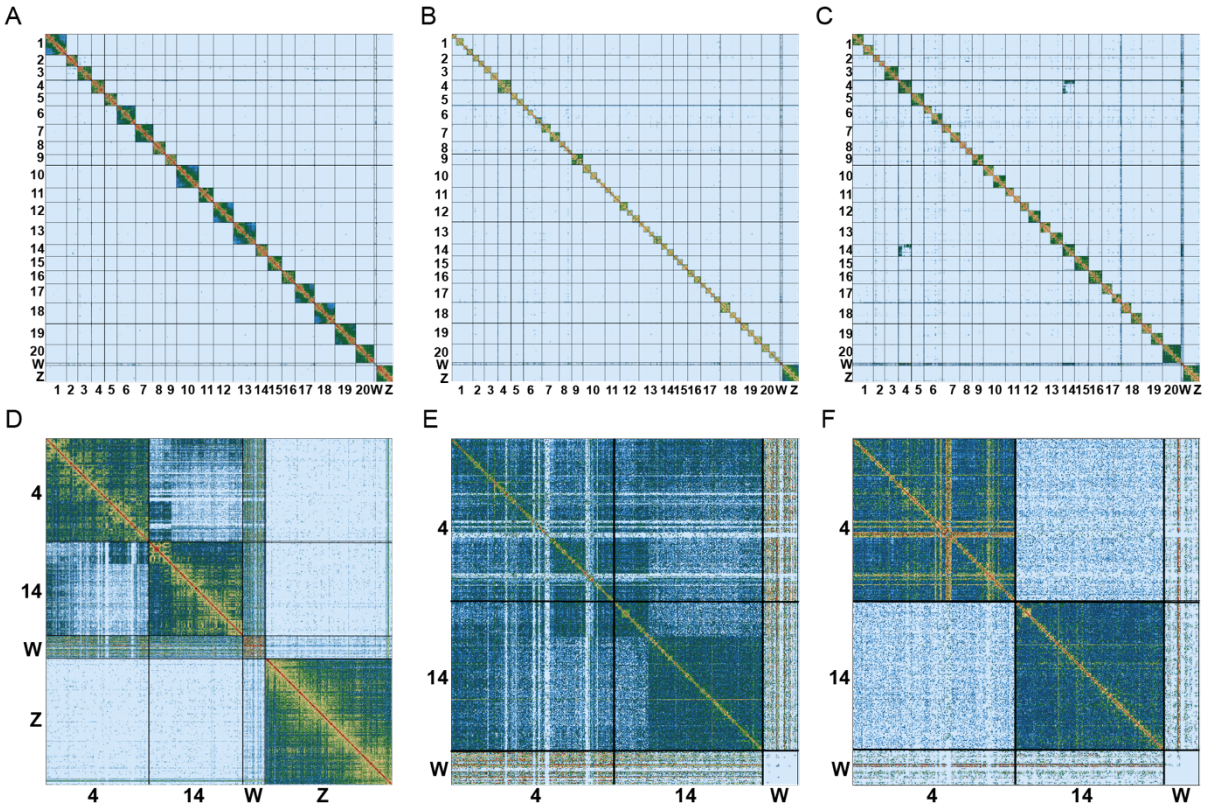


Fig. 7. Hi-C contact density map confirmed a high number of chromosome fissions and W-A fusions in the *sapho* subclade. Hi-C contact heatmaps of the *H. sara* reference genome of (A) the *H. sara* female, (B) the *H. sapho* male, and (C) the *H. congener* female. Note that the chromosomes of the *H. sara* reference genome are labelled and their boundaries are denoted with vertical and horizontal black lines. (D) Zoom in (C) showing Hi-C contacts on Chr4, Chr14, ChrW, and ChrZ. (E) Hi-C reads assigned to one of the two haplotypes in the *H. congener* female showing the W-A fusion. (F) Hi-C reads assigned to the other haplotype in the *H. congener* female showing no evidence of a fusion.

Discussion

We found evidence of three W-A fusions involving Chr4, Chr9, and Chr14 in the evolution of the *sara/sapho* clade in *Heliconius* butterflies. These autosomes seem to have fused with the W chromosome as supported by: (i) females with one haplotype each forming part of female-specific haplotype clade, (ii) low F_{ST} values in females, (iii)

high heterozygosity and low mean depth in females, and (iv) excess of Hi-C contacts in the *H. congener* female but not in the *H. sapho* male. Because females are the heterogametic sex in butterflies and show no recombination (45), the W-A fusion is restricted to females and generates female-specific haplotypes that do not recombine with the unfused chromosomes. This results in genealogies where one haplotype of the females clusters with male haplotypes while the other haplotype of the females, the fused one, forms a separate female-specific clade. Each female shows many heterozygous sites on the ancestral Chr4 and depending on the species also on Chr9 or Chr14, as it has inherited from the mother a fused W-4(-9/14) that never recombines and accumulates mutations and structural variants and from the father unfused Z chromosomes (Z1=ancestral Z, Z2=ancestral Chr4 (and Z3=ancestral Chr9 or Chr14). This, in turn, increases variation within species and reduces variation between them, as females share the fused haplotypes with females of other species, thus leading to reduced $F_{ST.}$ Chr4 showed female-specific haplotype clustering and high female heterozygosity throughout the chromosome in all five species of the *sapho* subclade, indicating that the W chromosome fused with Chr4 in the ancestor of this group. In line with an old fusion and increasing degeneration, females showed a high proportion of heterozygous sites in most of the chromosome, and in some regions, the fused haplotype had degenerated so much that mapping to the *H. sara* genome failed, leading to low sequencing depth and the absence of heterozygous sites. Similar, but less strong patterns of female-biased heterozygosity are found in Chr9 in *H. sapho* and *H. hewitsoni*, and in Chr14 in *H. eleuchia* and *H. congener.* In line with younger fusions and

420 weak degeneration, the sequencing depth of females matches that of the males on
421 those chromosomes, indicating that the female-specific haplotypes still map well to the
422 *H. sara* genome.

423 Sex-A fusions in Lepidoptera usually involve the Z chromosome (31,47), but this
424 does not seem to be the case in *Heliconius*. Our *H. sapho* male (ZZ) did not have any
425 chromosome fusions, indicating that neither Chr4 nor Chr9 in that species is fused to the
426 Z, whereas Hi-C data from the female *H. congener* supports a fusion on one of the
427 haplotypes between Chr4 andChr14 with the W chromosome. Thus, in this species, the
428 unfused Chr4 and Chr14 become neo-Z2 and neo-Z3 chromosomes, respectively, and
429 the fused chromosome becomes the neo-W chromosome. Although the absence of a
430 fusion with the Z in the female *H. congener* clearly points to a W fusion, Hi-C contacts
431 between Chr4-Chr14 and the W are weak. This is likely due to the divergence of the W
432 in the *sapho* subclade from that of *H. sara*, making it difficult to map reads against this
433 reference (S2 Fig). Alternatively, the original W in the *sapho* subclade may have an
434 independent origin to that in *H. sara* or it may have been lost completely, as suggested
435 for other Lepidoptera (47).

436 The finding of multiple W-A fusions in the *sapho* subclade is particularly striking,
437 since this group is known for its high number of chromosomes compared to all other
438 *Heliconius* species that was also confirmed with our Hi-C data. While most species in
439 this genus have 21 chromosomes, *H. sapho* and *H. eleuchia eleusinus and H. e.*
440 *primularis* has 56-57 chromosomes, and *H. eleuchia eleuchia* and *H. congener* have 37

and 33 chromosomes, respectively (36), indicative of high rates of chromosomal fission events in the group. However, even though in these species most autosomes are broken up (Fig 7B-C), the chromosomes fused with the W are the full-length ancestral chromosomes and thus likely did not undergo fissions. Interestingly, *H. antiochus* and *H. hewitsoni* which also show W-A fusions have 21 chromosomes. These findings could be explained by two alternative scenarios: a) The chromosomal fissions in *H. eleuchia*, *H. congener* and *H. sapho* happened independently in each (sub)species after the W-A fusions and fusing to the W protected Chr4 and Chr9/14 from fissions or b) the fissions could be ancestral, and the W-fusions could have involved the largest chromosomes that had not been broken up. In this second scenario, *H. hewitsoni* would have undergone additional autosomal fusion events restoring chromosome number to 21 after fission events. While we think the first scenario is more likely than the second one, full genome assemblies will be required to distinguish these hypotheses.

While we cannot test for an adaptive role of the W-A fusions in the *sapho* subclade, the fact that they occurred three times and remained fixed in multiple species suggests they are at least not deleterious, or that any deleterious effect is masked. In the latter scenario, low recombination around the W-A 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. This is the case in inversions in *H. numata* (3,48). Other alternative scenarios may have favoured the fixation of these fusions. First, a period of strong genetic drift could have led to the fixation of these fusions even if they carry some

463 deleterious mutations. Second, positive natural selection acting on fusions is possible
464 (49), though it might be hard to imagine that such beneficial effects would be found on
465 all three chromosomes that fused to the W. Third, transmission bias such as meiotic
466 drive (50,51), or coincidental linkage with endosymbionts transmitted via females (e.g.
467 male-killing spiroplasma) (52) could explain how the W-A fusions might have fixed.
468 However, if chromosomal fusions occur through non-homologous recombination, meiotic
469 drive might counteract or facilitate the spread of the fusions as observed for *Leptidea*
470 butterflies (53). Fourth, the W-A fusions might have spread due to reduction of sexual
471 conflict if there are sexually antagonistic loci on Chr4/9/14. In Danaini butterflies, sex-
472 biased gene expression is consistent with this hypothesis (11). Finally, the W-A fusions
473 may have contributed to the particularly high diversification rate in this clade if they
474 linked together barrier loci in regions with reduced recombination (14).

475 This is the first genomic study focused on the *sara/sapho* clade. The inclusion of
476 multiple species and subspecies of this clade from a broad geographic range also
477 allowed us to redefine some of the relations previously reported (38), and to identify the
478 effect of geography in shaping diversity. The phylogenetic position we found for *H.*
479 *antiochus* and *H. hewitsoni* contrasts with previous amplicon based phylogenies (38,54)
480 but agrees with a recent whole genome phylogeny based on *de novo* genome
481 assemblies (55), suggesting that the phylogenetic relations we describe for these two
482 species are the most plausible. We also identified *cis* and *trans*-Andean lineages for *H.*
483 *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.

Further studies are needed to understand the evolutionary drivers of the W-A fusions identified here, as well as their role (if any) in speciation or adaptation in this clade. Our study highlights the importance of including both sexes in short-read population WGS studies for identifying Sex-A fusions. Finally, we show what patterns to expect if the fusions are recent enough that the previously autosomal chromosomes are still diploid in both sexes. As W chromosomes are often not assembled in reference genomes due to their high repeat content, genome assemblies might not necessarily reveal W-A fusions. Our study thus demonstrates the power of short-read population data to discover sex-A fusions, particularly for taxa where one sex is achiasmic.

Materials and Methods

Genome assembly of *Heliconius sara*

We used two laboratory-reared females from a stock population from Panama to generate a reference genome for *H. sara* (BioSamples SAMEA8947140 and SAMEA8947139; S1 Table). We assembled the genome by combining PacBio, 10X data and Hi-C data, all generated by the Tree of Life Programme at the Wellcome Sanger Institute (). The BioSample SAMEA8947140 was used to generate the PacBio continuous long reads (CLR). Libraries were sequenced on four Single Molecule Real-

Time (SMRT) cells using the PacBio Sequel II system. The linked-reads from 10X Genomics Chromium technology were generated with the same sample and sequenced in four lanes on the Illumina HiSeq X Ten platform. The second BioSample SAMEA8947139 was used to produce Dovetail Hi-C data and sequenced on a HiSeq X Ten platform.

An initial contig assembly was generated from the PacBio (CLR) data using wtdbg2 v2.2 (56). The PacBio data was then used to polish the contigs using Arrow (https://github.com/PacificBiosciences/GenomicConsensus). We then retained haplotig identification with the Purge Haplotigs pipeline (57). The 10X data were mapped to this assembly using Longranger v2.2 (10X Genomics) and variant calling was performed using freebayes v1.1.0-3-g961e5f3 (58). Next, this first assembly was polished using BCFtools consensus v1.9 (59) 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 (60). Lastly, Dovetail Genomics Hi-C data was used to scaffold the assembly further using SALSA v2.2 (61), followed by another round of manual curation with gEVAL (60). 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 (62) using the Lepidoptera gene set. To obtain synteny plots between *H. sara* vs. *H. melpomene*, *H. erato,* and *H. charithonia*, we first performed pairwise alignments between these

525 genomes using minimap2 v. 2.24 (63). Subsequently, we plotted the minimap2 results
526 using custom scripts from (https://github.com/simonhmartin/asynt).

527 We used whole genome resequencing data from 114 individuals obtained in this
528 study (see sample collection for genome resequencing section) to identify the W
529 chromosome within the genome of *Heliconius sara*. We calculated the mean depth
530 across the scaffolds that were not yet assigned to a chromosome. We first generated a
531 file containing the mean depth per site averaged across all individuals of the same sex
532 and species using the --site-mean-depth option of vcftools v. 0.1.14 (64). Then, we used
533 the R package windowscanr v. 0.1 (https://github.com/tavareshugo/WindowScanR) to
534 calculate the average of the mean depth per species, per sex, and per 500 bp windows.
535 Scaffolds where we observed a higher mean depth in females compared to males were
536 assigned to the W chromosome.

537 ## Sample collection for genome resequencing

538 We collected 114 *Heliconius* individuals from 7 species and 18 subspecies in the
539 *sara/sapho* clade across their distribution range: 48 *H. sara*, 2 *H. leucadia*, 21 *H.*
540 *antiochus*, 13 *H. sapho*, 3 *H. hewitsoni*, 17 *H. eleuchia* and 10 *H. congener* (S1 Table).
541 The body of each individual was preserved in NaCl-saturated DMSO solution and stored
542 at -80°C; wings were kept for phenotype reference.

543

544

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. Our *H. sara* genome (HelSar1) was used as a reference to map the reads of each individual using BWA mem v0.7.12 (65) with default parameters. We then used samtools v1.12 to sort and index the alignment files (66). PCR-duplicate reads were identified and removed using Picard tools v2.9.2 (67), and variant calling was conducted with Haplotype Caller (GATK, v3.7.0) in BP-resolution mode (68). Then, samples were jointly genotyped using GATK's GenotypeGVCFs (68). We used vcftools v0.1.14 (64) 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 sequencing depth below 5 and mean depth per individual more than 1.5 times the mean to exclude paralogous regions. For this, we used the custom script *removeTooHighDepthSites.sh* from (https://github.com/joanam/VictoriaRegionSuperflock/BashPipelines). We additionally removed sites with excess heterozygosity across all individuals using the vcftools option

566 --*hardy* and a p-value cut-off of <1e-5 to remove reads from paralogous regions that are
567 collapsed in the reference genome.

568 Analysis of population structure within the *sara/sapho* clade

569 We performed a principal component analysis (PCA) to study the genetic
570 structure of populations. We filtered out monomorphic or multiallelic sites, and sites with
571 minor allele frequency (MAF) smaller than 0.1 with vcftools (64). To reduce the linkage
572 disequilibrium effect, we used the python script *ldPruning.sh* from
573 (https://github.com/joanam/scripts), which removes sites with r^2>0.2 in windows of 50
574 Kbp sliding by 10 Kbp. This resulted in a vcf file with 3,685,916 high-quality SNPs sites.
575 We conducted the PCA using Plink v2.0 with default parameters (69,70).

576 Phylogenetic relationships among *sara/sapho* clade species

577 We generated a whole-genome Maximum Likelihood (ML) tree using a vcf
578 containing all sites as input in RAXML v8.2.9 (71), with the GTRGAMMA model and 100
579 bootstrap replicates. We applied the same procedure to obtain ML trees for each
580 chromosome to study the phylogenetic incongruence across the genome. We also
581 inferred the species tree using the coalescent-based method ASTRAL (72). For this, we
582 used vfctools v0.1.14 (64) to: (i) extract two males per subspecies so Sex-A fusions
583 present only in females do not alter the species tree, and (ii) extract 2 kbp loci spaced at
584 least 10 kbp apart to ensure no linkage disequilibrium between them (73). Then,

samtools (66) was used to generate 271 multilocus blocks, each resulting from concatenating 100 loci. Each block was converted into PHYLIP format using our custom script *vcf2phylip.py* from (https://github.com/joanam/scripts/blob/master/vcf2phylip.py) and used to estimate a ML tree in IQ-tree (74,75) selecting the best model with ModelFinder and assessing node support with 1000 ultrafast bootstraps. The resulting 271 topologies were used as input in ASTRAL. We also investigated these 271 topologies with DensiTree (76) to visualize discordance.

Haplotype-based phylogenetic analysis on chromosomes 4, 9 and 14

To further investigate sex clustering in the phylogenies of autosomes 4, 9 and 14, we phased them to infer their haplotypes. Haplotype phasing was done by combining two methods: WhatsHap, which is a haplotype assembly technique (77), and SHAPEIT4, a statistical phasing method (78). To implement WhatsHap, we used the BAM file of each individual as well as the reference genome to group nearby genetic variants into fully resolved haplotype blocks or phase sets (77). Then, we used the WhatsHap output file to run SHAPEIT4, which further phases haplotypes based on population-level information using default parameters (78).

Next, we inferred SNP coalescent genealogies across each chromosome using Relate v.1.1.2 (79). Initially, the phased vcf file was transformed into haplotype format using RelateFileFormats with the “--mode ConvertFromVcf" flag. We performed this analysis using an effective population size of $1x10^7$ individuals, as estimated for *H. erato*

(80), and a mutation rate of 2.9×10^{-9} per site per generation from *H. melpomene* (81). As Relate further requires a genetic map, we created one using the average recombination rate of 6 cM/Mb calculated in *H. erato* (82), and our custom script (https://github.com/joanam/scripts/blob/master/createuniformrecmap.r). Finally, the ancestral allele was assigned as the one more common in the outgroup *H. charithonia*.

Because any possible Sex-A fusion would produce heterozygous females and homozygous males, we identified sites with this pattern (see hypotheses in Fig 2). We first calculated the number of heterozygous and homozygous individuals per species and site using the *--hardy* option in vcftools version 0.1.14 (46). Then, we used custom scripts to find sites where each species in the *sapho* subclade met the following criteria: (i) no males were heterozygous, (ii) at least one female was heterozygous, and (iii) not more than one female was homozygous (allowing for one female with one allele not called). Finally, we selected one SNP for every 1000 SNPs from these filtered sites and visualized their genealogies using the script *TreeView.sh* (58). This subsampling approach enabled us to examine genealogies from various sites evenly distributed along the chromosomes.

Patterns of genetic differentiation

We calculated F_{ST} by pairs of sister species 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). We also calculated F_{ST} per sex following the same methodology.

Patterns of heterozygosity and mean depth by chromosome

To study chromosomes with F_{ST} patterns different from the genome average, we used the options --het and --depth of vcftools v. 0.1.14 (62) tocalculate 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 proportion of heterozygous sites. 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 sequencing 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 (64). 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 the sexes. To assess differences between chromosomes, we applied a Kruskal-Wallis test and a *post hoc* test (pairwise Wilcoxon test for Kruskal-Wallis).

Identification of fusions with Hi-C data

To investigate whether autosomes 4, 9, and 14 are fused with the W or Z chromosomes, we constructed Hi-C libraries from the thorax of one female of *Heliconius congener* (BioSamples SAMEA112329098; S1 Table) and a male of *Heliconius sapho* (BioSamples SAMEA112696452, S1 Table) using the Arima2 kit (Arima Genomics, Inc). These libraries were then sequenced on an Illumina NovaSeq S4 platform with 150 bp paired-end reads. We used BWA mem2 v 2.2.1 (83) to map the reads against the genomes of *Heliconius sara* (this study) and *Heliconius charithonia* (41) using default parameters. We also mapped the Hi-C data of *H. sara* to both references to confirm the absence of the fusion in the *sara* subclade. Next, we removed PCR duplicates, eliminated poorly aligned reads, and filtered out reads with a mapping quality <10 from the resulting BAM files using samtools v1.12 (66). Additionally, we generated contact maps with pretextview and pretextSnapshot (https://github.com/wtsi-hpag/PretextView and https://github.com/wtsi-hpag/PretextSnapshot). All these steps were done using a custom Perl pipeline developed by Shane McCarthy at the Wellcome Sanger Institute.

To further investigate the excess of Hi-C contacts between autosomes 4, 14 and W in the female of *H. congener*, we generated haplotype-specific Hi-C maps for these

chromosomes. First, we created a version of the *H. sara* reference genome in which chromosomes 4, 14, and W were concatenated together to allow for phasing across chromosomes. Next, we mapped the Hi-C reads of *H. congener* to this modified reference following the mapping pipeline by Arima Genomics, Inc. (). Then we called heterozygous variants using freebayes v1.3.2-dirty (58). These variants were then normalized with bcftools v1.8 (59), decomposed with vcfallelicprimitives (84) and filtered for coverage (>21 and <141 reads) with vcftools v. 0.1.14 (64). Next, the remaining SNPs were phased using HAPCUT2 v1.3.3, using both the bam and the vcf files as input (85). We used the Python script *chomper.py* from () to separate haplotype aligned HiC reads. Finally, these haplotype-specific sets of Hi-C reads were realigned to the original *H. sara* assembly using a custom Perl pipeline developed by Shane McCarthy at the Wellcome Sanger Institute. Hi-C contact map was generated using pretextview and pretextSnapshot (and https://github.com/wtsi-hpag/PretextSnapshot).

Acknowledgments

We 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. We also thank the HPC Service of Universidad del Rosario (CALDAS) for computing time. We thank the core lab of the Tree of Life Programme and the sequencing centre of the Sanger Institute for support with the HiC sequencing.

References

1. Ma WJ, Rovatsos M. Sex chromosome evolution: The remarkable diversity in the evolutionary rates and mechanisms. J Evol Biol. 2022;35(12):1581–8.

2. Abbott JK, Nordén AK, Hansson B. Sex chromosome evolution: historical insights and future perspectives. Proc R Soc B Biol Sci. 2017 May 3;284(1854):20162806.

3. 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.

4. 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.

5. 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.

6. 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.

7. Charlesworth B. Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci U S A. 1978 Nov;75(11):5618–22.

8. 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.

9. Zhou Q, Bachtrog D. Sex-specific adaptation drives early sex chromosome evolution in Drosophila. Science. 2012 Jul 20;337(6092):341–5.

10. 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.

11. Mora P, Hospodářská M, Voleníková AC, Koutecký P, Štundlová J, Dalíková M, et al. Sex-biased gene content is associated with sex chromosome turnover in Danaini butterflies. Mol Ecol. n/a(n/a):e17256.

12. Krátká M, Šmerda J, Lojdová K, Bureš P, Zedek F. Holocentric Chromosomes Probably Do Not Prevent Centromere Drive in Cyperaceae. Front Plant Sci [Internet].

718 2021 [cited 2023 Mar 21];12. Available from:
719 https://www.frontiersin.org/articles/10.3389/fpls.2021.642661

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

722 14. Guerrero RF, Kirkpatrick M. Local adaptation and the evolution of chromosome
723 fusions. Evolution. 2014 Oct;68(10):2747–56.

724 15. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive
725 loci. Proc Natl Acad Sci U S A. 2013 May 7;110(19):E1743-1751.

726 16. Nguyen P, Sýkorová M, Šíchová J, Kůta V, Dalíková M, Čapková Frydrychová R, et
727 al. Neo-sex chromosomes and adaptive potential in tortricid pests. Proc Natl Acad Sci
728 U S A. 2013 Apr 23;110(17):6931–6.

729 17. Matsumoto T, Kitano J. The intricate relationship between sexually antagonistic
730 selection and the evolution of sex chromosome fusions. J Theor Biol. 2016 Sep
731 7;404:97–108.

732 18. Schmid M, Feichtinger W, Steinlein C, Visbal García R, Fernández Badillo A.
733 Chromosome banding in Amphibia. XXVIII. Homomorphic XY sex chromosomes and
734 a derived Y-autosome translocation in Eleutherodactylus riveroi (Anura,
735 Leptodactylidae). Cytogenet Genome Res. 2003;101(1):62–73.

736 19. Král J, Kořínková T, Krkavcová L, Musilová J, Forman M, Herrera IMÁ, et al. Evolution
737 of karyotype, sex chromosomes, and meiosis in mygalomorph spiders (Araneae:
738 Mygalomorphae). Biol J Linn Soc. 2013 Jun 1;109(2):377–408.

739 20. Maddison WP, Leduc-Robert G. Multiple Origins of Sex Chromosome Fusions
740 Correlated with Chiasma Localization in Habronattus Jumping Spiders (araneae:
741 Salticidae). Evolution. 2013;67(8):2258–72.

742 21. Flores SV, Evans AL, McAllister BF. Independent Origins of New Sex-Linked
743 Chromosomes in the melanica and robusta Species Groups of Drosophila. BMC Evol
744 Biol. 2008 Jan 29;8:33.

745 22. Charlesworth B, Coyne JA, Barton NH. The Relative Rates of Evolution of Sex
746 Chromosomes and Autosomes. Am Nat. 1987;130(1):113–46.

747 23. Bressa MJ, Papeschi AG, Vítková M, Kubícková S, Fuková I, Pigozzi MI, et al. Sex
748 chromosome evolution in cotton stainers of the genus Dysdercus (Heteroptera:
749 Pyrrhocoridae). Cytogenet Genome Res. 2009;125(4):292–305.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. Wright CJ, Stevens L, Mackintosh A, Lawniczak M, Blaxter M. Chromosome evolution in Lepidoptera [Internet]. Evolutionary Biology; 2023 May [cited 2023 Dec 28]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.05.12.540473

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

33. 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

784 34. Hill J, Rastas P, Hornett E, Neethiraj R, Clark N, Morehouse N, et al. Unprecedented
785 reorganization of holocentric chromosomes provides insights into the enigma of
786 lepidopteran chromosome evolution. Sci Adv. 2019 Jun 12;5:eaau3648.

787 35. Cicconardi F, Lewis JJ, Martin SH, Reed RD, Danko CG, Montgomery SH.
788 Chromosome Fusion Affects Genetic Diversity and Evolutionary Turnover of
789 Functional Loci but Consistently Depends on Chromosome Size. Mol Biol Evol.
790 2021;38(10):4449–62.

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

794 37. de Castro, Zagrobelny M, Zurano JP, Zikan Cardoso M, Feyereisen R, Bak S.
795 Sequestration and biosynthesis of cyanogenic glucosides in passion vine butterflies
796 and consequences for the diversification of their host plants. Ecol Evol.
797 2019;9(9):5079–93.

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

801 39. Belleghem SMV, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA, et al.
802 Complex modular architecture around a simple toolkit of wing pattern genes. Nat Ecol
803 Evol. 2017;1(52).

804 40. Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, et al. Major
805 Improvements to the Heliconius melpomene Genome Assembly Used to Confirm 10
806 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. G3
807 GenesGenomesGenetics. 2016 Mar 1;6(3):695–708.

808 41. Chakraborty M, Lara AG, Dang A, McCulloch KJ, Rainbow D, Carter D, et al. Sex-
809 linked gene traffic underlies the acquisition of sexually dimorphic UV color vision in
810 Heliconius butterflies [Internet]. bioRxiv; 2022 [cited 2023 Jul 26]. p.
811 2022.07.04.498748. Available from:
812 https://www.biorxiv.org/content/10.1101/2022.07.04.498748v1

813 42. Jiggins CD, Lamas G. The Ecology and Evolution of Heliconius Butterflies [Internet].
814 Oxford University Press; 2016 [cited 2022 Jul 18]. Available from:
815 https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/97801995665
816 70.001.0001/acprof-9780199566570

817 43. Rueda-M N, Salgado-Roa FC, Gantiva-Q CH, Pardo-Díaz C, Salazar C.
818 Environmental Drivers of Diversification and Hybridization in Neotropical Butterflies.

819 Front Ecol Evol [Internet]. 2021 [cited 2022 Oct 31];9. Available from:
820 https://www.frontiersin.org/articles/10.3389/fevo.2021.750703

821 44. QGIS Development Team. QGIS Geographic Information System. Open Source
822 Geospatial Found Proj. 2018;URL:http://qgis.osgeo.org.

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

825 46. Mackintosh A, Laetsch DR, Baril T, Foster RG, Dincă V, Vila R, et al. The genome
826 sequence of the lesser marbled fritillary, Brenthis ino, and evidence for a segregating
827 neo-Z chromosome. G3 GenesGenomesGenetics. 2022 Jun 1;12(6):jkac069.

828 47. Chen X, Wang Z, Zhang C, Hu J, Lu Y, Zhou H, et al. Unraveling the complex
829 evolutionary history of lepidopteran chromosomes through ancestral chromosome
830 reconstruction and novel chromosome nomenclature. BMC Biol. 2023 Nov
831 20;21(1):265.

832 48. Jay P, Chouteau M, Whibley A, Bastide H, Llaurens V, Parrinello H, et al. Mutation
833 accumulation in chromosomal inversions maintains wing pattern polymorphism in a
834 butterfly [Internet]. bioRxiv; 2019 [cited 2023 Oct 18]. p. 736504. Available from:
835 https://www.biorxiv.org/content/10.1101/736504v1

836 49. Mackintosh A, Vila R, Martin SH, Setter D, Lohse K. Do chromosome rearrangements
837 fix by genetic drift or natural selection? A test in Brenthis butterflies [Internet]. bioRxiv;
838 2023 [cited 2023 Aug 13]. p. 2023.06.16.545248. Available from:
839 https://www.biorxiv.org/content/10.1101/2023.06.16.545248v1

840 50. Stewart NB, Ahmed-Braimah YH, Cerne DG, McAllister BF. Female meiotic drive
841 preferentially segregates derived metacentric chromosomes in *Drosophila* [Internet].
842 Genetics; 2019 May [cited 2024 Feb 12]. Available from:
843 http://biorxiv.org/lookup/doi/10.1101/638684

844 51. White MJD. Models of Speciation. Science. 1968 Mar 8;159(3819):1065–70.

845 52. Martin SH, Singh KS, Gordon IJ, Omufwoko KS, Collins S, Warren IA, et al. Whole-
846 chromosome hitchhiking driven by a male-killing endosymbiont. PLOS Biol. 2020 Feb
847 27;18(2):e3000610.

848 53. Boman J, Wiklund C, Vila R, Backström N. Meiotic drive against chromosome fusions
849 in butterfly hybrids [Internet]. bioRxiv; 2024 [cited 2024 Feb 9]. p. 2024.02.02.578558.
850 Available from: https://www.biorxiv.org/content/10.1101/2024.02.02.578558v1

54. 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.

55. 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 30]. p. 2022.08.12.503723. Available from: https://www.biorxiv.org/content/10.1101/2022.08.12.503723v1

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

57. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018 Nov 29;19(1):460.

58. 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

59. 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.

60. 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.

61. 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 ago;15(8):e1007273.

62. 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.

63. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094–100.

64. 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.

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

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

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

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

893 69. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation
894 PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015 Dec
895 1;4(1):s13742-015-0047–8.

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

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

900 72. Mirarab S, Reaz R, Bayzid MdS, Zimmermann T, Swenson MS, Warnow T. ASTRAL:
901 genome-scale coalescent-based species tree estimation. Bioinformatics. 2014 Sep
902 1;30(17):i541–8.

903 73. Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, et al.
904 Butterfly genome reveals promiscuous exchange of mimicry adaptations among
905 species. Nature. 2012 Jul 5;487(7405):94–8.

906 74. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder:
907 fast model selection for accurate phylogenetic estimates. Nat Methods. 2017
908 Jun;14(6):587–9.

909 75. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
910 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol.
911 2015 Jan 1;32(1):268–74.

912 76. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics.
913 2010 May 15;26(10):1372–3.

914 77. Patterson M, Marschall T, Pisanti N, Van Iersel L, Stougie L, Klau GW, et al. W HATS
915 H AP : Weighted Haplotype Assembly for Future-Generation Sequencing Reads. J
916 Comput Biol. 2015 Jun;22(6):498–509.

917 78. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate,
918 scalable and integrative haplotype estimation. Nat Commun. 2019 Nov 28;10(1):5436.

919 79. Speidel L, Forest M, Shi S, Myers SR. A method for genome-wide genealogy
920 estimation for thousands of samples. Nat Genet. 2019 Sep;51(9):1321–9.

921 80. Van Belleghem SM, Baquero M, Papa R, Salazar C, McMillan WO, Counterman BA,
922 et al. Patterns of Z chromosome divergence among Heliconius species highlight the
923 importance of historical demography. Mol Ecol. 2018;27(19):3852–72.

924 81. Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, et al.
925 Estimation of the Spontaneous Mutation Rate in Heliconius melpomene. Mol Biol Evol.
926 2015 Jan 1;32(1):239–43.

927 82. Tobler A, Kapan D, Flanagan NS, Gonzalez C, Peterson E, Jiggins CD, et al. First-
928 generation linkage map of the warningly colored butterfly Heliconius erato. Heredity.
929 2005 Apr;94(4):408–17.

930 83. Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of
931 BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed
932 Processing Symposium (IPDPS) [Internet]. 2019 [cited 2024 Jan 20]. p. 314–24.
933 Available from: https://ieeexplore.ieee.org/document/8820962

934 84. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. Vcflib and tools for
935 processing the VCF variant call format [Internet]. bioRxiv; 2021 [cited 2024 Feb 12].
936 p. 2021.05.21.445151. Available from:
937 https://www.biorxiv.org/content/10.1101/2021.05.21.445151v1

938 85. Bansal V. HapCUT2: A Method for Phasing Genomes Using Experimental Sequence
939 Data. In: Peters BA, Drmanac R, editors. Haplotyping: Methods and Protocols
940 [Internet]. New York, NY: Springer US; 2023 [cited 2023 Nov 3]. p. 139–47. (Methods
941 in Molecular Biology). Available from: https://doi.org/10.1007/978-1-0716-2819-5_9

942

943

944

945

Supporting information captions

S1 Table. Sample information and genotyping statistics.

S2 Table. Lepidoptera genome assembly statistics.

S3 Table. BUSCO results statistics.

S1 Fig. Synteny plots showing high collinearity between *Heliconius* genomes.

S2 Fig. Identification of the W chromosome in the genome of *H. sara*.

S3 Fig. Missing data and mean depth per individual.

S4 Fig. Principal Component Analysis (PCA), performed with 3,685,916 SNPs.

S5 Fig. Species tree based on ASTRAL multi-species coalescence and Densitree.

S6 Fig. Maximum likelihood phylogeny of chromosome 1.

S7 Fig. Maximum likelihood phylogeny of chromosome 2.

S8 Fig. Maximum likelihood phylogeny of chromosome 3.

S9 Fig. Maximum likelihood phylogeny of chromosome 4.

S10 Fig. Maximum likelihood phylogeny of chromosome 5.

S11 Fig. Maximum likelihood phylogeny of chromosome 6.

S12 Fig. Maximum likelihood phylogeny of chromosome 7.

S13 Fig. Maximum likelihood phylogeny of chromosome 8.

S14 Fig. Maximum likelihood phylogeny of chromosome 9.

964 **S15 Fig. Maximum likelihood phylogeny of chromosome 10.**

965 **S16 Fig. Maximum likelihood phylogeny of chromosome 11.**

966 **S17 Fig. Maximum likelihood phylogeny of chromosome 12.**

967 **S18 Fig. Maximum likelihood phylogeny of chromosome 13.**

968 **S19 Fig. Maximum likelihood phylogeny of chromosome 14.**

969 **S20 Fig. Maximum likelihood phylogeny of chromosome 15.**

970 **S21 Fig. Maximum likelihood phylogeny of chromosome 16.**

971 **S22 Fig. Maximum likelihood phylogeny of chromosome 17.**

972 **S23 Fig. Maximum likelihood phylogeny of chromosome 18.**

973 **S24 Fig. Maximum likelihood phylogeny of chromosome 19.**

974 **S25 Fig. Maximum likelihood phylogeny of chromosome 20.**

975 **S26 Fig. Maximum likelihood phylogeny of chromosome Z.**

976 **S27 Fig. Maximum Likelihood (ML) phylogenies inferred genome-wide and per**
977 **chromosome.**

978 **S28 Fig. Marginal tree for one SNPs in Chr4.**

979 **S29 Fig. Marginal tree for one SNPs in Chr9.**

980 **S30 Fig. Marginal tree for one SNPs in Chr14.**

S31 Fig. Genome-wide divergence (F_{ST}) between pairs of species in the *sara/sapho* clade.

S32 Fig. Genome-wide divergence (F_{ST}) between pairs of subspecies of *H. eleuchia, H. congener* and *H. sapho*.

S33 Fig. Genome-wide divergence (F_{ST}) between pairs of subspecies of *H. antiochus*.

S34 Fig. Genome-wide divergence (F_{ST}) between pairs of subspecies of *H. sara*.

S35 Fig. Genome-wide divergence (F_{ST}) in the *sara/sapho* clade.

S36 Fig. Patterns of heterozygosity across the genome in the *sara/sapho* clade.

S37 Fig. Patterns of mean depth across the genome in the *sara/sapho* clade.

S38 Fig. Patterns of heterozygosity and depth across chromosome (A) 14 and (B) 9.

S39 Fig. Proportion of heterozygous sites and mean depth between sexes in sliding windows along chromosome 4.

S40 Fig. Proportion of heterozygous sites and mean depth between sexes in sliding windows along chromosome (A) 14 and (B) 9.

S41 Fig. The density map of Hi-C contacts confirms W-A fusions in the sapho subclade.