High-quality haploid genomes corroborate 29 chromosomes and highly conserved synteny of genes in *Hyles* hawkmoths (Lepidoptera: Sphingidae)

Anna K. Hundsdoerfer [1#], Tilman Schell [2], Franziska Patzold [1], Charlotte J. Wright [3], Atsuo Yoshido [4], František Marec [4], Hana Daneck [1], Sylke Winkler [5], Carola Greve [2], Lars Podsiadlowski [6], Michael Hiller [2] & Martin Pippel [5, 7]

[1] Senckenberg Natural History Collections Dresden, Königsbrücker Landstr. 159, 01109 Dresden, Germany

[2] LOEWE-Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany.

[3] Tree of Life, Wellcome Sanger Institute, Cambridge, CB10 1SA, UK

[4] Biology Centre of the Czech Academy of Sciences, Institute of Entomology, Branišovská 31, 370 05 České Budějovice, Czech Republic

[5] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany

[6] Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 127, 53113 Bonn, Germany

[7] Center for Systems Biology Dresden, Pfotenhauerstr. 108, 01307 Dresden, Germany

[#] Corresponding author: anna.hundsdoerfer@senckenberg.de

Key words: hawkmoth; PacBio sequencing; Hi-C assembly; chromosome-level scaffolding; wing pattern genes

Abstract

Background: Similarity of wing patterns does not correlate with phylogenetic relationships in the young genus *Hyles*. Speciation is an ongoing process, as is hybridization because reproductive barriers are still low. Genomic data promise insights into various aspects of speciation in the spurge hawkmoth (*Hyles euphorbiae*), we thus assembled a chromosome-level genome and investigated some characteristics.

Results: The *H. euphorbiae* male genome was sequenced using PacBio and amended by Hi-C Illumina data, yielding a 504 Mb assembly (scaffold N50 of 18.2 Mb) with 99.9% of the data represented by the 29 largest scaffolds forming the haploid chromosome set. Consistent with this, FISH analysis of the karyotype revealed *n* = 29 chromosomes and a WZ/ZZ (female/male) sex chromosome system in *H. euphorbiae*. Estimates of chromosome length based on the karyotype image provided an additional quality metric of assembled chromosome size and confirmed the WZ sex chromosomes as the largest chromosome pair in the karyotype. Additional Hi-C data were used to scaffold the published *H. vespertilio* male genome into chromosomes, resulting in a high-quality assembly (651 Mb) with a scaffold N50 of 22 Mb and 98% of sequence data placed in the 29 largest scaffolds (chromosomes). The measured genome sizes of *H. euphorbiae* and *H. vespertilio* had an average 1C DNA value of 472 Mb and 562 Mb, respectively. The larger genome size of *H. vespertilio* was accompanied by a proportional increase in repeats from 45% in *H. euphorbiae* to almost 55% in *H. vespertilio*. In both *Hyles* species, the three wing pattern genes, *optix, WntA*, and *cortex*, were found on chromosomes 23, 28, and 17, respectively, with varying amounts and positions of indels possibly corrupting their function.

Conclusions: The comparison of high-quality annotated chromosome-level assemblies in *H. euphorbiae* and *H. vespertilio* revealed highly conserved synteny of

genes between these two species, but significant differences in repetitive DNA content. The identification of wing pattern genes promises future understanding of the evolution of forewing patterns in the genus *Hyles*, although further sequencing data from more individuals are needed. The genomic data obtained provide additional reliable references for further comparative studies in hawkmoths (Sphingidae).

Background

The genus *Hyles* comprises about 30 species, is distributed worldwide and, with an estimated divergence time of 5.7-9.3 million years (Mya) [1], is rather young from an evolutionary perspective [2]. Some species are still in the process of speciation, others hybridize because reproductive barriers are still low. These processes are difficult to discern in *Hyles euphorbiae*, which has been overly split into five species in the past [3]. Genomic data promise insights into numerous aspects of speciation, including ongoing gene flow, food plant utilization, and the genetic basis for morphological differentiation. Therefore, we sequenced and examined the genome of *Hyles euphorbiae*. The spurge hawkmoth *Hyles euphorbiae* Linnaeus 1758 (family Sphingidae) is a charismatic Palearctic species with large, colorful, and polymorphic larvae and camouflaged, heavy adults with strong flight abilities. Despite their aposematic coloration, the larvae do not sequester the toxic spurge diterpene esters [4] from their food plants. Instead, they protect themselves by spewing the plant slurry that has not yet been detoxified. Larval specialization on toxic host plants of the genus *Euphorbia* has evolved twice independently within the genus *Hyles* [5, 6]. The impressively high morphological variability of larvae has complicated the taxonomy of

H. euphorbiae by contributing to an overestimation of species diversity (overview in Hundsdoerfer et al. [3]). Similarly, the high intraspecific diversity of mitochondrial marker genes bedeviled the reconstruction of the molecular phylogeny of the formerly five species (synonymized with *H. euphorbiae* in 2019 [5]), but provided valuable resolution for phylogeography [7]. While some larval patterns are correlated with geography [7] and are thus expected to be based on underlying genetic diversity causing phenotypic variability, others appear to be environmentally driven by phenotypic plasticity.

Early studies have shown that wing pattern similarity does not correlate with phylogenetic relationships in *Hyles* [1, 2]. Seven basic wing patterns are observed in the Central Palearctic *Hyles* species, which neither correlate with currently defined species, nor reflect phylogenetic relationships within the genus. Recently, we standardized forewing patterns for members of the genus, defined morphological characters and coded these for about 200 individuals in a matrix [8]. Grouping in the tree reconstructions of these morphological data could be characterized as being based completely on wing pattern. One such group consists of moths with many stripes on the forewings (the former subgenus ‘Danneria’ according to Danner et al. [9]), but this has been refuted as a clade by previous molecular phylogenetic work [2]. Another large clade includes moths that predominantly show a pattern of dark brown spots and stripes on a lighter, cream-colored background. This corresponds to the typical forewing pattern of *H. euphorbiae*, and thus the group included all species with a similar wing pattern [8] (or slight variations thereof), although many did not show a close molecular phylogenetic relationship [10] to this species. Another clade consists of species with forewing patterns that lack many or most of the pattern elements described above, including *H. vespertilio*, whose genome was recently

published [11]. The forewings of this species lack a wing pattern [11], raising the possibility that the gene(s) controlling wing pattern formation are inactive, making genomic comparisons with this species particularly intriguing.

Interspecific differences in forewing patterns within the genus *Hyles* should be based on detectable genetic differences, as patterns are stable within species in the well-separated, oldest Neotropical (and Nearctic) taxa [2]. In the Palearctic, incomplete lineage sorting and ongoing hybridization impede such insights, justifying ongoing systematic, phylogenetic and taxonomic research (e.g. Patzold et al. [12]).

Numerous Sphingidae genomes are currently being published (e.g. Smerinthinae *Hyles vespertilio*, Macroglossinae, by Pippel et al. [11], *Mimas tiliae*, Smerinthinae, by Boyes et al. [13], *Laothoe populi*, Smerinthinae [14]). This wealth of data will provide insights into the evolution of wing patterns in hawkmoths. Pioneering studies of the genes underlying the wing pattern of Lepidoptera in the genus *Heliconius* (family Nymphalidae) [15] have revealed a modular architecture with narrow stretches of the genome associated with specific differences in color and pattern. It has been previously proposed that the evolution of lepidopteran wing pattern stripes occurred through the repeated gain, loss, and modification of only a handful of serially repeated elements [16]. In addition, variation in colors and certain color patterns of species is expected to be primarily driven by biotic and abiotic factors [17] acting as a selection pressure on individuals with differential expression of genotypes.

Optix is a single-exon gene in *Heliconius* [15, 18] encoding a homeobox transcription factor that regulates the production of ommochrome pigments [19] and is associated with red and orange forewing patterning in these butterflies. Nearby non-coding regions control its expression in *Heliconius* wing development [15] and these are the

regions of divergence, while the coding region is a conserved homeobox gene. The forewings of *H.euphorbiae* are known for their pink flush in some regions of its distribution range (see moths in Fig. 3 of Hundsdoerfer, Lee et al. [3]), whereas red is not prominent in the grey wings of *H. vespertilio*.

The *WntA* signaling ligand is highly conserved at the amino acid level within *Heliconius* [15]) and is associated with the forewing black band. Variation is again more likely to be influenced by expression during wing development and is therefore found in the nearby control region [15].

The wing pattern gene *cortex* is associated with the yellow patterns on *Heliconius* wings [20] and has been suggested to underlie switches in color pattern across Lepidoptera [21]. For example, it has been shown that insertion of a transposable element (TE) into an intron of *cortex* results in industrial melanism [22], i.e. black morphs in the peppered moth (*Biston betularia*; Geometridae), corroborating its function across Lepidoptera, due to, e.g., the *cortex* TE insertion. This insertion [22], estimated to have occurred around 1819, could only manifest itself in the population through a selection pressure against the light-colored moths showing up on the dark trees and thus experiencing higher predation. Overall, it is expected that this set of conserved, flexible wing patterning genes has driven the rapid morphological diversification in the genus *Hyles*.

In contrast to *H. vespertilio* [11], in which the forewings have a naturally occurring knock-out appearance (as if the wing pattern genes were dysfunctional) of near-uniform grey wings lacking high-contrast patterning, *H. euphorbiae* shows important elements of the typical ground forewing pattern of the genus, which has been reconstructed as the ancestral set of characters for proto-*Hyles* [8]. By comparing the chromosome-level genomes of *H. euphorbiae* and *H. vespertilio*, the present work

aims to provide reference genome data for future studies to understand the origin of *Hyles* wing patterns as phenotypic variability starting with the wing pattern control genes described for *Heliconius* and other nymphalid butterflies (e.g. [15, 16, 20]).

Results

Hyles euphorbiae genome assembly with chromosome-level scaffolding

The *H. euphorbiae* genome was assembled with Pac-Bio HiFi reads [23] combined with Hi-C data. Assembly contiguity statistics are summarized in Table 1. The final contig assembly consists of 56 scaffolds with an N50 of 18.2 Mb (321 contigs with an N50 of 2.76 Mb) and a size of 504 Mb and is available under the NCBI SRA accession number SRR17432892, BioProject PRJNA794104, BioSample SAMN24610150 and genome JALBCW000000000, JALBCX000000000. The annotated circular mitochondrial genome has length of 15,322 bp. Final assembly statistics are summarized in Table 2.

Table 1: Contiguity statistics of *Hyles euphorbiae* genome.

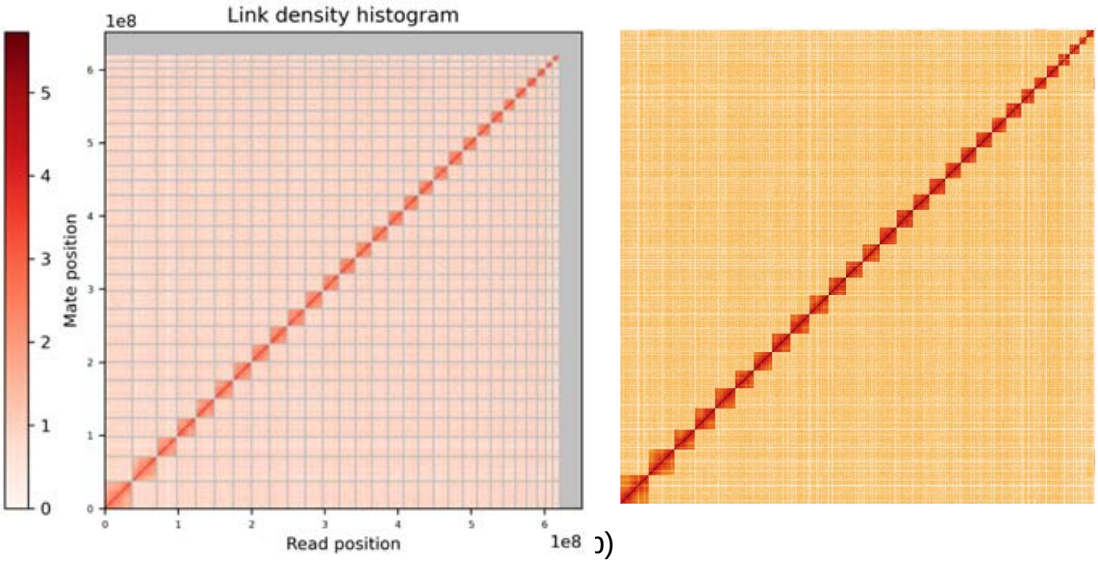
	Contigs	Initial scaffolds	Curated scaffolds	Curated contigs
#Sequences	592	81	56	322
Total length	513,190,344	513,292,544	504,259,600	504,323,440
Largest sequence	6,062,269	30,760,442	30,347,856	10,621,849
N50	1,441,051	18,456,932	18,182,747	2,758,341

Table 2: Assembly statistics of presented and related species. *Ms*: *Manduca sexta* [24], *Mt*: *Mimas tiliae* [13], *Lp*: *Laothoe populi* [14], *Hf*: *Hemaris* fuciformis [25], *Dp: Deilephila porcellus* [26], *He*: *H. euphorbiae, Hv*: *Hyles vespertilio* , the new assembly of *H. vespertilio* was used (based on PacBio sequence data from Pippel et al. [11]).

	Ms	*Mt*	*Lp*	*Hf*	*Dp*	*He*	*Hv*
# contigs (>= 0 bp)	4056	30	33	43	31	56	391
# contigs (>= 1000 bp)	4056	30	33	42	31	56	390
# contigs (>= 5000 bp)	3527	30	33	36	31	49	389
# contigs (>= 10000 bp)	2822	30	33	34	31	47	380
# contigs (>= 50000 bp)	61	29	29	30	29	34	251
Total length (>= 0 bp) Mb	470.0	478.0	576.4	448.9	402.1	504.3	651.4
Total length (>= 1000 bp) Mb	470.0	478.0	576.4	448.9	402.1	504.3	651.4
Total length (>= 5000 bp) Mb	468.5	478.0	576.4	448.8	402.1	504.3	651.4
Total length (>= 10000 bp) Mb	463.1	478.0	576.4	448.8	402.1	504.3	651.4
Total length (>= 50000 bp) Mb	410.9	478.0	576.3	448.7	402.0	504.0	647.4
Largest contig (Mb)	21.11	25.98	29.55	25.57	22.66	30.35	37.22
Total length (Mb)	470.0	478.0	576.4	448.9	402.1	504.3	651.4
GC (%)	35.59	38.21	38.02	37.09	36.34	36.24	36.58
N50 (Mb)	14.25	17.90	21.13	16.55	15.07	18.18	22.14
N75 (Mb)	12.41	15.70	18.59	14.30	12.99	16.57	20.07
L50	14	12	13	12	12	12	13
L75	23	19	20	20	20	19	20
# N's per 100 kb	230.91	0.65	5.29	3.26	0.4	12.66	2.24
BUSCO (lepidoptera_odb10):							
complete, single-copy (%)	91.8	98.5	98.5	98.4	98.6	97.9	95.4
complete, duplicated (%)	6.5	0.3	0.4	0.3	0.2	0.3	2.9
fragmented (%)	0.7	0.3	0.3	0.3	0.6	0.5	0.7
missing (%)	1.0	0.9	0.8	1.0	0.9	1.3	1.0

Chromosome scaffolding with *Hyles vespertilio* Hi-C data

HiRise [27] scaffolding of *H. vespertilio* using more than 132 million read pairs yielded a scaffold N50 of 22.1 Mb. In total, 146 joins and only one break of the input assembly were introduced. The contact map clearly shows 29 well-supported scaffolds representing chromosomes (Fig. 1). Sizes of chromosomes of both *Hyles* species are of the same order of magnitude, but not identical (Table 3). The assembly of *H. vespertilio* is almost 147 Mb larger than that of *H. euphorbiae*.



a) b)

Fig. 1. Contact map of a) *Hyles vespertilio* (Dovetail HiRise analysis) and b) *Hyles euphorbiae* (as described in M&M). Chromosome-level scaffolding clearly supports 29 scaffolds representing chromosomes. Chromosomes are ordered by size from bottom left (largest) to top right (smallest).

Table 3: Chromosome sizes of final assemblies and chromosome numbering in homology to *Bombyx mori*. Chromosome 1 corresponds to the Z chromosome [28]. Footnotes: * also contains parts of BmChr26, ** also contains parts of BmChr11, § also contains parts of BmChr23, # also contains part of BmChr24.

***B. mori* Chr. number**	***H. vesper-tilio* Chr. number**	***H. vespertilio* scaffold name in browser**	**Genebank Accession number**	**Assembly Length (bp) incl. gaps**	***H. eu-phorbiae* Chr. Number**	***H. euphorbiae* scaffold name in browser**	**Genebank Accession number**	**Assembly Length (bp) incl. gaps**
BMChrZ	HvChrZ	ScWp86a_195_HRSCAF_280	CM042833.1	37,218,281	HeChrZ	sc_1	CM041058.1	30,347,856
BmChr2	HvChr2*	ScWp86a_159_HRSCAF_226	CM042805.1	24,553,817	HeChr2	sc_7	CM041030.1	19,904,380
BmChr3	HvChr3	ScWp86a_257_HRSCAF_395	CM042806.1	20,065,448	HeChr3	sc_18	CM041031.1	16,855,035
BmChr4	HvChr4	ScWp86a_21_HRSCAF_27	CM042807.1	25,450,662	HeChr4	sc_6	CM041032.1	20,022,273
BmChr5	HvChr5	ScWp86a_119_HRSCAF_165	CM042808.1	26,040,715	HeChr5	sc_4	CM041033.1	21,578,258
BmChr6	HvChr6	ScWp86a_136_HRSCAF_194	CM042809.1	22,136,963	HeChr6	sc_13	CM041034.1	17,936,700
BmChr7	HvChr7	ScWp86a_95_HRSCAF_134	CM042810.1	17,182,831	HeChr7	sc_23	CM041035.1	14,557,976
BmChr8	HvChr8	ScWp86a_214_HRSCAF_310	CM042811.1	21,873,546	HeChr8	sc_14	CM041036.1	17,505,000
BmChr9	HvChr9	ScWp86a_232_HRSCAF_355	CM042812.1	24,188,504	HeChr9	sc_11	CM041037.1	19,134,990
BmChr10	HvChr10	ScWp86a_42_HRSCAF_61	CM042813.1	24,738,224	HeChr10	sc_8	CM041038.1	19,757,600
BmChr11	HvChr11	ScWp86a_178_HRSCAF_255	CM042814.1	22,098,621	HeChr11	sc_15	CM041039.1	17,388,153
BmChr12	HvChr12	ScWp86a_182_HRSCAF_262	CM042815.1	26,852,440	HeChr12	sc_5	CM041040.1	21,432,827
BmChr13	HvChr13	ScWp86a_68_HRSCAF_94	CM042816.1	24,145,169	HeChr13	sc_10	CM041041.1	19,486,524
BmChr14	HvChr14	ScWp86a_111_HRSCAF_155	CM042817.1	20,079,433	HeChr14	sc_21	CM041042.1	15,684,570
BmChr15	HvChr15	ScWp86a_62_HRSCAF_86	CM042818.1	26,014,704	HeChr15	sc_3	CM041043.1	21,768,765
BmChr16	HvChr16	ScWp86a_55_HRSCAF_78	CM042819.1	18,679,767	HeChr16	sc_22	CM041044.1	15,570,054
BmChr17	HvChr17	ScWp86a_210_HRSCAF_302	CM042820.1	24,352,496	HeChr17	sc_9	CM041045.1	19,686,965
BmChr18	HvChr18	ScWp86a_1_HRSCAF_1	CM042821.1	20,909,092	HeChr18	sc_16	CM041046.1	17,373,644
BmChr19	HvChr19	ScWp86a_207_HRSCAF_299	CM042822.1	20,353,206	HeChr19	sc_19	CM041047.1	16,571,156
BmChr20	HvChr20	ScWp86a_378_HRSCAF_523	CM042823.1	16,497,537	HeChr20	sc_24	CM041048.1	12,812,799
BmChr21	HvChr21	ScWp86a_162_HRSCAF_230	CM042824.1	21,114,399	HeChr21	sc_17	CM041049.1	17,181,161
BmChr22	HvChr22#	ScWp86a_157_HRSCAF_223	CM042825.1	33,800,775	HeChr22	sc_2	CM041050.1	27,093,441

BmChr23	HvChr23	ScWp86a_31_HRSCAF_41	CM042826.1	22,933,977	HeChr23	sc_12	CM041051.1	18,182,747
BmChr24	HvChr24	ScWp86a_166_HRSCAF_234	CM042827.1	11,894,872	HeChr24	sc_27	CM041052.1	10,049,873
BmChr25	HvChr25	ScWp86a_7_HRSCAF_7	CM042828.1	19,247,757	HeChr25	sc_20	CM041053.1	15,917,517
BmChr26	HvChr26**	ScWp86a_81_HRSCAF_111	CM042829.1	9,224,630	HeChr26	sc_29	CM041054.1	7,206,840
BmChr27	HvChr27	ScWp86a_289_HRSCAF_430	CM042830.1	14,304,874	HeChr27	sc_26	CM041055.1	12,048,215
BmChr28	HvChr28#	ScWp86a_140_HRSCAF_198	CM042831.1	15,210,208	HeChr28	sc_25	CM041056.1	12,771,623
n.a.	HvChr29§	ScWp86a_137_HRSCAF_195	CM042832.1	9,587,503	HeChr29	sc_28	CM041057.1	7,786,824

Chromosome structure in Sphingidae

Comparison of single copy orthologs in the *H. euphorbiae* genome assembly to those in *H. vespertilio* shows strong conservation of both synteny and chromosome structure, defined as conservation of gene order. This suggests that the last common ancestor of *Hyles* had 29 chromosomes that subsequently remained intact (Fig. 2a). Next, the orthologs of *H. euphorbiae* were compared to those of *Deilephila porcellus* (Fig. 2b), both of which belong to Macroglossinae, a subfamily of Sphingidae. This also revealed a set of 29 orthologous chromosomes, suggesting that chromosome structure has remained stable throughout the evolution of this subfamily. To determine whether the stability of this set of 29 chromosomes extends throughout Sphingidae, we compared the orthologs of two further species, *Manduca sexta* (*n* = 28; Fig. 2c) of the subfamily Sphinginae and *Laothoe populi* (*n* = 28; Fig. 2d), which belongs to the third subfamily Smerinthinae. This revealed that 27 chromosomes of *M. sexta* are orthologous and highly syntenic compared to *H. euphorbiae.* The remaining chromosome in *M. sexta* (CM026259.1) maps to two autosomes in *H. euphorbiae* (CM041057.1, CM041033.1), thus accounting for the difference in karyotype. Similarly, one chromosome of *Laothoe populi* (HG992147.1) is orthologous to two autosomes in *H. euphorbiae* (CM041039.1, CM041057.1). However, the two chromosomes are different in *L. populi* and *M. sexta*, suggesting two independent fusion events in these two species. This would suggest that the Sphingidae originally had a karyotype of *n* = 29 that has remained largely intact with a limited number of fusion events in extant species. Moreover, the strong conservation of gene order in *H. euphorbiae* compared to *M. sexta* illustrates limited

intrachromosomal rearrangements since the last common ancestor of the Sphingidae.

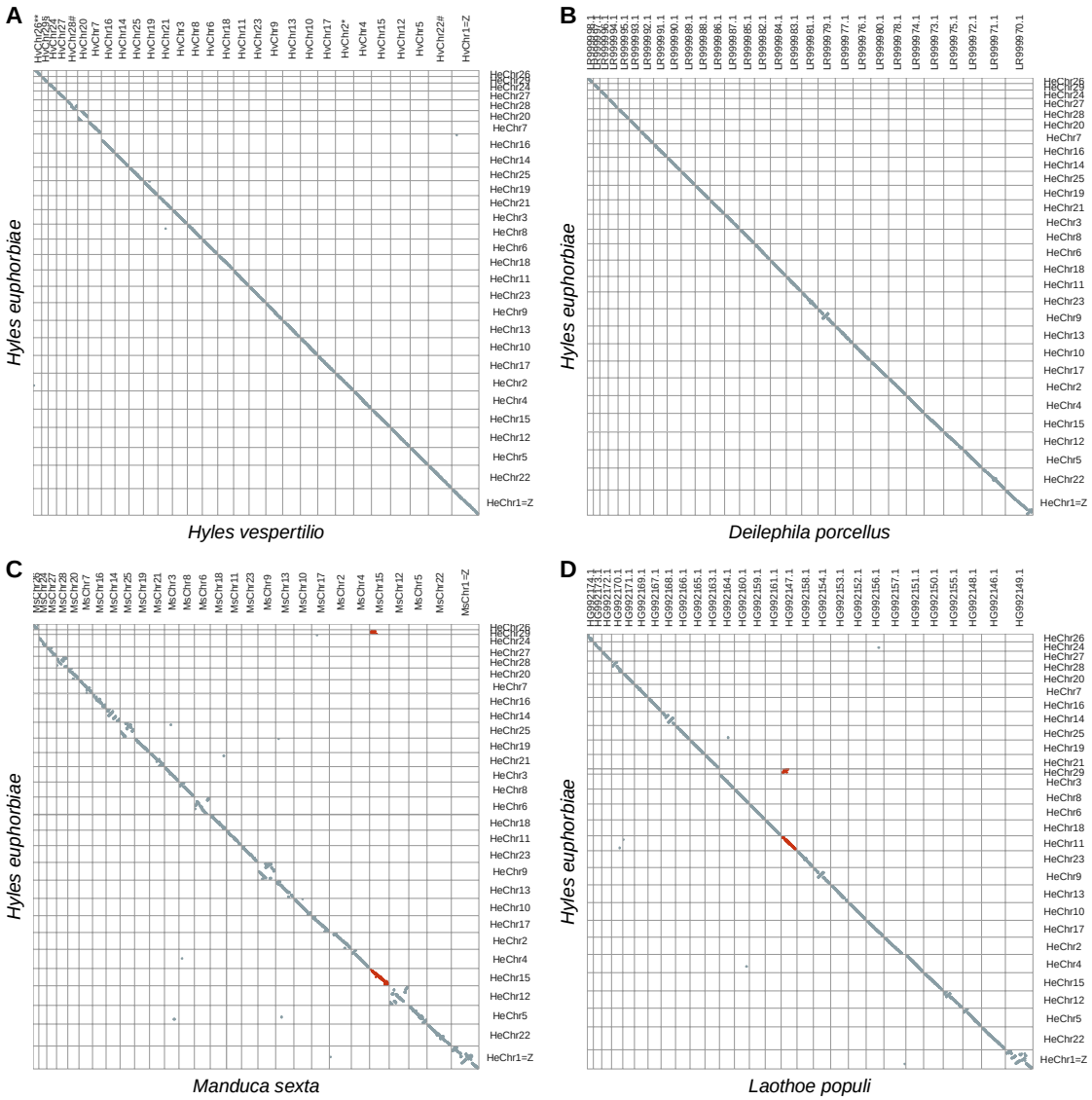


Fig. 2. Oxford plots for a.) *Hyles vespertilio,* b.) *Deilephila porcellus* and c.) *Manduca sexta* d.) *Laothoe populi* versus *Hyles euphorbiae*. Comparison of assemblies demonstrate strong conservation of both linkage groups and gene order in Sphingidae. Orthologs colored in red belong to chromosomes which are inferred to be involved in a fusion event in the species displayed on the x-axis.

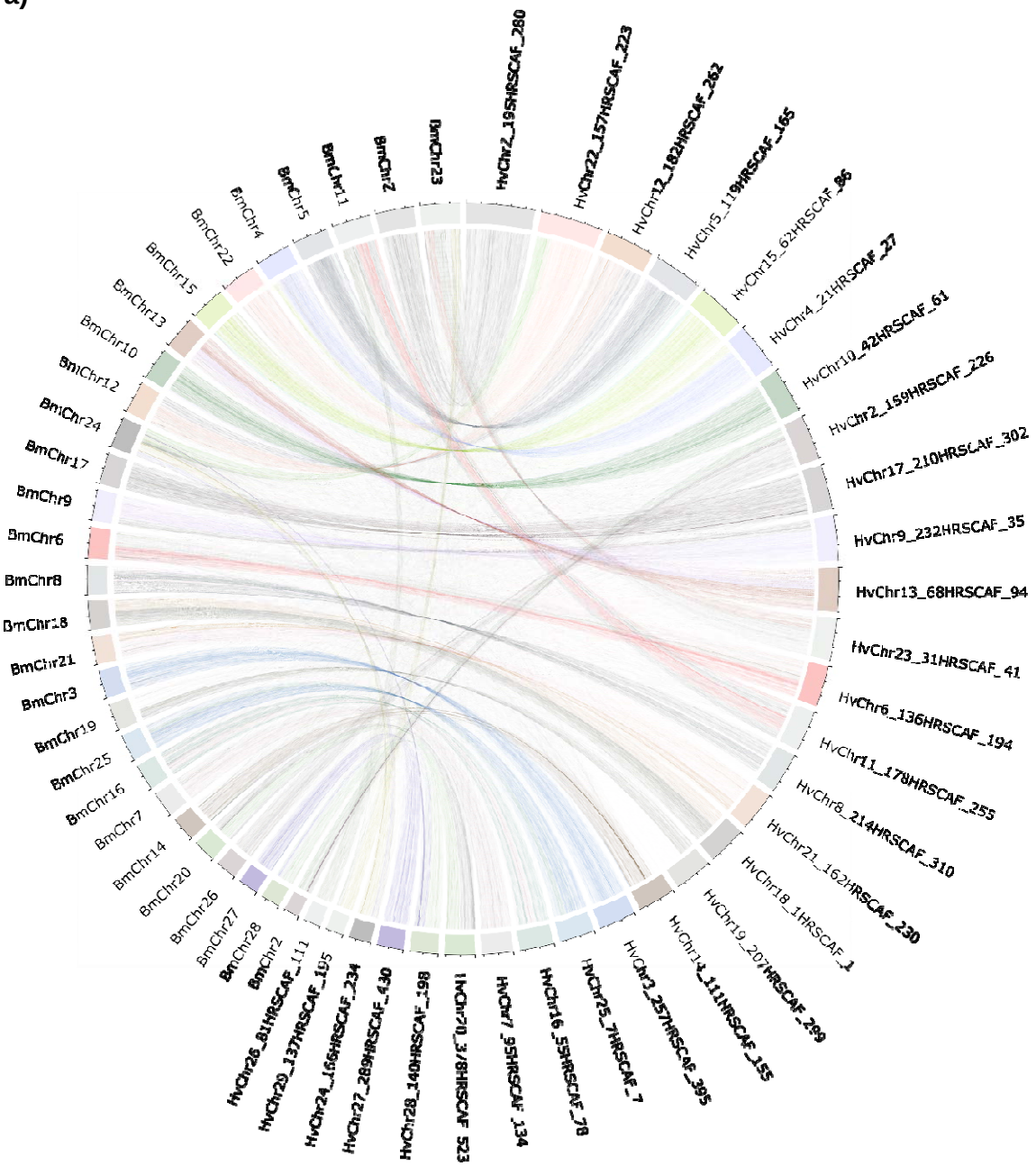
Genome alignments

The two species of the genus *Hyles* have one more chromosome than *B. mori* (*n* = 28), but the alignment between *H. vespertilio* and *B. mori* can still be illustrated (Fig. 3a) to allow comparison by eye.

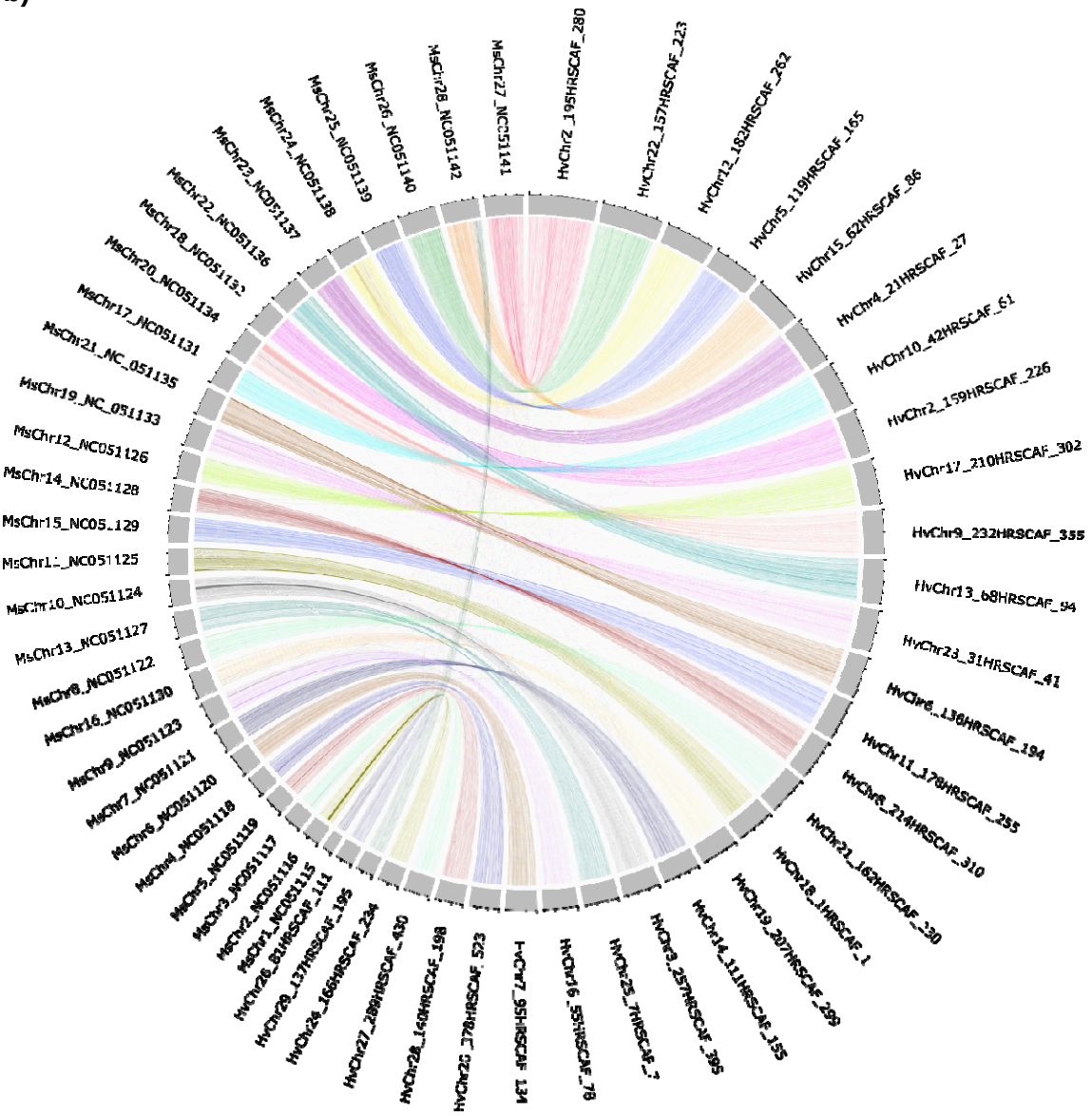
CIRCOS plots of the alignment between the two species show well-preserved synteny, except for some chromosomes that could be involved in chromosome rearrangements. Chromosome 2 (HvChr2) of *H. vespertilio* corresponds to chromosomes 2 and 26 of *B. mori* (BmChr2, BmChr26). Chromosome 24 of *B. mori*, BmChr24, corresponds to parts of *H. vespertilio* chromosomes 22, 24, and 28 (HvChr22, HvChr24, HvChr28), and chromosome 11 of *B. mori* (BmChr11) corresponds to parts of *H. vespertilio* chromosomes 11 and 26 (HvChr11, HvChr26). These chromosome rearrangements between *B. mori* and *H. vespertilio* are consistent with those between *B. mori* and *M. sexta* [24, 29], corroborating the strong synteny between the two hawkmoth species, *H. vespertilio* and *M. sexta* (Fig. 3b). Chromosome 23 of *B. mori* (BmChr23) is split into *H. vespertilio* chromosomes 23 and 29 (HvChr23, HvChr29), resulting in an additional chromosome in *H. vespertilio* (HvChr29). In *M. sexta*, which has 28 chromosomes, HvChr29 and HvChr15 correspond to MsChr28, thus showing a different combination.

The plot comparing the 29 chromosome sequences of *H. vespertilio* and *H. euphorbiae* (Fig. 3c) illustrates the highly conserved synteny within the genus *Hyles* in the definition of collinearity, the conservation of block order in the two sets of chromosomes. The larger size of the *H. vespertilio* genome based on the genome size estimation and its longer assembly than that of *H. euphorbiae* are reflected by the larger size of every chromosome (Fig. 3c).

a)


BmChr23
HvChr2_195HRSCAF_280
HvChr22_157HRSCAF_223
HvChr12_182HRSCAF_262
HvChr5_119HRSCAF_165
HvChr15_62HRSCAF_86
HvChr4_21HRSCAF_27
HvChr10_42HRSCAF_61
HvChr2_159HRSCAF_226
HvChr17_210HRSCAF_302
HvChr9_232HRSCAF_35
HvChr13_68HRSCAF_94
HvChr23_31HRSCAF_41
HvChr6_136HRSCAF_194
HvChr11_178HRSCAF_255
HvChr8_214HRSCAF_310
HvChr21_162HRSCAF_230
HvChr18_1HRSCAF_1
HvChr19_207HRSCAF_299
HvChr14_111HRSCAF_155
HvChr3_257HRSCAF_395
HvChr25_7HRSCAF_7
HvChr16_55HRSCAF_78
HvChr7_95HRSCAF_134
HvChr20_378HRSCAF_523
HvChr28_140HRSCAF_198
HvChr27_289HRSCAF_430
HvChr24_166HRSCAF_234
HvChr29_137HRSCAF_195
HvChr26_81HRSCAF_111
BmChr2
BmChr28
BmChr27
BmChr26
BmChr20
BmChr14
BmChr7
BmChr16
BmChr25
BmChr19
BmChr3
BmChr21
BmChr18
BmChr8
BmChr6
BmChr9
BmChr17
BmChr24
BmChr12
BmChr10
BmChr13
BmChr15
BmChr22
BmChr4
BmChr5
BmChr11
BmChr2

b)


MsChr27_NC051141
HvChr2_195HRSCAF_280
HvChr22_157HRSCAF_223
HvChr12_182HRSCAF_262
HvChr5_119HRSCAF_165
HvChr15_62HRSCAF_86
HvChr4_21HRSCAF_27
HvChr10_42HRSCAF_61
HvChr2_159HRSCAF_226
HvChr17_210HRSCAF_302
HvChr9_232HRSCAF_355
HvChr13_68HRSCAF_94
HvChr23_31HRSCAF_41
HvChr6_138HRSCAF_194
HvChr11_178HRSCAF_255
HvChr8_214HRSCAF_310
HvChr21_162HRSCAF_230
HvChr18_1HRSCAF_1
HvChr19_207HRSCAF_299
HvChr14_111HRSCAF_155
HvChr3_257HRSCAF_395
HvChr25_7HRSCAF_7
HvChr16_55HRSCAF_78
HvChr7_95HRSCAF_134
HvChr20_378HRSCAF_523
HvChr28_148HRSCAF_198
HvChr27_289HRSCAF_430
HvChr24_166HRSCAF_224
HvChr29_137HRSCAF_195
HvChr26_81HRSCAF_111
MsChr1_NC051115
MsChr2_NC051116
MsChr3_NC051117
MsChr5_NC051119
MsChr4_NC051118
MsChr6_NC051120
MsChr7_NC051121
MsChr9_NC051123
MsChr16_NC051130
MsChr8_NC051122
MsChr13_NC051127
MsChr10_NC051124
MsChr11_NC051125
MsChr15_NC05_129
MsChr14_NC051128
MsChr12_NC051126
MsChr19_NC_051133
MsChr21_NC_051135
MsChr17_NC051131
MsChr20_NC051134
MsChr18_NC051132
MsChr22_NC051136
MsChr23_NC051137
MsChr24_NC051138
MsChr25_NC051139
MsChr26_NC051140
MsChr28_NC051142

c)

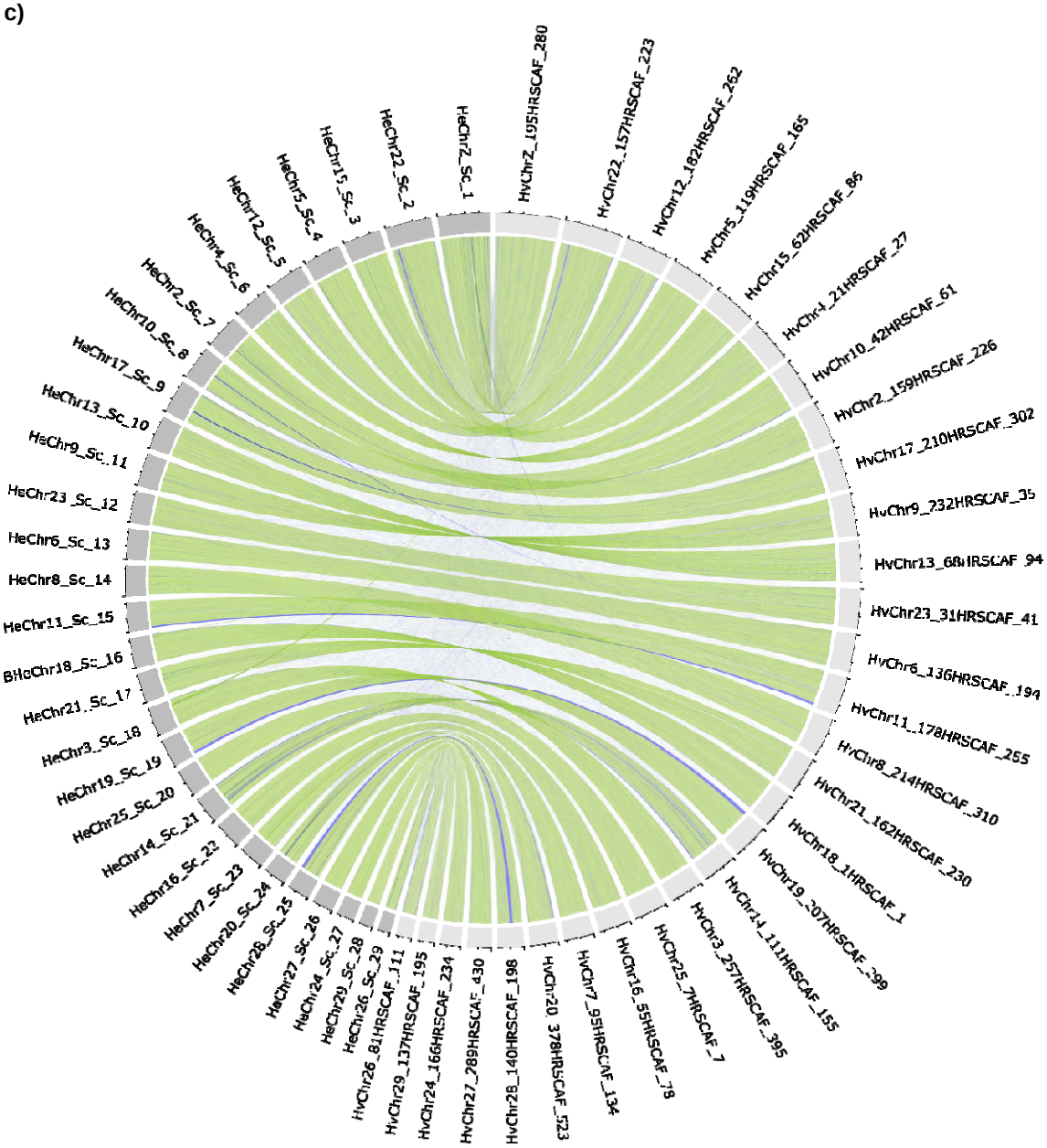


Fig. 3. CIRCOS plots of genome alignments at the chromosome level. The Z chromosome corresponds to chromosome 1. Chromosomes of *H. vespertilio* are ordered by size. The letters "ScWp86a" in the scaffold names of *Hyles vespertilio* are omitted for clarity. **a**) *Bombyx mori* and *H. vespertilio* with automatic coloring using shinyCircos to facilitate correlation. **b**) *Manduca sexta* and *H. vespertilio* with coloring by chromosome. **c**) *H. euphorbiae* and *H. vespertilio*, with the 29 chromosomes of each species (*H. vespertilio* dark grey boxes in the outer rim, *H. euphorbiae* light grey boxes). The color of the links represents the strand orientation of the chromosomes of *H. euphorbiae* compared to the *H. vespertilio* strand orientation (green means +; blue means -).

Annotation and methylation profile

The proportion of the genome covered by the major classes of repetitive elements is 37% and 47% for *H. euphorbiae* and *H. vespertilio*, respectively (Table 4, Fig. 4). Thus, *Hyles vespertilio* has a much larger genome (651 Mbp to 504 Mbp in *H. euphorbiae*) and a higher proportion of repetitive DNA than *H. euphorbiae*. A comprehensive comparison of transposon content in seven species of the family Sphingidae (Fig. 4) demonstrates the dynamics of repetitive elements and the contribution of recent bursts in transposable element families to genome size expansion in *H. vespertilio* and independently in *Laothoe populi*.

The number of genes per chromosome is very similar in the two species (Fig. 5), but exhibits a wide range from a minimum of 169 genes in both species on chromosome 29 to 826 / 798 genes in *H. vespertilio* / *H. euphorbiae* on chromosome 22.

Table 4. Assembly lengths and proportion of repeats of the available genomes in comparison (a new assembly of *H. vespertilio* was used, based on PacBio sequence data from Pippel et al. [11]).

Species	**Total length**	**Masked [%]**
Manduca sexta [24]	470,036,997	27.8
Mimas tiliae [13]	472,718,480	46.7
L. populi [13]	549,962,493	47.7
H. fuciformis [25]	441,141,813	33.2
D. porcellus [26]	389,730,088	27.5
H. euphorbiae	504,310,614	37.1
H. vespertilio	651,427,907	46.6

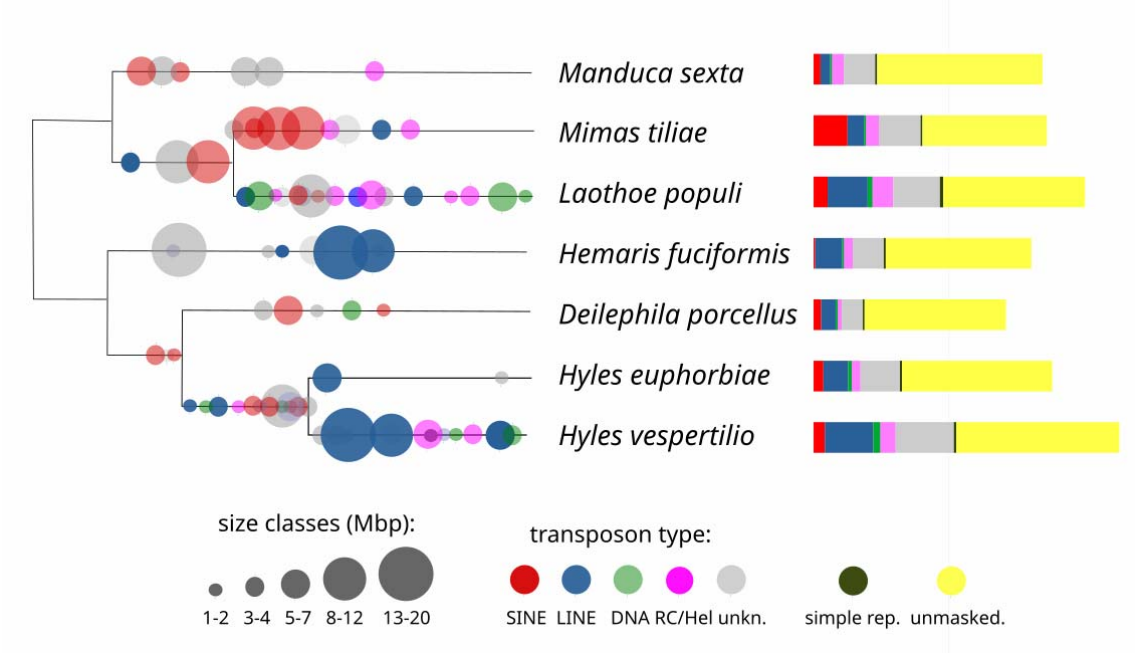


Fig. 4. Transposon bursts mapped on a phylogenetic tree drawn by hand according to backbone topology in [30] and proportional repeat contents of seven sphingid species with available genome data (transposon types: SINE: short interspersed elements, LINE: long interspersed elements, DNA: DNA transposons, RC/Hel: rolling-circle transposons, e.g. helitrons).

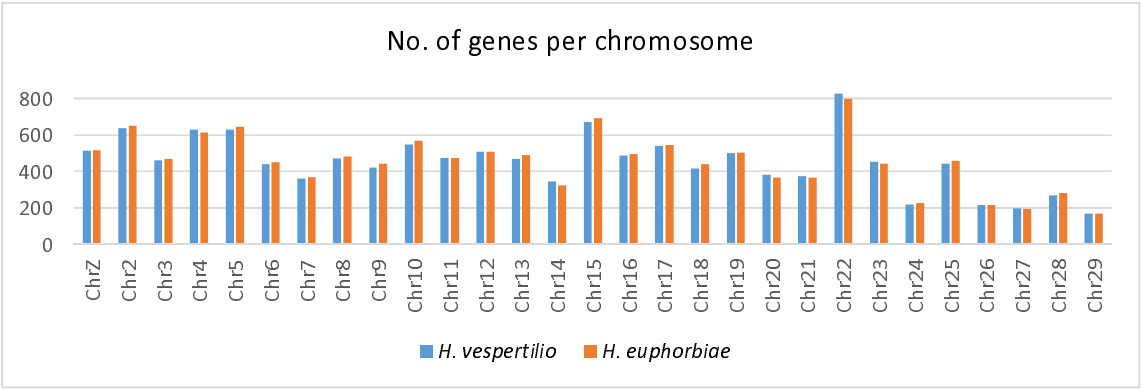


Fig. 5. Number of genes per chromosome for *H. euphorbiae* and *H. vespertilio* in comparison.

The mitochondrial genome of *H. euphorbiae* (Fig. S1) contains 13 protein coding genes, two ribosomal RNA (rRNA) genes plus 22 transfer RNA (tRNA) sequences and the control region.

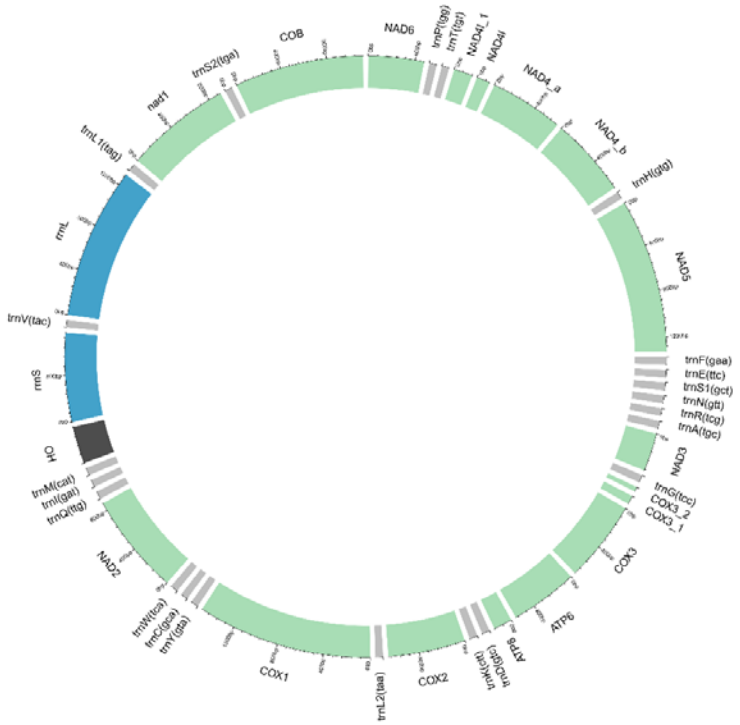


Fig. S1. The annotated mitochondrial genome of *Hyles euphorbiae*.

The methylation profile of *H. euphorbiae* (track “Expression and Regulation“ in the genome browser; e.g. Fig. S3b) shows modification probabilities for every site, whereby base positions that have a high probability of modification / to be methylated and thus not be expressed are marked red (> 50% probability).

Wing pattern genes

The wing pattern genes *optix, WntA*, and *cortex* were identified with high confidence using BLAT [31, 32], only one hit was found each with thresholds of identity above 85% and bit score above 550.

The hits of the *H. erato optix* protein sequence (Table S1) revealed a position on chromosome 23 in *H. euphorbiae* (HeChr23, "sc_12:1,582,862-1,583,680", size: 819 bp) and *H. vespertilio* (HvChr23, "ScWp86a_31_HRSCAF_41:20,799,551-20,800,375", size: 825 bp). No repetitive elements were annotated in the area of the coding sequence in either species, however upon zooming out 10x or 100x a high amount of indels are shown in the genome browser (400 KB up-& downstream of the coding sequence).

Table S1: Accession numbers of reference wing pattern sequences and data type used for blat in the genome browser. *Manduca sexta WntA* and *optix* are not displayed on the NCBI web portal.

Gene	**Reference species**	**Accession number**	**Blat**
WntA	*Heliconius himera*	AFC75686.1	Protein sequence
optix	*Heliconius erato*	JN102349.1	Protein sequence
cortex (isoform X1)	*Manduca sexta*	XM_030179809	mRNA sequence

The *Heliconius himera WntA* amino acid sequence (Table S1) mapped to the *Hyles* genomes returning a position on chromosome 28 in *H. euphorbiae* (HeChr28, “sc_25:5,829,059-5,837,439”, spanning six exons and 8,381 bp; Fig. S2a; annotated by transcript LOC115447391_rna-XM037443222.1.47 of *Manduca sexta* spanning seven exons and 17,247 bp) and *H. vespertilio* (HvChr28, “ScWp86a_140_HRSCAF_198:8,200,403-8,210,485”, spanning six exons and 10,083 base pairs, Fig. S2b; annotated by evm07703.91 spanning seven exons and 14,395 bp). There are numerous differences in the repeats between the two *Hyles* species. For example, *H. vespertilio* contains a 443 bp long DNA repeat insertion (“ScWp86a_140_HRSCAF_198:8,200,940-8,201,382”; rnd-4_family-526; family TcMar-Tc1, pos. 512-1311 in the repeat) in intron 2, which *H. euphorbiae* lacks. The

genomic interval spanning the exons and introns of *WntA* is inverted in *H. euphorbiae* (or *H. vespertilio*, depending on the polarity definition).

Fig. S2a)

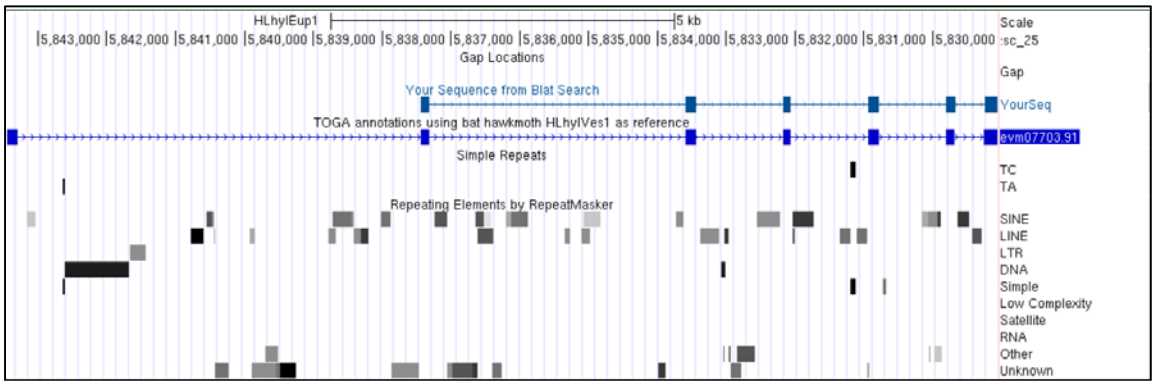


Fig. S2b)

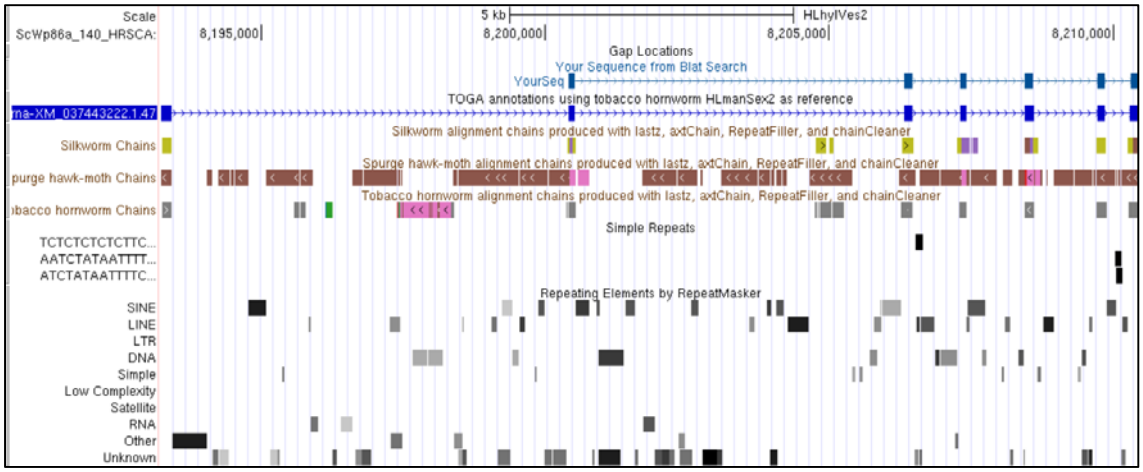


Fig. S2. View of the *Hyles* genome alignment, annotation, and repeat occurrence in the vicinity of the BLAT result of the *Heliconius WntA* protein sequence AFC75686 in **a**) *H. euphorbiae* (reverse) and **b**) *H. vespertilio* showing the genomic region in *B. mori*, *H. euphorbiae* (inverted), and *M. sexta*.

BLAT search of the *M. sexta cortex* protein sequence resulted in a match corresponding to 8 exons on chromosome 17 in both species [Fig. S3; *H. euphorbiae* (HeChr17, “sc_9:7,671,034-7,674,090”) and *H. vespertilio* (HvChr17, “ScWp86a_210_HRSCAF_302:14,798,018-14,801,349”)]. The methylation profile in

H. euphorbiae (Fig. S3b) indicates normal transcription activity in this stretch of exons, as the probabilities of base modifications are all < 50% (blue). The position, quality, and quantity of repeats in introns differ between the two species. For example, the first intron in *H. euphorbiae* contains no repeats, whereas in *H. vespertilio* it contains two LINEs in addition to a 211 bp stretch (ScWp86a_210_HRSCAF_302:14,798,399-14,798,609; Fig. S3a) of unknown repeats (named rnd-1_family-40). It overlaps downstream with the LINE of family L2, which shows an irregular simple repeat of 5' TATT. It is in intron 1 of the cortex X1 sequence of *M. sexta* (annotated with *M. sexta* transcript LOC115451467_rna-XM_030179809.2.8; Fig. S3a, c). Again, the area surrounding the gene cortex is inverted in *H. euphorbiae* (or *H. vespertilio*, depending on the polarity definition). The 100 kb view reveals a high number of repeats in the vicinity of the stretch of gene exons in both species (Fig. S4).

Fig. S3a)

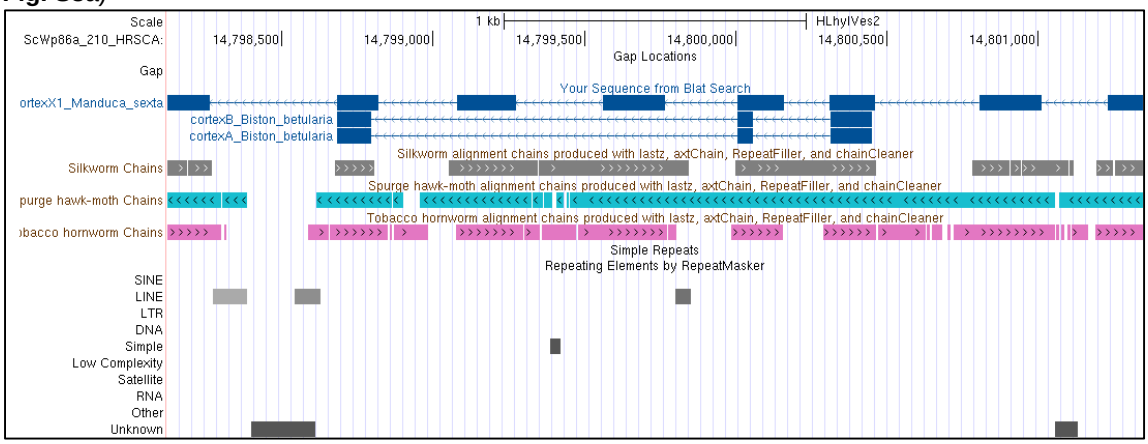


Fig. S3b)

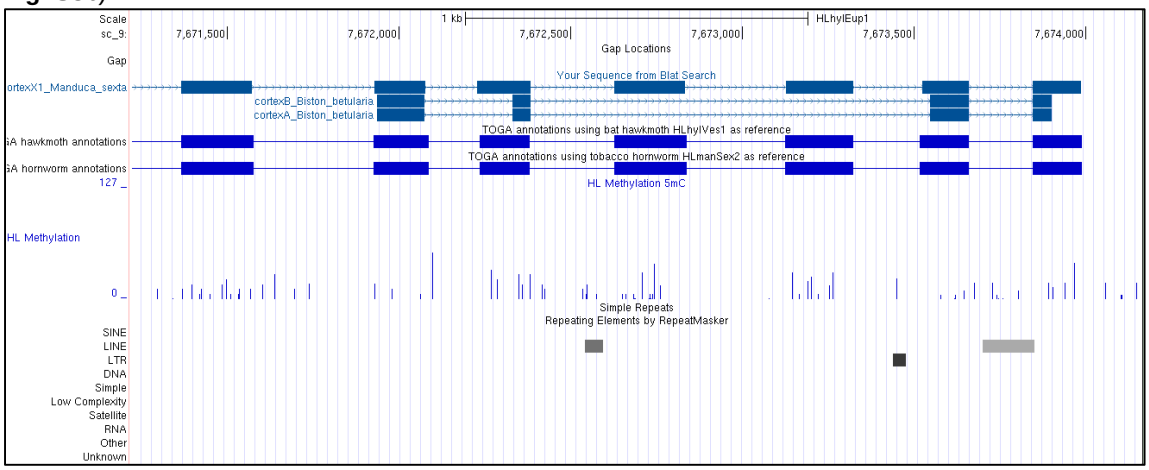


Fig. S3. BLAT result from the genome browser showing the exon/intron structures of the *Manduca sexta* and *Biston betularia cortex* protein (Table S1) on chromosome 17 in (**a**) *Hyles euphorbiae* and (**b**) *H. vespertilio*. The track "HL Methylation" shows the *H. vespertilio* methylation pattern analyzed. The exon-intron structure annotated with TOGA (bright blue). Other sequence colors are pink (*M. sexta*), blue-green (*H. euphorbiae*), grey (*B. mori*).

Fig. S4a)

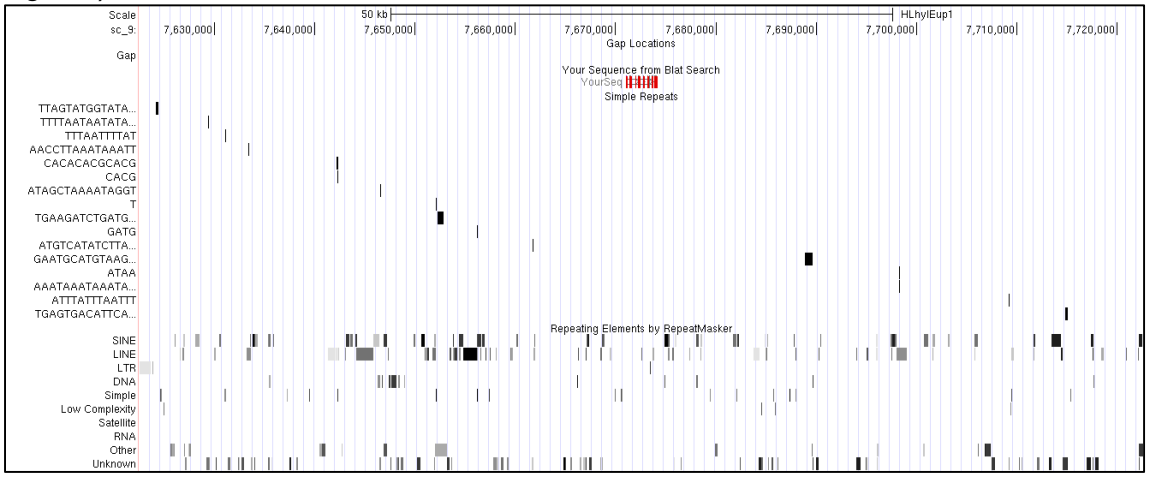


Fig. S4b)

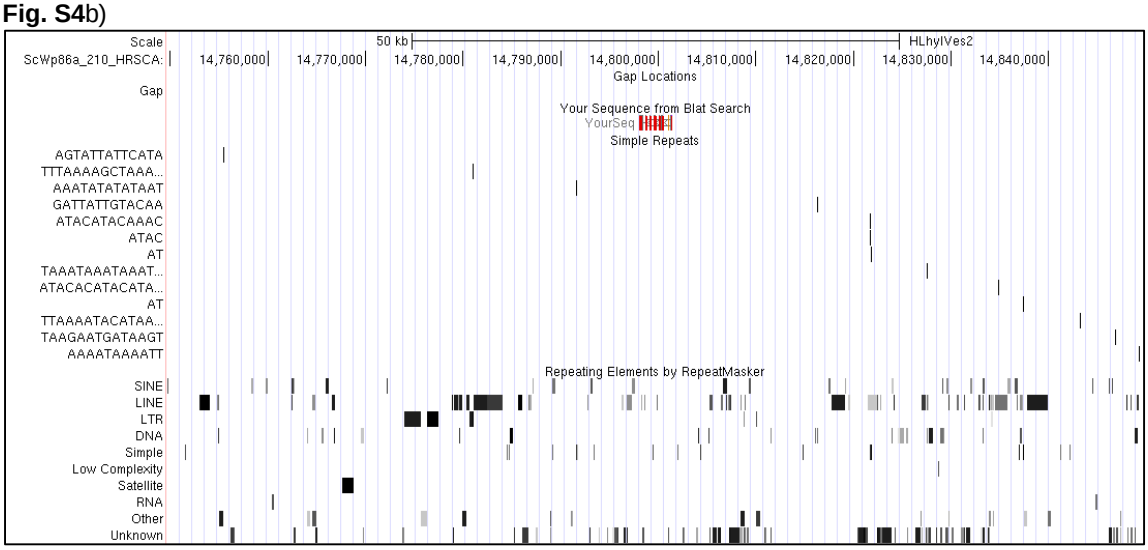


Fig. S4. BLAT result of *Manduca sexta* mRNA of *cortex* (Table S1) on chromosome 17 in (**a**) *Hyles euphorbiae* and (**b**) *Hyles vespertilio* in a 100 kb view showing a large number of repeats in the vicinity, with position, length, and type differing between species.

Karyotype

Analysis of male mitotic chromosomes stained by FISH with a telomeric probe (telomere-FISH) showed that the karyotype of *H. euphorbiae* is composed of 2*n* = 58 chromosomes (Fig. 6a). As typical for Lepidoptera, the chromosomes are of the holokinetic type, i.e. they lack a primary constriction (centromere) and are morphologically uniform, differing only in size. Chromosome number was confirmed by analysis of meiotic nuclei at the pachytene stage, where homologous chromosomes pair and form elongated bivalents. Pachytene complements, stained by GISH in combination with telomere-FISH, showed a haploid number of 29 bivalents in both sexes (Fig. 6b, c). In addition, GISH identified a WZ sex chromosome bivalent in pachytene oocytes by labeling a large portion of the W chromosome with the female genomic DNA (gDNA) probe (Fig. 6c), whereas no bivalent was identified in pachytene spermatocytes (Fig. 6b). These results clearly

show that *H. euphorbiae* has a WZ/ZZ (female/male) sex chromosome system, which is common in Lepidoptera. It should be noted that the WZ bivalent is relatively long (Fig. 6c), suggesting that the W and Z chromosomes are among the largest chromosomes in the *H. euphorbiae* karyotype.

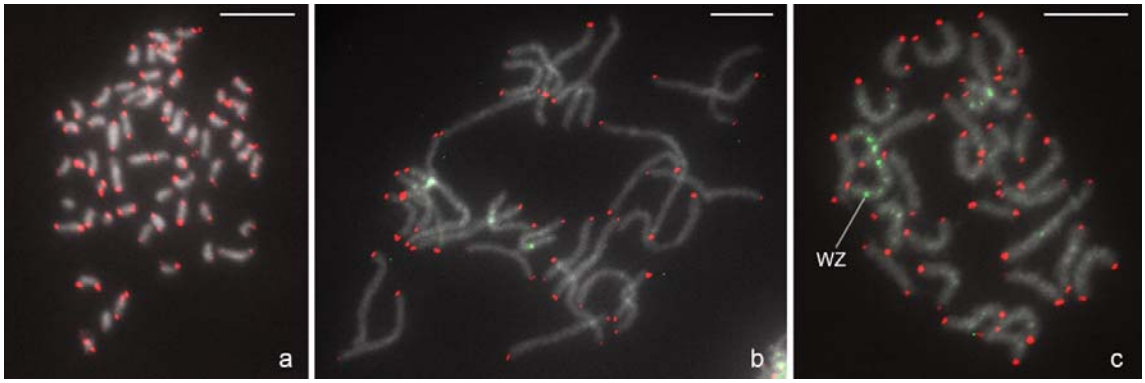


Fig. 6. **Molecular cytogenetic analysis of *Hyles euphorbiae* chromosomes.** Hybridization signals of the Cy3-labeled $(TTAGG)_n$ telomeric probe (red) indicate chromosomal ends (**a**–**c**), and the fluorescein-labeled female gDNA probe (green) identifies the sex chromosome system (**b** and **c**). Chromosomes were stained with DAPI (grey). (**a**) Male mitotic prometaphase stained with telomere-FISH showing a diploid chromosome number of 2*n* = 58. (**b**) Male pachytene complement stained with a combination of GISH and telomere-FISH showing 29 bivalents, but without any bivalent highlighted, indicating a ZZ sex chromosome constitution. (**c**) Female pachytene complement stained with a combination of GISH and telomere-FISH, showing 29 bivalents, including the WZ sex chromosome pair identified by the W chromosome highlighted with the female gDNA probe. Bar = 10 µm.

Chromosome size estimation from karyotype image data

Chromosome size estimation is based on Fig. 6c, bivalents from a female gonad cell at the pachytene stage. The chromosome size estimates are shown in Fig. 7. They corroborate the WZ bivalent as the largest chromosome. Based on semi-automated image processing, the software package napari-karyotype [33] relies on threshold-based image segmentation to detect chromosome-related components. Identified chromosomal objects are surrounded by red rectangles and labeled with the estimates.

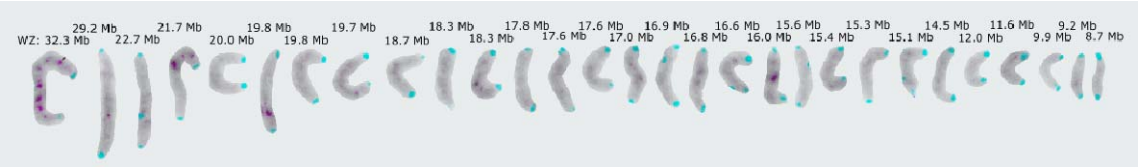


Fig. 7. Annotated chromosome bivalents from a female pachytene image of *Hyles euphorbiae* using the software package napari-karyotype. Chromosome bivalents were aligned manually according to the estimated size. The largest chromosome pair represents the sex chromosomes, in this case a WZ bivalent.

Genome size estimation

Genome sizes of *H. vespertilio* and *H. euphorbiae* were measured in three replicates, each from a single individual. The average results of 0.575 pg and 0.484 pg showed that the 1C DNA values were 562 Mb and 472 Mb, respectively (1 pg = 978 Mb). Genome size estimation for *H. euphorbiae* using ModEst revealed 535 Mb.

Discussion

Assembly quality

The chromosomal genome assemblies of *H. euphorbiae* and *H. vespertilio* presented here are highly contiguous, as demonstrated by the scaffold N50 values exceeding those of the sequenced close relatives *M. sexta* [24] and *B. mori* [28]. The percentage of assembly length that could be placed in chromosome scaffolds is highest in *Hyles euphorbiae* (Table 2). Both metrics underline the high quality of the two new assemblies for the *Hyles* genus. Gene content is similarly high in all compared assemblies.

Comparison of two *Hyles* genomes

The large difference in genome size estimates between the two species within the genus *Hyles*, i.e. *H. vespertilio* with 562 Mb being ~20% larger than that of *H. euphorbiae* with 472 Mb in the flow cytometry measurements, was unexpected due to the short divergence time between the two species [34]. *Hyles lineata* is one of the first *Hyles* species to diverge within the genus [5] and has an even smaller flow cytometry genome size of 450 Mb (0.46 pg) [35]. The potentially smaller genome size of this third *Hyles* species suggests that the larger genome of *H. vespertilio* is the result of a recent proliferation of repetitive elements and is a derived character.

Although the values of genome sizes based on assembly lengths are not directly comparable to genome size values based on flow cytometry due to technical differences in methodology, the relative sizes are in the same order of magnitude. The assembly length of *H. vespertilio* with 651 Mb is ~30% longer than that of *H. euphorbiae* with 504 MB, although the number of genes is very similar in both species (Fig. 5).

The genome of *H. euphorbiae* presented in this work has much fewer repetitive elements, in particular LINEs, than *H. vespertilio*. The high repeat content in all four species compared, *H. euphorbiae, H. vespertilio, M. sexta*, and *B. mori,* is typical for lepidopteran genomes [36, 37]. However, the number of repeats varies between species. Previous research has shown a correlation between the repetitive fraction of the genome and genome size within and between species [38, 39]. Indeed, the genome of *H. vespertilio* is the largest in our comparison and the genome with the most extensive repeat content, whereas the other sphingid genomes show lower

amounts of repeats, correlating with their smaller genome size. Therefore, as described in previous research, the repetitive elements found here are thus likely the driving force for the expansion of genome size, possibly due to a positive feedback loop that allows these elements to spread more easily in large genomes [40]. In addition, the repeatome is thought to play a significant role in genetic innovation and speciation [41-45]. Since *H. vespertilio* is known to be one of the most isolated species in the genus of hawkmoths, this could be a reasonable argument in our case. In contrast, *H. euphorbiae* is well known for its wide distribution and frequent hybridization with even distantly related species [2, 34, 46].

Since the two species are separated by a *p*-distance of only 2.1% in the neutral marker genes *COI* and *COII* [34], corresponding to a split timing of only 4.2-5.6 Mya (according to the respective molecular clock 1.1–1.2% per million years per lineage, resulting in 2.3% sequence divergence between species pairs per million years of separation [47]), high synteny of genes between *H. vespertilio* and *H. euphorbiae* in the nuclear chromosomes set was expected (Figs. 2, 3). Indeed, the illustrated genome alignments (Figs. 2, 3) show how similar the two *Hyles* genomes are. The mitochondrial genome of *H. euphorbiae* (Fig. S1) is also very similar to that of *H. vespertilio* (see [12]). Comparison of single copy orthologs in each genome demonstrates conservation of 29 orthologous chromosomes and a high degree in synteny, defined as conservation of gene order within each chromosome. Extending the comparisons beyond *Hyles* suggests that these 29 chromosomes have remained stable throughout Sphingidae, with the exception of two independent fusion events in *M. sexta* and *L. populi,* which are absent in *Hyles* and *D. porcellus.* Such strong conservation of chromosome structure has been previously observed in Nymphalidae

[48]. Our analyzes presented here suggest that the Sphingidae are also remarkably syntenic over long periods of evolution.

Wing pattern genes

The location of the three major wing pattern genes has been well studied in *Heliconius*. Here, we infer the likely locations of *optix*, *WntA*, and *cortex* in *Hyles* based on amino acid similarity and characterize their genomic context.

optix

optix and surrounding genes are highly conserved in Lepidoptera [49]. Zhang et al. [18] have shown that *optix* knock-outs exhibit complete replacement of color pigments by melanins, resulting in black and grey butterflies. In contrast, the high percentage of indels found in the alignment up- & downstream of *optix* between the genomes of *H. vespertilio* and *H. euphorbiae* may support the previously proposed hypothesis [50] that wing patterns are indeed controlled by cis-regulatory elements near the position of *optix*. As more sequence data from more individuals become available, F_{ST} plots will be calculated to corroborate correlations between particular candidate regions. However, the causal relationship between grey wings in *H. vespertilio* and *optix* and surrounding gene regions will only be answered by analyzing knock-outs generated using the CRISPR/Cas9 system.

WntA

The highly varying amounts, classes, and families of repeats between the two *Hyles* species sequenced suggest that the function of this gene could also have been

corrupted in the grey *H. vespertilio*, as this species has the larger genome and more repeats. Sequence data from more individuals are needed to pinpoint a candidate region and establish a robust correlation.

cortex

Another striking difference between the two species near the stretch of 8 *cortex* exons is the insertion of 229 bp, which was marked as an unknown repeat (Fig. S3c) in *H. vespertlilio*. Mapping this fragment to the genome of *H. euphorbiae* using BLAT yielded 200 hits between 89.8-96.7% identity on each chromosome and 203 hits on the new genome of *H. vespertilio* (data not shown), indicating the wide distribution of this repeat region across the genomes. It was also found in the very closely related macroglossine species *Deilephila porcellus* [26] and *Hemaris fuciformis* [25], as well as several hits in the genome of the smerinthine *Mimas tiliae* [13], a phylogenetically somewhat more distant species. Further analyzes and comparisons with the genomes presented here are expected to lead to a better understanding of the evolution of wing patterns in the family Sphingidae.

Karyotype, chromosome taxonomy

Genome alignment of the *B. mori* and *H. vespertilio* assemblies (Fig. 3a) allowed homology assessment of the chromosomes, resulting in the robust chromosome taxonomy that we presented for the two *Hyles* species studied (Fig. 3c). Given the young age of the genus (the oldest intra-genus split is 10.8-12.4 Mya, based on *p*-distance [2]), it can be assumed that it also applies to the entire genus *Hyles*. We propose our protocol of homology inference based on genome alignments as the gold standard for chromosome taxonomy of the Bombycoidea. The most recent

hawkmoth assemblies [13, 26], including that of *M. sexta* [24], use only arbitrary numbers (see Table S2) according to chromosome size. A standard in bombycoid chromosome taxonomy will greatly facilitate discussions based on genomic comparisons for specific research questions, such as the homologies of wing pattern genes in the present study.

Table S2 (separate spreadsheet): Chromosome and scaffold name conversion table of *Bombyx mori*, *Manduca sexta*, *Hyles euphorbiae*, and *H. vespertilio*.

In the karyotype evolution of Lepidoptera, chromosomes BmChr11, BmChr23 and BmChr24 are often split in other species, increasing the chromosome number compared to *B. mori*, e.g. [29, 51, 52]. The ancestral number of chromosomes in the haploid genome of Lepidoptera is thought to be *n* = 31 [48, 52, 53]. Thus, the karyotypes of *Bombyx mori* (n = 28) and *Hyles* (n = 29) underwent a reduction in chromosome number from this ancestral number, independently of e.g. *Heliconius* (*n* = 21). The well-known chromosome architecture in this genus was studied using linkage maps [54-56].

The *H. euphorbiae* chromosome images obtained had a sufficient clarity to be annotated with a size estimate by napari-karyotype (Tab. 3) [33]. However, it should be noted that it was more difficult than expected because the chromosome bivalents at the pachytene stage were touching each other in the image and thus had to be extracted manually. Without this step, the assignment of object vs. background would have been inaccurate. Furthermore, chromosomes are flexible structures and their length depends on the stage, the degree of condensation and also on the preparation methods. Therefore, their measured length does not always correspond to their size, which is especially true for meiotic chromosomes at the pachytene stage ([57]). In

order to annotate the chromosome images of the karyotype with the chromosome numbers correlated to *B. mori* chromosome taxonomy, it will be necessary to implement further in-situ-hybridization with gene-specific fluorochrome-labeled probes following Yasukochi et al. [29] (their Figure 2) in the future.

Conclusions

Previous studies had already shown that similarity of wing patterns does not correlate with phylogenetic relationships in the genus *Hyles* [1, 2, 5], neither as currently defined by morphology nor as reflected by molecular phylogeny. Morphologists have long argued that they must rely on phenotypic characters in the absence of molecular data, and have often included striking differences in wing patterns in their species descriptions. In *Hyles*, however, the evolution of wing pattern characters does not reflect the evolution of the species. Of course, gene trees are not species trees [58], so even traditional genetic analyzes, e.g. [1, 2, 5], do not necessarily reflect the true tree.

In this study, we present two high-quality annotated chromosome-level assemblies and report the presence, sequence, and location of wing pattern genes. Using reference genomes of *H. euphorbiae* and *H. vespertilio*, which have very different forewing pattern and coloration, we open up possibilities for studying the evolution of wing patterns based on a numerically evaluable, objective data source. The three genes *optix*, *WntA* and *cortex* promise to be of interest in a comparative study of numerous *Hyles* individuals, as the genomic regions surrounding these genes show

high numbers of indels potentially corrupting their function. The results of our comparative repeatome analysis confirm that repetitive elements are likely the driving force for the expansion of genome size in *Hyles vespertilio* compared to *H. euphorbiae*. The chromosome-level genomic data provided in this study for these two species provide reliable references in the family Sphingidae for future studies involving as many species as possible to elucidate the evolution of forewing patterns in this group of Lepidoptera.

Material and Methods

Insects

For karyotype analysis, specimens of the *H. euphorbiae* population from Greece (leg. P. Mazzei, Serifos) were reared in the laboratory (summer 2019). Several larvae and young pupae were used to make chromosome preparations (see below). For genome sequencing, one male specimen of *H. euphorbiae* (Fig. 8) was collected near Berbisdorf (Germany) on 27 July 2021 and a second, similar male moth from the same locality was placed as a voucher specimen in the SNSD collection (Senckenberg Naturhistorische Sammlungen Dresden).

Fig. 7. A male of the spurge hawkmoth (*Hyles euphorbiae*) collected near Berbisdorf.

Karyotype analysis

Spread chromosome preparations were made as described by Yoshido et al. [59]. Mitotic chromosomes were obtained from wing imaginal discs or testes of last instar larvae. Meiotic chromosomes in the pachytene stage of prophase I were obtained either from the testes of last instar larvae or from the ovaries of 3–5-day old pupae. Briefly, tissues were dissected in saline solution, swollen in hypotonic solution (75 mM KCl) for either 5 min (ovaries) or 15 min (testes and wing imaginal discs) and then fixed in Carnoy's fixative (ethanol, chloroform, acetic acid, 6:3:1) for 10 min. Cells dissociated in 60% acetic acid were spread on a heating plate at 45°C. All chromosome preparations were passed through a graded ethanol series (70%, 80%, and 100%, 30 s each) and stored at –80°C.

Fluorescence in situ hybridization (FISH) with the $(TTAGG)_n$ telomeric probe and genomic in situ hybridization (GISH) were performed according to the procedure described by Yoshido et al. [60]. $(TTAGG)_n$ telomeric sequences were generated by

non-template PCR following the protocol of Sahara et al. [61]. Male and female genomic DNAs (gDNAs) of *H. euphorbiae* were obtained separately from last instar larvae by standard phenol-chloroform extraction. DNA probes were labeled by nick translation using a mixture of DNase I and DNA polymerase I (both Thermo Fisher Scientific, Waltham, MA, USA) with either aminoallyl-dUTP-Cy3 or fluorescein-12-dUTP (both Jena Bioscience, Jena, Germany).

Chromosome preparations were removed from the freezer, passed through the graded ethanol series, air-dried and then denatured in 70% formamide in 2× SSC for 3.5 min at 70°C. For one preparation, the probe cocktail contained 500 ng of fluorescein-labeled female gDNA, 100 ng of Cy3-labeled telomeric probe, 3 µg of unlabeled sonicated male gDNA, and 25 µg of sonicated salmon sperm DNA (Sigma-Aldrich, St. Louis, MO, USA) in 10 µl hybridization buffer (50% formamide, 10% dextran sulfate in 2× SSC). Denaturation of the probe cocktail was performed for 5 min at 90°C. The preparations were examined under a Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany). Digital images were captured with an Olympus CCD monochrome camera XM10 equipped with cellSens 1.9 digital imaging software (Olympus Europa Holding, Hamburg, Germany) and processed with Adobe Photoshop CS4.

Karyotype-based automated chromosome annotation and size estimation

The meiotic chromosome image was preprocessed using the image processing software GIMP (version 2.10) [62] to manually cut out individual chromosome bivalents. The processed image was loaded into the tool napari-karyotype (version

c41103e) [33]. The image segmentation threshold, blur factor and genome size were set to 0.13, 0.5 and 504 Mb, respectively.

Genome size estimation

Genome sizes of two hawkmoth species were estimated following the flow cytometry protocol with propidium iodide-stained nuclei as described previously [63]. Neural tissue from frozen (−80°C) adult samples of *H. vespertilio* and *H. euphorbiae* (from one head each; the same individuals used for the PacBio genome and Hi-C sequencing) and neural tissue from the internal reference standard *Acheta domesticus* (female, 1C = 2 Gb) were each chopped with a razor blade in a Petri dish containing 2 ml of ice-cold Galbraith buffer. The suspension was filtered through a 42-µm nylon mesh, then stained with the intercalating fluorochrome propidium iodide (PI, Thermo Fisher Scientific) and treated with RNase A (Sigma-Aldrich), each at a final concentration of 25 µg/ml. The mean red PI fluorescence of the stained nuclei was quantified using a Beckman-Coulter CytoFLEX flow cytometer with a solid-state laser emitting at 488 nm. Fluorescence intensities of 10,000 nuclei per sample were recorded. Subsequently, the nuclei suspensions of *H. vespertilio* and *H. euphorbiae* were each mixed with the nuclei suspension of the internal reference standard (see above) and again the fluorescence intensities of 10,000 nuclei per mixed sample were recorded. For the histogram analyses, we used CytExpert 2.3 software. The total amount of DNA in each sample of the two *Hyles* species was calculated as the ratio of the mean fluorescence signal of the 2C peak of the stained nuclei of the respective species divided by the mean fluorescence signal of the 2C peak of the stained nuclei of the reference standard times the 1C amount of DNA in the reference standard. Three replicates, each from the same individual of *H. vespertilio* and *H.*

euphorbiae, were measured on three different days to minimize possible random instrumental errors. Genome size was reported as 1C, the mean amount of DNA in Mb in a haploid nucleus.

In addition, genome size was estimated by mapping coverage using ModEst [64]. Estimates were calculated using backmap.pl 0.5 (https://github.com/schellt/backmap), in combination with bwa mem 0.7.17 [65], minimap 2.24 [66], samtools 1.15 [67], qualimap 2.2.1 [68], bedtools 2.30.0 [69], R 4.0.3 [70] and multiqc 1.12 [71]. Briefly, the reads used for assembly were mapped back to the assembly itself. Then, the number of mapped nucleotides was divided by the mode of the mapping coverage distribution.

PacBio genome DNA and sequencing

Head tissue (38 mg) from a single male *Hyles euphorbiae* specimen (the same single individual used for genome size estimation) was used for high molecular weight DNA extraction using an adaptation of the protocol of Miller et al. [72]. Final purity and concentrations of DNA were measured using the NanoPhotometer (Implen GmbH, Munich, Germany) and the Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA). One SMRTbell library was constructed following the instructions of the SMRTbell Express Prep kit v2.0 with Low DNA Input Protocol (Pacific Biosciences, Menlo Park, CA). The total input DNA for the library was 3 µg. The library was loaded at an on-plate concentration of 80 pM using diffusion loading. One SMRT cell sequencing run was performed on the Sequel System II in CCS mode using 30-hour movie time with 2 hours of pre-extension and sequencing chemistry V2.0.

Genome assembly of *Hyles euphorbiae*

We created PacBio CCS reads (rq > 0.99) from the *Hyles euphorbiae* subreads.bam file using PacBio's ccs command line tool (version 6.3.0). We obtained 7.9 Gb of high-quality CCS reads (HiFi reads) with an N50 of 11.74 kb. To further increase the read coverage, we applied the DeepConsensus tool (v0.2 with default settings) [73] and obtained a total yield of 8.8 Gb (N50: 11.83 kb). We ran HiFiasm (version 0. 16.1-r375) [23] to create the contig assembly. Remaining haplotypic duplications in the primary contig set were removed using purge-dups (v.1.2.3) [74]. The assembly was scaffolded with HiC data using yahs (v 1.1a) and manually curated with higlass. Remaining gaps in the scaffolds were filled by mapping the raw PacBio subreads with pbmm2 (version 1.7.0), and for alignment piles that fully spanned the gap regions with 1000 anchor bases on both sides, a consensus sequence was produced with gcpp (version 2.0.2). The consensus sequence was used to fill a gap only if: 1) the complete consensus sequence was covered at least 5x; and 2) the coverage profile of the closed gaps fully supported the consensus sequence (i.e. no alignment breaks or huge repeat alignment piles occurred). Two rounds of error correction were performed by aligning the DeepConsensus reads to the assembly using pbmm2, calling variants using DeepVariant (version 1.3.0) [75] and correcting homozygous errors using bcftools consensus. The assembly was checked for contamination using blobtoolkit (version 1.1) and an in-house pipeline that screens several blast databases. Assessment of completeness regarding single copy orthologs was performed via BUSCO 4.1.4 [76] together with the lepidoptera_odb10 set and the options "`--long --offline`". BUSCO (version 5.2.2) [76] scores and merqury

(version 1.3) [77]. QV values were generated for the final scaffolds (*H. euphorbiae* QV=58.1).

The mitochondrial genome was created using the mitoHifi pipeline (version 2) [78] based on CCS reads and the closely related mitochondrial reference genome of *Theretra oldenlandiae* (NCBI accession: MN885801.1).

Hi-C sequence data

Dovetail Hi-C libraries for *H. vespertilio* and *H. euphorbiae* were prepared from head tissues (52.2 mg and 40.6 mg; the same single individuals used for PacBio genome sequencing and genome size estimation) using the Dovetail Hi-C kit (Dovetail Genomics, Scotts Valley, CA, USA) following the manufacturer's protocol version 1.4 for insect samples. Briefly, chromatin was fixed with formaldehyde and then extracted. The fixed chromatin was digested with DpnII, the 5' overhangs were filled with biotinylated nucleotides and the free blunt ends were ligated. After ligation, the crosslinks were reversed, the associated proteins were degraded and the DNA was purified. DNA was then sheared to a mean fragment size of ~350 bp and sequencing libraries were generated using Illumina-compatible adapters. Biotinylated fragments were captured with streptavidin beads before PCR amplification.

The Hi-C libraries were sequenced on a NovaSeq 6000 platform at Novogene (UK), generating 100 million 2 × 150 bp paired-end reads each with a total volume of 30 Gb. The fragment size distribution and concentration of the final PacBio and Dovetail Hi-C libraries were assessed using the TapeStation (Agilent Technologies) and the Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA), respectively.

Scaffolding the assembly of *H. vespertilio* with HiRise

The *H. vespertilio* input assembly of Pippel et al. [11] and the *H. vespertilio* Dovetail Hi-C library reads were used as input data for HiRise, a software pipeline specifically designed for using proximity ligation data to scaffold genome assemblies [27]. The Dovetail Hi-C library sequences were aligned to the draft input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). The separations of the Hi-C read pairs mapped within draft scaffolds were analyzed using HiRise version 2.1.7 to produce a likelihood model for the genomic distance between the read pairs. The model was used to identify and break putative misjoins, score prospective joins, and make joins above a threshold.

Annotation and methylation profile

Structural annotation of protein coding genes was performed using TOGA [79], a method that uses pairwise genome alignment chains between an annotated reference genome (here *Manduca sexta*) and other query species (here *Hyles euphorbiae* and *H. vespertilio*). Briefly, TOGA uses machine learning to infer orthologous loci for each reference transcript, using the concept that orthologous genes display more alignments between intronic and flanking intergenic regions [79]. TOGA then projects each reference transcript to its orthologous query locus using CESAR 2.0 [80], a hidden Markov model method that takes reading frame and splice site annotation of the reference exons into account. CESAR avoids spurious frameshifts and is able to detect evolutionary splice site shifts and precise intron deletions [80, 81]. Using the CESAR alignment, TOGA determines whether the

transcript has inactivating mutations (frameshifting mutations, premature stop codons, splice site disrupting mutations, deletions of entire coding exons).

The mitochondrial genome of *H. euphorbiae* was annotated using the MITOS WebServer [82] and the result illustrated using shinyCircos [83]. This software was also used for the CIRCOS-Plots of the aligned genomes.

The assemblies of *H. euphorbiae* and *H. vespertilio* were aligned to those of the model species *Manduca sexta* (GCF_014839805.1) [24] and *Bombyx mori* (GCF_014905235.1) [28]. To compare contiguity between *H. euphorbiae*, *H. vespertilio*, *M. sexta* and *B. mori*, Quast 5.0.2 [84] was used.
Genomes were aligned using LASTZ 1.04.03 [85] with the parameters (K = 2400, L = 3000, Y = 9400, H = 2000 and the lastz default scoring matrix). Then, we used axtChain [31] (default parameters except linearGap=loose) to compute co-linear alignment chains, RepeatFiller [86] (default parameters) to capture previously missed alignments between repetitive regions and chainCleaner [87] (default parameters except minBrokenChainScore=75000 and -doPairs) to improve alignment specificity. *H. vespertilio* was used as reference and *H. euphorbiae*, *B. mori* and *M. sexta* as queries.

The *Hyles euphorbiae* genome assembly (with annotations and probabilities of site modification/methylation, see below) are accessible in the Senckenberg Genome Browser (https://genome.senckenberg.de/cgi-bin/hgTracks?db=HLhylEup1), as is this genome in an alignment with *Bombyx mori*, *Manduca sexta* and *H. vespertilio* (https://genome.senckenberg.de/cgi-bin/hgTracks?db=HLhylVes2). The genome alignment of *H. vespertilio* to *B. mori* was used to postulate chromosome homologies

and name chromosomes accordingly. Detailed values of the proportions of homologous regions per chromosome are provided in Table S3 (supplementary file).

Table S3 (separate spreadsheet): Chromosome proportion values of the *Bombyx mori – Hyles vespertilio* alignment.

The genomes of *Mimas tiliae* [13], *Laothoe populi* [14], *Hemaris fuciformis* [25] and *Deilephila porcellus* [26] were included for assembly statistics comparisons and a phylogenetic comparison of repeat evolution.

Comparative annotation of repeats for *H. vespertilio*, *H. euphorbiae*, *D. porcellus, Mimas tiliae, L. populi, H. fuciformis,* and *Manduca sexta* was performed using RepeatModeler 2.0.3 [88] and RepeatMasker 4.1.2-p1 [89] as provided in Dfam TETools (https://github.com/Dfam-consortium/TETools), in combination with rmblastn 2.11.0+. The repeat library for RepeatMasker contains a combination of all repeat families identified with RepeatModeler, combined and made non-redundant using Cdhit [90] and manual curation of models according to guidelines set up in [91].

For the methylation profile, the *H. euphorbiae* PacBio HiFi reads containing 5mC base modification tags were converted to CpG/5mC data using pb-CpG-tools (https://github.com/PacificBiosciences/pb-CpG-tools) and uploaded to the genome browser (s.a.). Sites with a modification probability of <50% are marked blue and >50% red.

Chromosome structure analysis

To compare chromosome structure across species, single copy orthologs were inferred using BUCSOs (v5.4) ('lepidoptera odb 10' dataset, metaeuk mode) [92]. Synteny between pairs of genomes was then visualized using oxford plots which were made using custom code (https://github.com/charlottewright/hyles_manuscript) and filtering for all chromosomes with three or more single copy orthologs.

Wing pattern genes

The positions of the wing pattern genes *optix, WntA* and *cortex* (see Table S1 for accession numbers of references) were identified in the two *Hyles* genomes using the BLAT tool [31, 32] with default options as implemented in the Senckenberg Genome browser. The Blat results are presented sorted by alignment length, with the longest selected for each gene.

For the *cortex* gene, we additionally compared the *Hyles* data with the sequences of *Biston betularia* (Geometridae; KT182637), in which the common pale (*typica*) form was replaced by a previously unknown black (*carbonaria*) form during the Industrial Revolution, driven by the interaction between bird predation and smoke pollution [93] and caused by a transposon insertion in a *cortex* intron [22, 94].

Declarations

Ethics approval and consent to participate

The individual of the protected species *H. euphorbiae* was collected with an exception issued in a permit by “Landratsamt Meißen, Kreisumweltamt, Untere Naturschutzbehörde” (AZ 364.621-377/2009-4026/2017-45577/2021) for the purpose of scientific study.

Consent for publication: not applicable

Availability of data and materials

The data generated and analyzed during the current study are available in the NCBI repository, [https://www.ncbi.nlm.nih.gov/genome/?term=hyles].

Competing interests: The authors declare that they have no competing interests.

Funding

This study was supported by grants from the German Research Foundation (DFG) in the framework of the priority program SPP 1991: Taxon-OMICS (HU 1561/5-2). MP was partially funded by the BMBF (grant 01IS18026C). AY and FM acknowledge support from grant 20-13784S of the Czech Science Foundation. CJW was funded by the Wellcome Trust (grants 206194 and 218328). HiRise scaffolding service by Dovetail Genomics for *H. vespertilio* was funded by LOEWE-TBG.

Authors' contributions

AKH conceived the study and partnerships, did field work, performed analyses, produced and edited figures, and wrote the text. TS performed assembly, scaffolding, annotation, and wrote text. FP, CW, LP, AY & FM proposed and performed analyses, produced figures, and wrote text. HD contributed field and laboratory work and SW &

CG took responsibility for DNA isolation, sequencing, and contributed text. MH provided bioinformatics infrastructure and supervision, scripts, intermediate results, advice, and performed alignments, MP performed assembly, scaffolding, numerous analyses and provided the basis for a figure. All authors read and approved the final manuscript.

Acknowledgements

This study benefitted from the sharing of expertise within the DFG priority program SPP 1991 Taxon-Omics. Dovetail Genomics provided HiRise scaffolding service. Thanks go to Alexander Ben Hamadou (LOEWE-TBG) for DNA isolation and library preparations. We thank Ian J. Kitching for language correction and helpful suggestions.

References

1. Hundsdoerfer AK, Päckert M, Kehlmaier C, Strutzenberger P, Kitching IJ: **Museum archives revisited: Central Asiatic hawkmoths reveal exceptionally high late Pliocene species diversification (Lepidoptera, Sphingidae)**. *Zoologica Scripta* 2017, **46**(5):552-570.
2. Hundsdoerfer AK, Kitching IJ, Wink M: **A molecular phylogeny of the hawkmoth genus *Hyles* (Lepidoptera: Sphingidae, Macroglossinae).** *Molecular Phylogenetics and Evolution* 2005, **35**:442-458.
3. Hundsdoerfer AK, Lee KM, Kitching IJ, Mutanen M: **Genome-wide SNP data reveal an overestimation of species diversity in a group of hawkmoths**. *Genome Biology and Evolution* 2019, **11**(8):2136–2150.
4. Hundsdoerfer AK, Tshibangu JN, Wetterauer B, Wink M: **Sequestration of phorbol esters by aposematic larvae of *Hyles euphorbiae* (Lepidoptera: Sphingidae)?** *Chemoecology* 2005, **15**(4):261-267.
5. Hundsdoerfer AK, Rubinoff D, Attié M, Kitching IJ, Wink M: **A revised molecular phylogeny of the globally distributed hawkmoth genus *Hyles* (Lepidoptera: Sphingidae), based on mitochondrial and nuclear DNA sequences**. *Molecular Phylogenetics and Evolution* 2009, **52**:852–865.

6. Hundsdoerfer AK, Buchwalder K, O'Neill MA, Dobler S: **Chemical ecology traits in an adaptive radiation: TPA-sensitivity and detoxification in *Hyles* and *Hippotion* (Sphingidae, Lepidoptera) larvae**. *Chemoecology* 2019, **29**(1):35-47.
7. Mende MB, Bartel M, Hundsdoerfer AK: **A comprehensive phylogeography of the *Hyles euphorbiae* complex (Lepidoptera: Sphingidae) indicates a 'glacial refuge belt'**. *Scientific Reports* 2016, **6**:29527
8. Hundsdoerfer AK, Kitching IJ: **Morphological evolution in *Hyles* Hübner, 1819 hawkmoths (Lepidoptera, Sphingidae): reconstructing the ancestral *Hyles* habitus**. *Nota Lepidopterologica* 2020, **43**:181.
9. Danner F, Eitschberger U, Surholt B: **Die Schwärmer der westlichen Palaearktis. Bausteine zu einer Revision (Lepidoptera: Sphingidae).** *Herbipoliana* 1998, **4**:1-368 (Textband), 361-720 (Tafelband).
10. Hundsdoerfer AK, Kitching IJ: **Ancient incomplete lineage sorting of *Hyles* and *Rhodafra* (Lepidoptera: Sphingidae)**. *Organisms Diversity & Evolution* 2020, **20**(3):527-536.
11. Pippel M, Jebb D, Patzold F, Winkler S, Vogel H, Myers G, Hiller M, Hundsdoerfer AK: **A highly contiguous genome assembly of the bat hawkmoth *Hyles vespertilio* (Lepidoptera: Sphingidae)**. *GigaScience* 2020, **9**(1):giaa001.
12. Patzold F, Zilli A, Hundsdoerfer AK: **Advantages of an easy-to-use DNA extraction method for minimal-destructive analysis of collection specimens**. *PLoS One* 2020, **15**(7):e0235222.
13. Boyes D, Holland P, University of Oxford and Wytham Woods Genome Acquisition Lab L, al. e: **The genome sequence of the lime hawk-moth, *Mimas tiliae* (Linnaeus, 1758)**. *Wellcome Open Research* 2021, **6**.
14. Boyes D, Holland PW, Consortium DToL: **The genome sequence of the poplar hawk-moth, *Laothoe populi* (Linnaeus, 1758).** *Wellcome Open Research* 2021 [version 1 peer review: 1 approved, 1 approved with reservations], **6**:237.
15. Van Belleghem SM, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA, Hanly JJ, Mallet J, Lewis JJ, Hines HM *et al*: **Complex modular architecture around a simple toolkit of wing pattern genes**. *Nat Ecol Evol* 2017, **1**(3):52.
16. Martin A, Reed RD: **Wingless and aristaless2 define a developmental ground plan for moth and butterfly wing pattern evolution**. *Molecular Biology and Evolution* 2010, **27**(12):2864-2878.
17. Stevens M, Merilaita S: **Animal camouflage: current issues and new perspectives**. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2009, **364**(1516):423-427.
18. Zhang L, Mazo-Vargas A, Reed RD: **Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence**. *Proc Natl Acad Sci U S A* 2017, **114**(40):10707-10712.
19. Livraghi L, Martin A, Gibbs M, Braak N, Arif S, Breuker CJ: **CRISPR/Cas9 as the key to unlocking the secrets of butterfly wing pattern development and its evolution**. In: *Advances in Insect Physiology*. vol. 54: Elsevier; 2018: 85-115.
20. Jiggins CD, Wallbank RW, Hanly JJ: **Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns?** *Philos Trans R Soc Lond B Biol Sci* 2017, **372**(1713).
21. Nadeau NJ, Pardo-Diaz C, Whibley A, Supple MA, Saenko SV, Wallbank RW, Wu GC, Maroja L, Ferguson L, Hanly JJ *et al*: **The gene *cortex* controls mimicry and crypsis in butterflies and moths.** *Nature* 2016, **534**(7605):106-110.
22. van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ: **The industrial melanism mutation in British peppered moths is a transposable element**. *Nature* 2016, **534**(7605):102-105.

23. Cheng H, Concepcion GT, Feng X, Zhang H, Li H: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm**. *Nature methods* 2021, **18**(2):170-175.
24. Gershman A, Romer TG, Fan Y, Razaghi R, Smith WA, Timp W: **De novo genome assembly of the tobacco hornworm moth (*Manduca sexta*)**. *G3* 2021, **11**(1):1-9.
25. Mulhair PO, Holland PW: **Evolution of the insect Hox gene cluster: Comparative analysis across 243 species**. *Seminars in Cell & Developmental Biology* 2022, https://doi.org/10.1016/j.semcdb.2022.11.010.
26. Boyes D, University of Oxford and Wytham Woods Genome Acquisition Lab L, Darwin Tree of Life Barcoding collective c, al. e: **The genome sequence of the small elephant hawk moth, *Deilephila porcellus* (Linnaeus, 1758)** *Wellcome Open Research* 2022, **7**:80.
27. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW: **Chromosome-scale shotgun assembly using an in vitro method for long-range linkage**. *Genome Research* 2016, **26**(3):342-350.
28. Kawamoto M, Jouraku A, Toyoda A, Yokoi K, Minakuchi Y, Katsuma S, Fujiyama A, Kiuchi T, Yamamoto K, Shimada T: **High-quality genome assembly of the silkworm, *Bombyx mori*.** *Insect Biochemistry and Molecular Biology* 2019, **107**:53-62.
29. Yasukochi Y, Tanaka-Okuyama M, Shibata F, Yoshido A, Marec F, Wu C, Zhang H, Goldsmith MR, Sahara K: **Extensive conserved synteny of genes between the karyotypes of *Manduca sexta* and *Bombyx mori* revealed by BAC-FISH mapping**. *PLoS One* 2009, **4**(10):e7465.
30. Kawahara AY, Barber JR: **Tempo and mode of antibat ultrasound production and sonar jamming in the diverse hawkmoth radiation**. *Proceedings of the National Academy of Sciences* 2015, **112**(20):6407-6412.
31. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D: **Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(20):11484-11489.
32. Lee BT, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM *et al*: **The UCSC Genome Browser database: 2022 update**. *Nucleic Acids Res* 2022, **50**(D1):D1115-D1122.
33. napari, contributors: **napari: a multi-dimensional image viewer for python**. In.; 2019.
34. Hundsdoerfer AK, Kitching IJ, Wink M: **The phylogeny of the *Hyles euphorbiae*-complex (Lepidoptera: Sphingidae): molecular evidence from sequence data and ISSR-PCR fingerprints.** *Organisms Diversity and Evolution* 2005, **5**:173–198.
35. Hanrahan SJ, Johnston JS: **New genome size estimates of 134 species of arthropods**. *Chromosome Research* 2011, **19**(6):809-823.
36. d'Alencon E, Sezutsu H, Legeai F, Permal E, Bernard-Samain S, Gimenez S, Gagneur C, Cousserans F, Shimomura M, Brun-Barale A *et al*: **Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements**. *Proc Natl Acad Sci U S A* 2010, **107**(17):7680-7685.
37. Lavoie CA, Platt RN, Novick PA, Counterman BA, Ray DA: **Transposable element evolution in *Heliconius* suggests genome diversity within Lepidoptera**. *Mob DNA* 2013, **4**(1):21.
38. Lynch M, Walsh B: **The origins of genome architecture** vol. 98. Sunderland, MA: Sinauer Associates; 2007.
39. Shah A, Hoffman JI, Schielzeth H: **Comparative Analysis of Genomic Repeat Content in Gomphocerine Grasshoppers Reveals Expansion of Satellite DNA and**

Helitrons in Species with Unusually Large Genomes. *Genome Biol Evol* 2020, **12**(7):1180-1193.

40. Hollister JD, Gaut BS: **Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression**. *Genome Research* 2009, **19**(8):1419-1428.

41. Charlesworth B, Sniegowski P, Stephan W: **The evolutionary dynamics of repetitive DNA in eukaryotes**. *Nature* 1994, **371**(6494):215-220.

42. Talla V, Suh A, Kalsoom F, Dinca V, Vila R, Friberg M, Wiklund C, Backstrom N: **Rapid Increase in Genome Size as a Consequence of Transposable Element Hyperactivity in Wood-White (Leptidea) Butterflies**. *Genome Biol Evol* 2017, **9**(10):2491-2505.

43. Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Kawakami T, Kunstner A, Makinen H, Nadachowska-Brzyska K, Qvarnstrom A *et al*: **The genomic landscape of species divergence in *Ficedula* flycatchers**. *Nature* 2012, **491**(7426):756-760.

44. Feliciello I, Akrap I, Brajkovic J, Zlatar I, Ugarkovic D: **Satellite DNA as a driver of population divergence in the red flour beetle *Tribolium castaneum***. *Genome Biol Evol* 2014, **7**(1):228-239.

45. Maumus F, Fiston-Lavier AS, Quesneville H: **Impact of transposable elements on insect genomes and biology**. *Curr Opin Insect Sci* 2015, **7**:30-36.

46. Pittaway AR: **The hawkmoths of the Western Palaearctic**. Colchester: Harley Books; 1993.

47. Brower AVZ: **Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution**. *Proceedings of the National Academy of Science of the United States of America* 1994, **91**:6491-6495.

48. Ahola V LR, Somervuo P, Salmela L, Koskinen P, Rastas P et al.: **The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera**. *Nature Communications* 2014(5):4737.

49. Jiggins CD: **What can we learn about adaptation from the wing pattern genetics of *Heliconius* butterflies?** In: *Diversity and Evolution of Butterfly Wing Patterns*. Springer, Singapore; 2017: 173-188.

50. Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Chen R *et al*: **optix drives the repeated convergent evolution of butterfly wing pattern mimicry**. *Science* 2011, **333**(6046):1137-1141.

51. Sahara K, Yoshido A, Shibata F, Fujikawa-Kojima N, Okabe T, Tanaka-Okuyama M, Yasukochi Y: **FISH identification of *Helicoverpa armigera* and *Mamestra brassicae* chromosomes by BAC and fosmid probes**. *Insect Biochem Mol Biol* 2013, **43**(8):644-653.

52. van't Hof AE, Nguyen P, Dalikova M, Edmonds N, Marec F, Saccheri IJ: **Linkage map of the peppered moth, Biston betularia (Lepidoptera, Geometridae): a model of industrial melanism**. *Heredity (Edinb)* 2013, **110**(3):283-295.

53. Lukhtanov V: **Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta)**. *Journal of Zoological Systematics and Evolutionary Research* 2000(38):73–79.

54. Davey JW, Barker SL, Rastas PM, Pinharanda A, Martin SH, Durbin R, McMillan WO, Merrill RM, Jiggins CD: **No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions**. *Evol Lett* 2017, **1**(3):138-154.

55. Jiggins CD, Mavarez J, Beltran M, McMillan WO, Johnston JS, Bermingham E: **A genetic linkage map of the mimetic butterfly Heliconius melpomene**. *Genetics* 2005, **171**(2):557-570.

56. Pringle EG, Baxter SW, Webster CL, Papanicolaou A, Lee SF, Jiggins CD: **Synteny and chromosome evolution in the lepidoptera: evidence from mapping in *Heliconius melpomene***. *Genetics* 2007, **177**(1):417-426.
57. Yoshido A, Bando H, Yasukochi Y, Sahara K: **The *Bombyx mori* karyotype and the assignment of linkage group.** . *Genetics* 2005, **170**:675-685.
58. Maddison WP: **Gene trees in species trees**. *Systematic Biology* 1997, **46**(3):523-536.
59. Yoshido A, Sahara K, Yasukochi Y, Sharakhov I: **Silk moths (Lepidoptera)**. In: *Protocols for Cytogenetic Mapping of Arthropod Genomes* Edited by Sharakhov IV. Boca Ranton, FL, USA: CRC Press; 2015: 219-256.
60. 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, **114**(3):193-202.
61. Sahara K, Marec F, Traut W: **TTAGG telomeric repeats in chromosomes of some insects and other arthropods**. *Chromosome Research* 1999, **7**(6):449-460.
62. Team TGD: **GIMP. Retrieved from https://www.gimp.org**. In.; 2019.
63. Hare EE, Johnston JS: **Genome size determination using flow cytometry of propidium iodide-stained nuclei**. In: *Molecular methods for evolutionary genetics*. Springer; 2012: 3-12.
64. Pfenninger M, Schönnenbeck P, Schell T: **ModEst: Accurate estimation of genome size from next generation sequencing data**. *Molecular Ecology Resources* 2021.
65. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**. *arXiv preprint arXiv:13033997* 2013.
66. Li H: **Minimap2: pairwise alignment for nucleotide sequences**. *Bioinformatics* 2018, **34**(18):3094-3100.
67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
68. Okonechnikov K, Conesa A, Garcia-Alcalde F: **Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data**. *Bioinformatics* 2016, **32**(2):292-294.
69. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics* 2010, **26**(6):841-842.
70. R Core Team R: **A Language and Environment for Statistical Computing**. 2021.
71. Ewels P, Magnusson M, Lundin S, Kaller M: **MultiQC: summarize analysis results for multiple tools and samples in a single report**. *Bioinformatics* 2016, **32**(19):3047-3048.
72. Miller S, Dykes D, Polesky H: **A simple salting out procedure for extracting DNA from human nucleated cells**. *Nucleic Acids Research* 1988, **16**(3):1215.
73. Baid G, Cook DE, Shafin K, Yun T, Llinares-Lopez F, Berthet Q, Wenger AM, Rowell WJ, Nattestad M, Yang H: **DeepConsensus: Gap-Aware Sequence Transformers for Sequence Correction**. *bioRxiv* 2021.
74. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R: **Identifying and removing haplotypic duplication in primary genome assemblies**. *Bioinformatics* 2020, **36**(9):2896-2898.
75. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT: **A universal SNP and small-indel variant caller using deep neural networks**. *Nature Biotechnology* 2018, **36**(10):983-987.
76. Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage**

for scoring of eukaryotic, prokaryotic, and viral genomes. *arXiv preprint arXiv:210611799* 2021.

77. Rhie A, Walenz BP, Koren S, Phillippy AM: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies**. *Genome Biology* 2020, **21**(1):1-27.

78. Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F: **MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. https://github.com/marcelauliano/MitoHiFi**. *Mol Ecol Resour* 2020, **20**(4):892-905.

79. Kirilenko B, Munegowda C, Osipova E, Jebb D, Sharma V, Blumer M, Morales A, Ahmed A, Kontopoulos D, Hilgers L *et al*: **TOGA integrates gene annotation with orthology inference at scale.** under review.

80. Sharma V, Schwede P, Hiller M: **CESAR 2.0 substantially improves speed and accuracy of comparative gene annotation**. *Bioinformatics* 2017, **33**(24):3985-3987.

81. Sharma V, Elghafari A, Hiller M: **Coding exon-structure aware realigner (CESAR) utilizes genome alignments for accurate comparative gene annotation**. *Nucleic Acids Res* 2016, **44**(11):e103.

82. Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsch G, Putz J, Middendorf M, Stadler PF: **MITOS: improved de novo metazoan mitochondrial genome annotation**. *Mol Phylogenet Evol* 2013, **69**(2):313-319.

83. Yu Y, Ouyang Y, Yao W: **shinyCircos: an R/Shiny application for interactive creation of Circos plot**. *Bioinformatics* 2018, **34**(7):1229-1231.

84. Gurevich A, Saveliev V, Vyahhi N, Tesler G: **QUAST: quality assessment tool for genome assemblies**. *Bioinformatics* 2013, **29**(8):1072-1075.

85. Harris RS: **Improved pairwise alignment of genomic DNA. A Thesis in Computer Science and Engineering.**: The Pennsylvania State University; 2007.

86. Osipova E, Hecker N, Hiller M: **RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements**. *bioRxiv* 2019:696922.

87. Suarez HG, Langer BE, Ladde P, Hiller M: **chainCleaner improves genome alignment specificity and sensitivity**. *Bioinformatics* 2017, **33**(11):1596-1603.

88. Smit A, Hubley R, Green P: **RepeatModeler Open-1.0. 2008–2015**. *Institute for Systems Biology, Seattle, USA Available from: http://wwwrepeatmaskerorg* 2015.

89. Smit A, Hubley R, Green P: **RepeatMasker Open-4.0. 2013–2015. Available from: http://www.repeatmasker.org**. In.; 2015.

90. Fu L, Niu B, Zhu Z, Wu S, Li W: **CD-HIT: accelerated for clustering the next-generation sequencing data**. *Bioinformatics* 2012, **28**:3150–3152.

91. Goubert C, Craig RJ, Bilat AF, Peona V, Vogan AA, Protasio AV: **A beginner's guide to manual curation of transposable elements. Mobile DNA**. *Mobile DNA* 2022, **13**:1-19.

92. Simao Neto F, Waterhouse R, Ioannidis P, Kriventseva E, Zdobnov E: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics* 2015, **31**:3210-3212.

93. Cook LM: **The rise and fall of the Carbonaria form of the peppered moth**. *Q Rev Biol* 2003, **78**(4):399-417.

94. van't Hof AE, Edmonds N, Dalíková M, Marec F, Saccheri IJ: **Industrial melanism in British peppered moths has a singular and recent mutational origin**. *Science* 2011, **332**(6032):958-960.