High quality genomes corroborate 29 chromosomes of the haploid *Hyles* (Lepidoptera: Sphingidae) karyotype

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

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

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

[5] 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

FISH analysis of the karyotype revealed n = 29 chromosomes in *Hyles euphorbiae*. The measured genome sizes of *H. euphorbiae* and *H. vespertilio* are estimated to have average 1C DNA values of 472 and 562 Mb respectively. The *H. euphorbiae* genome was PacBio sequenced and amended by Hi-C Illumina data yielding a 504

Mb assembly with a scaffold N50 of 18.2 Mb and 99.9% of the data being represented by the 29 largest scaffolds, corroborating the haploid karyotype. Chromosome length estimations based on karyotype image data provide an additional quality metric of the assembled chromosome sizes. Hi-C data was also used for chromosome-level scaffolding of the published *H. vespertilio* genome, leading to a second assembly (651 Mb) with scaffold N50 of 22 Mb, 98% in the 29 largest scaffolds representing the chromosomes. The larger *H. vespertilio* genome size was accompanied by a proportional increase of repeats from 45% in *H. euphorbiae* to nearly 55% in *H. vespertilio*.

In both *Hyles* species, the three wing pattern genes, *optix, wingless/wint-1* and *cortex*, were found on chromosomes 23, 4 and 17, respectively. Peaks of divergence surrounding *wingless/wnt-1* and *cortex* provide candidate genomic areas in which wing patterns are determined in this genus.

Introduction

The spurge hawkmoth *Hyles euphorbiae* Linnaeus 1758 is a charismatic Palearctic species of the genus *Hyles* (family Sphingidae) with large, colorful, aposematic and polymorphic larvae and camouflaged, heavy moths with strong flight abilities. Surprisingly, its larvae do not sequester the toxic spurge diterpene esters [1] and monophagy of larvae on toxic *Euphorbia* host plants has evolved twice independently within the genus [2, 3]. The impressively high morphological variability of larvae has complicated its taxonomy by contributing to an overestimation of species diversity (overview in Hundsdoerfer et al. [4]). Similarly, high intraspecific mitochondrial marker gene diversity bedeviled reconstruction of the molecular phylogeny of the former five

species [2], but provided valuable resolution for phylogeography [5]. Whereas some larval patterns are correlated to geography [5] and are thus expected to be based on underlying genetic diversity causing phenotypic variability, others appear to be environmentally determined (phenotypic plasticity).

Early studies have already demonstrated that wing pattern similarity does not correlate with phylogenetic relatedness in this genus [6, 7]. Seven basic wing patterns are observed in the Central Palearctic *Hyles* species and they do not correlate with species as currently defined and also do not reflect the phylogenetic relationships within the genus. Recently, we standardized forewing patterns for members of the genus, defined morphological characters and coded these for around 200 individuals in a matrix [8]. Group formation in 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 forewing (the former subgenus 'Danneria' according to Danner et al. [9]), but this had to be refuted as a clade by earlier molecular phylogenetic work [6]. Another large clade encompasses moths that largely show a pattern of dark brown spots and stripes on a lighter, cream-colored background. This corresponds to the typical *H. euphorbiae* forewing pattern, and thus the group included all species with a similar wing pattern [8] (or slight variations of it), despite many lacking a close molecular phylogenetic relationship [10] to this species. Another clade consists of species with forewing patterns that lack many or most of these pattern elements described above, including *H. vespertilio*, for which the genome has been published recently [11]. This species' forewings have the appearance of a naturally occurring lack of wing pattern, as if the gene(s) for the wing pattern were naturally knocked-out, which makes a genomic comparison to it particularly intriguing.

Interspecific differences in forewing patterns within the genus *Hyles* should be based on detectable genetic differences, since the patterns are stable within species in the well separated, oldest Neotropical (and Nearctic) taxa [6]. 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]).

Currently, numerous Sphingidae genomes are being published (e.g. Pippel et al. [11]), for which peer-review processes are partly still underway (e.g. *Mimas tiliae*, Smerinthinae [13]). This wealth of data will enable insight into the evolution of wing patterns in hawkmoths. Pioneer studies of genes underlying the Lepidoptera wing pattern in the genus *Heliconius* (family Nymphalidae) [14] have revealed a modular architecture with narrow stretches of the genome associated with specific differences in color and pattern. *Optix* is a single-exon gene on chromosome 18 in *Heliconius* [14, 15] encoding a transcription factor and thus not directly involved in ommochrome pigmentation [16], but is nevertheless associated with red and orange forewing patterning in these butterflies. Nearby non-coding regions control its expression in *Heliconius* wing development [14] and these are the regions of divergence, whereas the coding area is a conserved homeobox gene. *Hyles euphorbiae* forewings are known for their pink flush in some regions of its distribution range (see moths in Fig. 2 of Hundsdoerfer, Lee et al. [4]), whereas red is not prominent in the grey wings of *H. vespertilio*.

The gene *wingless* is necessary for wing and haltere development in *Drosophila melanogaster* [17] and was identified as similar to mouse secretory glycoprotein *int-1*, leading to the new nomenclature in which both genes are referred to as *wnt-1* (*wingless*-type integration site gene family [18]). The *wntA* signalling ligand on

chromosome 10 is highly conserved at the amino-acid level within *Heliconius* [14]) and associated with the forewing black band. Variation is again rather influenced via the expression during wing development and thus found in the nearby control region [14]. Within *Hyles*, *wingless/wnt-1* sequences show variability and have thus been used as a source of characters for phylogenetic inference [19].

Another wing pattern gene, *cortex* (on chromosome 17 and coding for yellow patterns on *Heliconius* wings) [20], has been suggested to regulate pattern switches across Lepidoptera [21]. The same flexible mechanism for rapid morphological diversification is expected to apply to the genus *Hyles*. The insertion of a transposable element (TE) into an intron of the gene *cortex* was shown to give rise to industrial melanism [22], i.e. darker forewings in the Peppered Moth (*Biston betularia*; Geometridae), corroborating the function across Lepidoptera. It had already been 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 [23]. In addition, variation in the colors and certain color patterns of species is primarily expected to be driven by biotic and abiotic factors [24] acting as selection pressure on individuals with differential expression of genotypes, due to, e.g., the *cortex* TE insertion. Indeed, this insertion [22], estimated to have occurred around 1819, could only be genetically manifested in the population via selection pressure against the light-colored moths showing up on the dark trees and thus experiencing higher predation.

In contrast to the genome-sequenced *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 current paper aims at providing a data basis and a directed track for future studies to understand the origin of *Hyles* wing patterns as phenotypic variability coded by the genetic wing pattern modules described in *Heliconius* (e.g. [14, 20, 23]).

Results

Karyotype

Analysis of male mitotic chromosomes stained by FISH with telomeric probe (telomere-FISH) showed that the karyotype of *H. euphorbiae* is composed of $2n = 58$ chromosomes (Fig. 1a). As is 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. The chromosome number was confirmed by analysis of meiotic nuclei in 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. 1b, c). In addition, GISH identified a WZ sex chromosome bivalent in pachytene oocytes by labelling the major portion of the W chromosome with the female gDNA probe (Fig. 1c), whereas no bivalent was identified in pachytene spermatocytes (Fig. 1b). 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. 1c),

suggesting that the W and Z chromosomes are among the largest chromosomes in the *H. euphorbiae* karyotype.

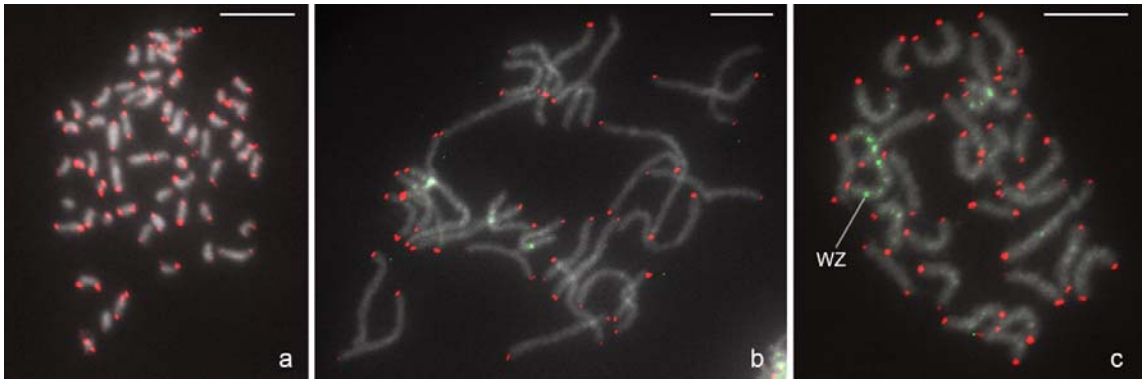


Fig. 1. **Molecular cytogenetic analysis of *Hyles euphorbiae* chromosomes.** Hybridization signals of the Cy3-labelled $(TTAGG)_n$ telomeric probe (red) indicate the chromosomal ends (**a**–**c**), and the fluorescein-labelled female gDNA probe (green) identifies the sex chromosome system (**b** and **c**). Chromosomes were stained with DAPI (grey). (**a**) Male mitotic prometaphase stained by telomere-FISH showing a diploid chromosome number of $2n = 58$. (**b**) Male pachytene complement stained by combination of GISH and telomere-FISH showing 29 bivalents, but without any bivalent highlighted, thus indicating a ZZ sex chromosome constitution. (**c**) Female pachytene complement stained by 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

The chromosome size estimation is based on Fig. 1c, bivalents from a female pupal gonad cell in the pachytene stage. The chromosome size estimates are shown in Fig. 2, they corroborate the WZ bivalent as the largest chromosome. Based on semi-automated image processing, the software package napari-karyotype [25] relies on threshold-based image segmentation to detect chromosome-related components. Identified chromosomal objects are surrounded by red rectangles and labeled with the estimates.

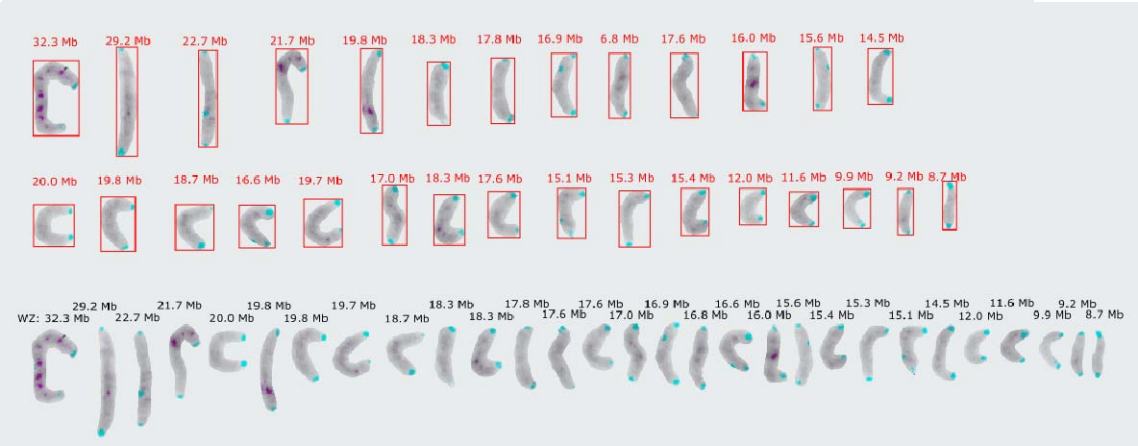


Fig. 2. Annotated chromosomes from a female pachytene *Hyles euphorbiae* karyotype image using the software package napari-karyotype. The top and middle rows show the input image, red containment boxes and red size estimates. The bottom row shows the chromosomes aligned manually according to the size estimated. The largest chromosome represents the sex chromosome, in this case a WZ-bivalent.

Genome size estimation

The 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 estimate for *H. euphorbiae* using ModEst resulted in 535 Mb.

Hyles euphorbiae genome assembly

Assembly contiguity statistics are summarized in Table 1. The final contig assembly consists of 321 contigs, has a N50 of 2.76 Mb and a size of 504 Mb and is available under the NCBI SRA accession SRR17432892, BioProject PRJNA794104,

BioSample SAMN24610150 and genome JALBCW000000000, JALBCX000000000. The final annotated circular mitochondrial genome has length of 15,322 bp.

Final assembly statistics are summarized in Table 2. The assembly of *H. euphorbiae* is almost 124 Mb smaller than that of *H. vespertilio*.

Table 1: Contiguity statistics of different assembly steps from *H. euphorbiae* assembly.

	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: Available statistics of presented and related species. Genome size estimates for *B. mori* and *M. sexta* are taken from the Animal Genome Size Database (Gregory, 2021).

	H. euphorbiae	*H. vespertilio*	*M. sexta*[26]	*B. mori*[27]
#sequences	322	390	4,057	697
Genome size	472,000,000	562,000,000	420,540,000[28]	508,560,000[29]
Total length	504,323,440	651,427,907	470,036,997	460,349,660
Scaffold N50	18,182,747	22,136,963	14,248,853	16,796,068
Contig N50	2,758,341	7,263,332	424,948	12,201,325
Karyotype (haploid)	29	29	28	28
Total length in longest scaffolds according karyotype [%]	99.9	95.3	86.1	96.7
BUSCO (N=5286)				
Complete	98.2	98.3	98.3	98.7
Single copy	97.9	95.4	91.8	97.8
Duplicated	0.3	2.9	6.5	0.9
Fragmented	0.5	0.7	0.7	0.4
Missing	1.3	1.0	1.0	0.9

Chromosome scaffolding *H. vespertilio* Hi-C data

HiRise scaffolding of *H. vespertilio* using more than 132M 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 the chromosomes (Fig. 3). Sizes of chromosomes of both *Hyles* species are of the same order of magnitude, but not identical (Table 3).

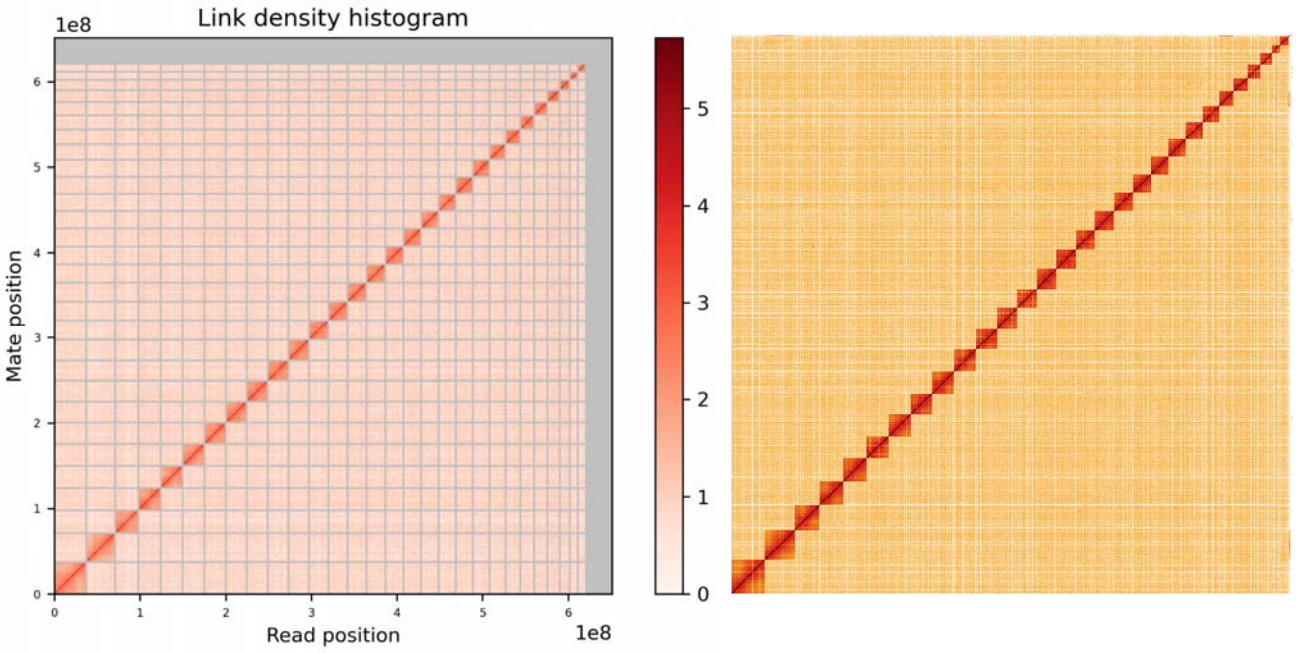


Fig. 3. Contact map of a) *H. vespertilio* and b) *H. euphorbiae*. Chromosome-level scaffolding clearly supports 29 scaffolds representing the chromosomes. Chromosomes are given in order of size from the bottom left to top right.

Table 3: Chromosome sizes based on bioinformatic analyzes of sequence data and chromosome numbering in homology to *B. mori*. Chromosome 1 corresponds to chromosome Z [56]. 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**	**Assembly Length (bp) incl. gaps**	***H. eu-phorbiae*** **Chr. number**	***H. euphorbiae*** **scaffold name in browser**	**Assembly Length (bp) incl. gaps**
BMChrZ	HvChrZ	ScWp86a_195_HRSCAF_280	37,218,281	HeChrZ	sc_1	30,347,856
BmChr2	HvChr2*	ScWp86a_159_HRSCAF_226	24,553,817	HeChr2	sc_7	19,904,380
BmChr3	HvChr3	ScWp86a_257_HRSCAF_395	20,065,448	HeChr3	sc_18	16,855,035
BmChr4	HvChr4	ScWp86a_21_HRSCAF_27	25,450,662	HeChr4	sc_6	20,022,273
BmChr5	HvChr5	ScWp86a_119_HRSCAF_165	26,040,715	HeChr5	sc_4	21,578,258
BmChr6	HvChr6	ScWp86a_136_HRSCAF_194	22,136,963	HeChr6	sc_13	17,936,700
BmChr7	HvChr7	ScWp86a_95_HRSCAF_134	17,182,831	HeChr7	sc_23	14,557,976
BmChr8	HvChr8	ScWp86a_214_HRSCAF_310	21,873,546	HeChr8	sc_14	17,505,000
BmChr9	HvChr9	ScWp86a_232_HRSCAF_355	24,188,504	HeChr9	sc_11	19,134,990
BmChr10	HvChr10	ScWp86a_42_HRSCAF_61	24,738,224	HeChr10	sc_8	19,757,600
BmChr11	HvChr11	ScWp86a_178_HRSCAF_255	22,098,621	HeChr11	sc_15	17,388,153
BmChr12	HvChr12	ScWp86a_182_HRSCAF_262	26,852,440	HeChr12	sc_5	21,432,827
BmChr13	HvChr13	ScWp86a_68_HRSCAF_94	24,145,169	HeChr13	sc_10	19,486,524
BmChr14	HvChr14	ScWp86a_111_HRSCAF_155	20,079,433	HeChr14	sc_21	15,684,570
BmChr15	HvChr15	ScWp86a_62_HRSCAF_86	26,014,704	HeChr15	sc_3	21,768,765
BmChr16	HvChr16	ScWp86a_55_HRSCAF_78	18,679,767	HeChr16	sc_22	15,570,054
BmChr17	HvChr17	ScWp86a_210_HRSCAF_302	24,352,496	HeChr17	sc_9	19,686,965
BmChr18	HvChr18	ScWp86a_1_HRSCAF_1	20,909,092	HeChr18	sc_16	17,373,644
BmChr19	HvChr19	ScWp86a_207_HRSCAF_299	20,353,206	HeChr19	sc_19	16,571,156
BmChr20	HvChr20	ScWp86a_378_HRSCAF_523	16,497,537	HeChr20	sc_24	12,812,799
BmChr21	HvChr21	ScWp86a_162_HRSCAF_230	21,114,399	HeChr21	sc_17	17,181,161
BmChr22	HvChr22#	ScWp86a_157_HRSCAF_223	33,800,775	HeChr22	sc_2	27,093,441
BmChr23	HvChr23	ScWp86a_31_HRSCAF_41	22,933,977	HeChr23	sc_12	18,182,747
BmChr24	HvChr24	ScWp86a_166_HRSCAF_234	11,894,872	HeChr24	sc_27	10,049,873
BmChr25	HvChr25	ScWp86a_7_HRSCAF_7	19,247,757	HeChr25	sc_20	15,917,517
BmChr26	HvChr26**	ScWp86a_81_HRSCAF_111	9,224,630	HeChr26	sc_29	7,206,840
BmChr27	HvChr27	ScWp86a_289_HRSCAF_430	14,304,874	HeChr27	sc_26	12,048,215
BmChr28	HvChr28#	ScWp86a_140_HRSCAF_198	15,210,208	HeChr28	sc_25	12,771,623
n.a.	HvChr29§	ScWp86a_137_HRSCAF_195	9,587,503	HeChr29	sc_28	7,786,824

Annotation

The proportion of the *H. euphorbiae* genome assembly covered by major classes of repetitive elements is 45% in total (Table 4), illustrated by stacked bar charts (Fig. 4). *Hyles vespertilio* thus exhibits nearly 10% more non-repetitive, potentially informative DNA than *H. euphorbiae*. The mitochondrial genome of *H. euphorbiae* (Fig. S1) contains 15 protein coding genes, two ribosomal RNA genes (rRNA) plus 22 transfer RNA sequences (tRNA) and the control region.

Table 4. Assembly lengths and proportion of repeats.

Species	Total length	Masked [%]
H. euphorbiae	504,310,614	47.1
H. vespertilio	651,427,907	53.39
M. sexta	470,036,997	34.13
B. mori	460,349,660	50.99

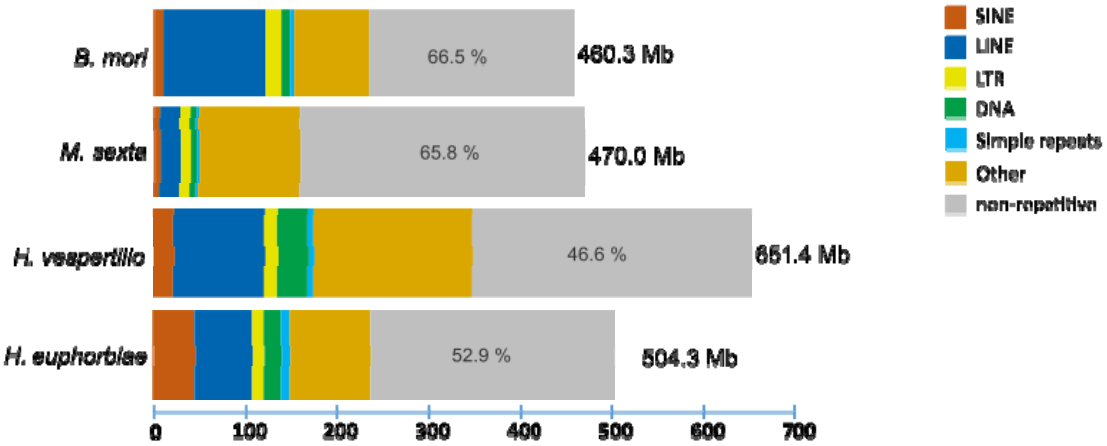


Fig. 4. Proportional repeat contents of *H. euphorbiae*, *H. vespertilio*, *M. sexta* (all Sphingidae) and *B. mori* (Bombycidae). Scale in megabases (Mb).

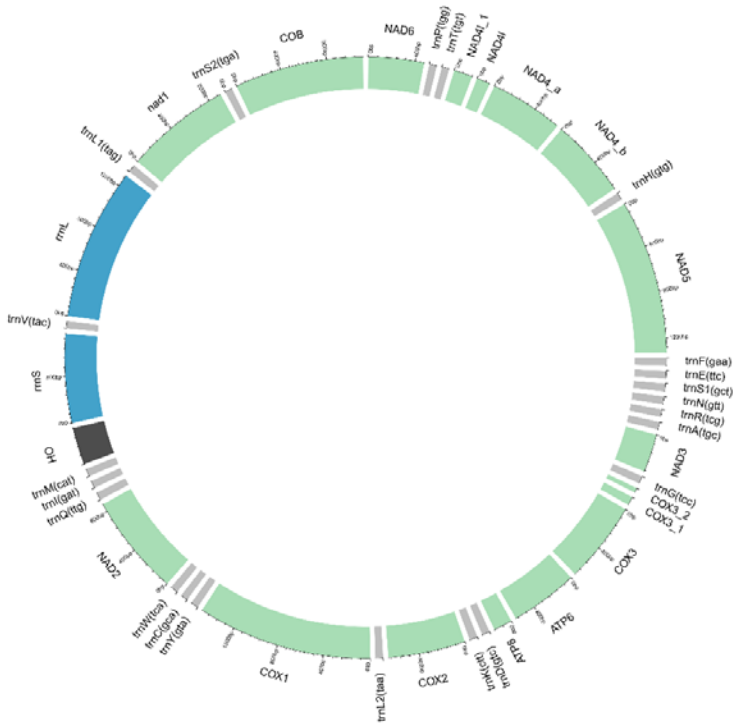


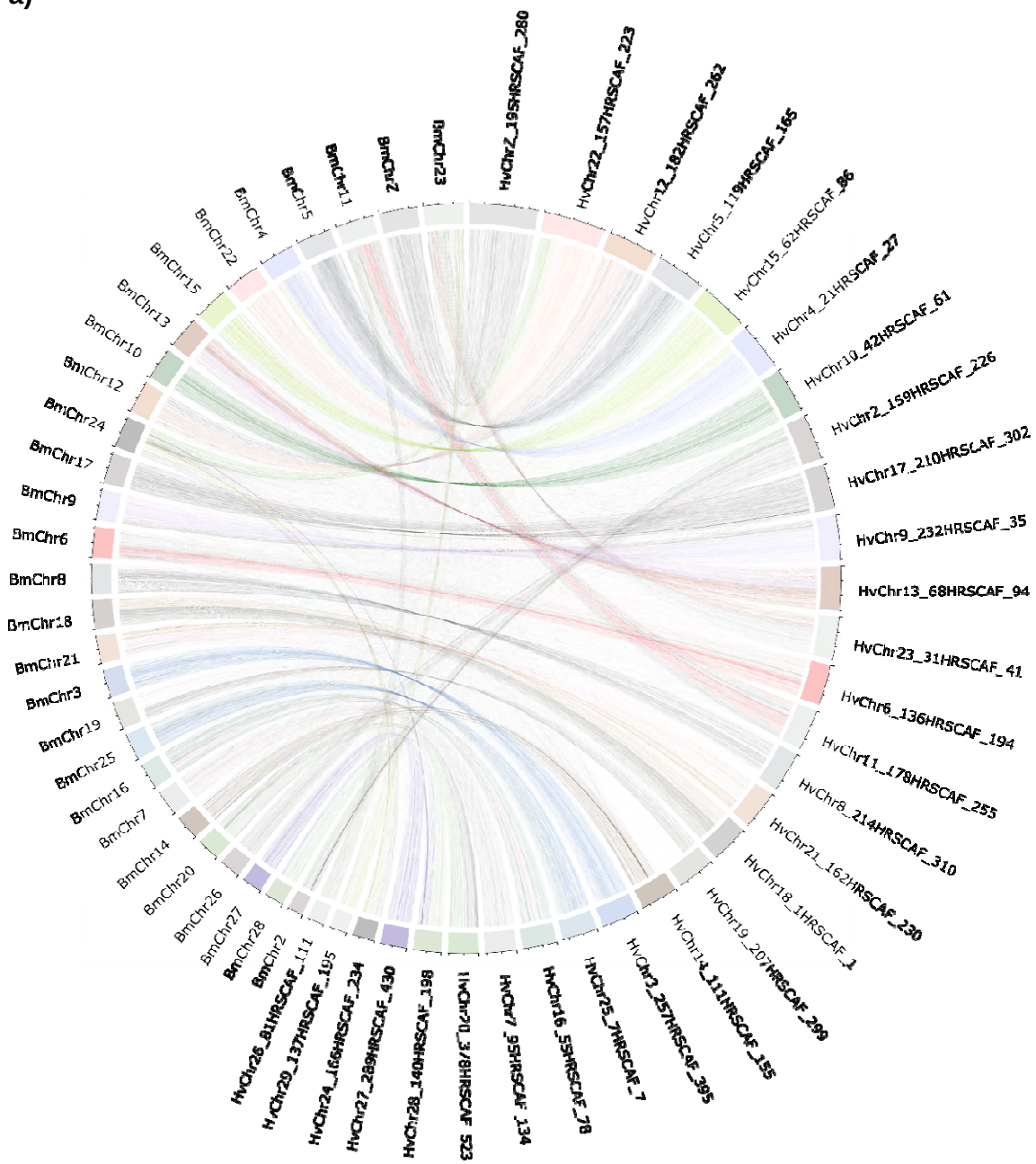
Fig. S1. The annotated mitochondrial genome of *H. euphorbiae*.

The *Hyles* genome alignment and annotations are accessible in the Senckenberg Genome Browser (https://genome.senckenberg.de/cgi-bin/hgTracks?db=HLhylVes2 in an alignment with *B. mori*, *M. sexta* and *H. euphorbiae,* and https://genome.senckenberg.de/cgi-bin/hgTracks?db=HLhylEup1). 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. 5a) to allow comparison by eye. CIRCOS plots of the alignment between the two species showed high chromosome homology, except for some chromosomes that might be involved in chromosome rearrangements. The *H. vespertilio* chromosome 2 (HvChr2) corresponds to chromosomes 2 and 26 in *B. mori* (BmChr2, BmChr26). *Bombyx mori* chromosome 24, BmChr24, is split among *H. vespertilio* chromosomes 22, 24 and 28 (HvChr22, HvChr24, HvChr28) and chromosome 11 of *B. mori* (BmChr11) is split among *H. vespertilio* chromosomes 11 and 26 (HvChr11, HvChr26). These

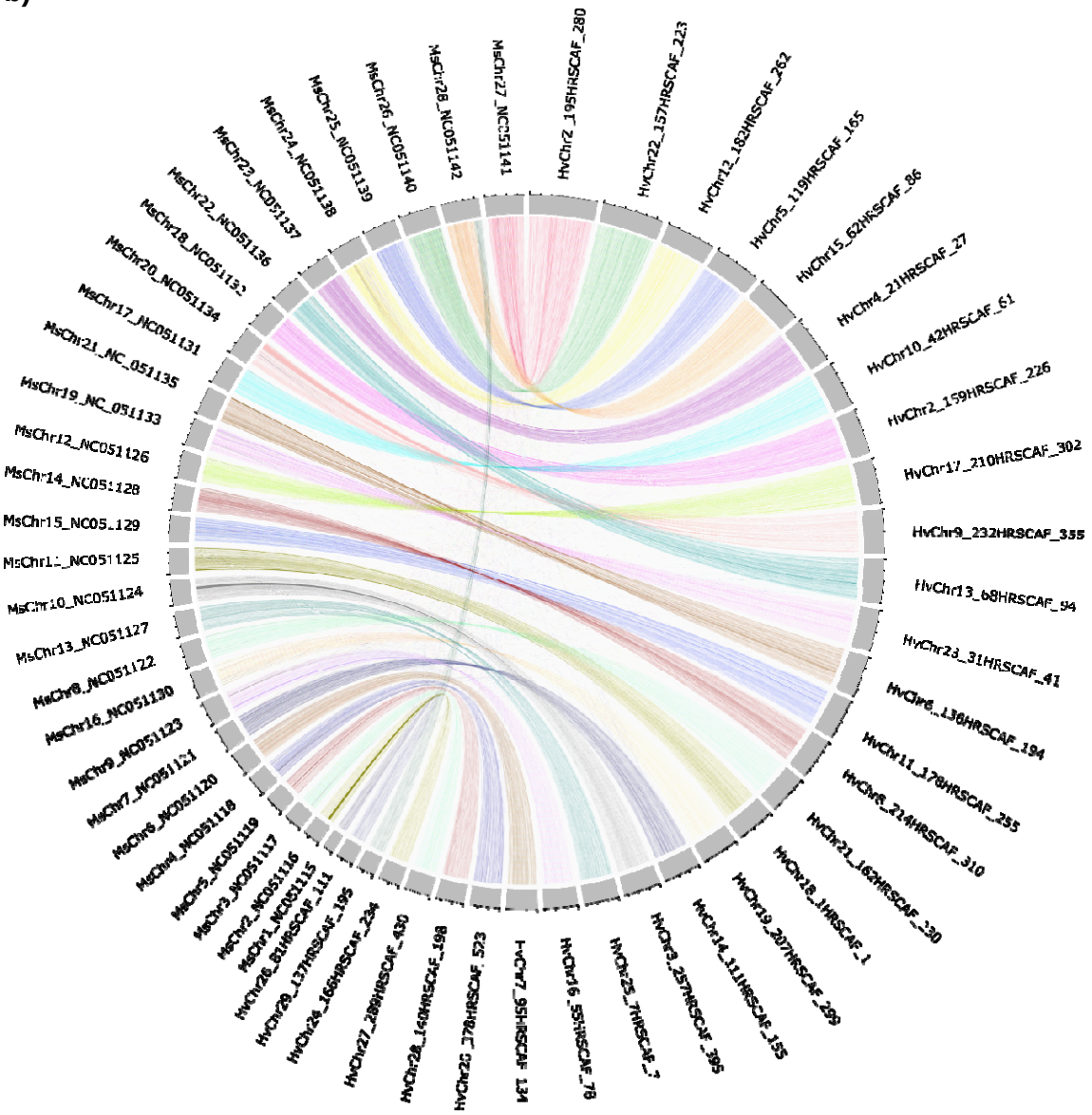
chromosome rearrangements between *B. mori* and *H. vespertilio* are consistent with those between *B. mori* and *M. sexta* [69], corroborating high chromosome homology between the two hawkmoth species, *H. vespertilio* and *M. sexta* (Fig. 5b). Chromosome 23 of *B. mori* (BmChr23) is split up into *H. vespertilio* chromosomes 23 and 29 (HvChr23, HvChr29), resulting in an additional chromosome in *H. vespertilio* (HvChr29).

The plot comparing the 29 chromosome sequences (Fig. 5c) illustrates the high synteny within the genus *Hyles* in the definition of colinearity, the conservation of blocks of order within 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. 5c).

a)


BmChr23
HvChr2_195HRSCAF_280
HvChr22_157HRSCAF_223
HvChr12_182HRSCAF_262
HvChr5_119HRSCAF_165
HvChr15_62HRSCAF_96
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)

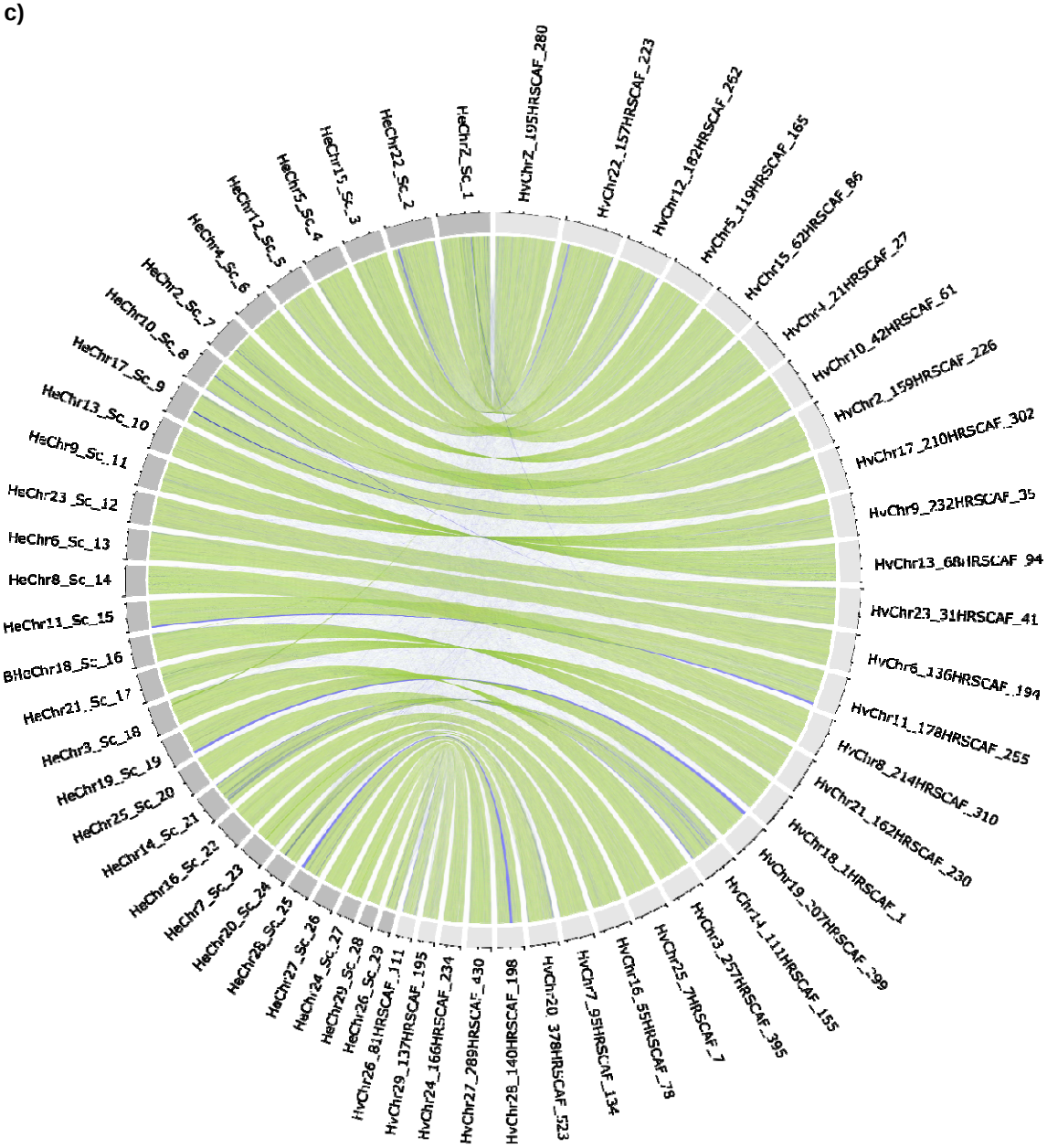


Fig. 5. CIRCOS plots of genome alignments at chromosome level. Chromosome Z corresponds to chromosome 1. Chromosomes are ordered according to size. The letters "ScWp86a" in *H. vespertilio* scaffold names are omitted for clarity. **a**) *B. mori* and *H. vespertilio* with automatic coloration by shinyCircos to facilitate correlation. **b**) *M. sexta* and *H. vespertilio* with coloration 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 strandiness of the chromosomes of *H. euphorbiae* in comparison to the *H. vespertilio* strand orientation (green= +; blue= -).

Wing pattern genes

The wing pattern genes *optix, wingless/wnt-1* and *cortex* were identified with high confidence using BLAT [30, 31] (see above), since identity, bit score and alignment length were highest for one sequence only with sufficient distance to the second best hits to distinguish them from random hits.

The hits of the *M. sexta optix* gene sequence (Table S1) revealed a position on chromosome 23 in *H. euphorbiae* (HeChr23, "sc_12:1,583,003-1,583,475 ", size: 473 bp) and *H. vespertilio* (HvChr23, "ScWp86a_31_HRSCAF_41:20,799,756-20,800,228 ", size: 473 bp). No repeats are noted in the area of the exon in either species. The divergence plot in 2 Kb windows of the 150 Kb surrounding the gene showed a very high level of sequence identity, the p-distance always below 2%. (Fig. 6a). However, a high percentage of InDels (49.97%) was found in the alignment downstream of the *optix* gene, which is why the alignment shown ends at 64 Kb. Stretches with gaps in either sequence were removed for the calculation of the p-distance.

Table S1: Accession numbers of *Manduca sexta* reference wing pattern gene or protein sequences.

Gene	**Accession number**	**Length**
wingless	XM_037446381.1	2.683 bp
optix	JH668350.1	551 bp
cortex (isoform X1)	XP_030035669.1	475 amino acids

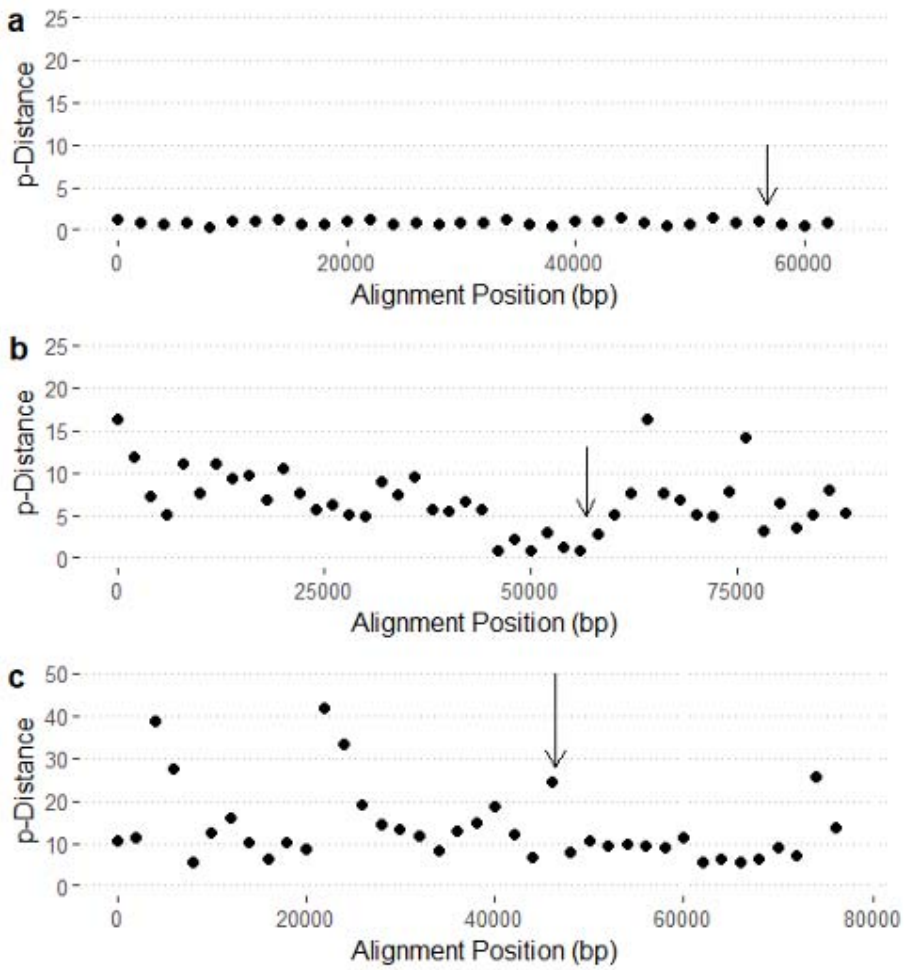


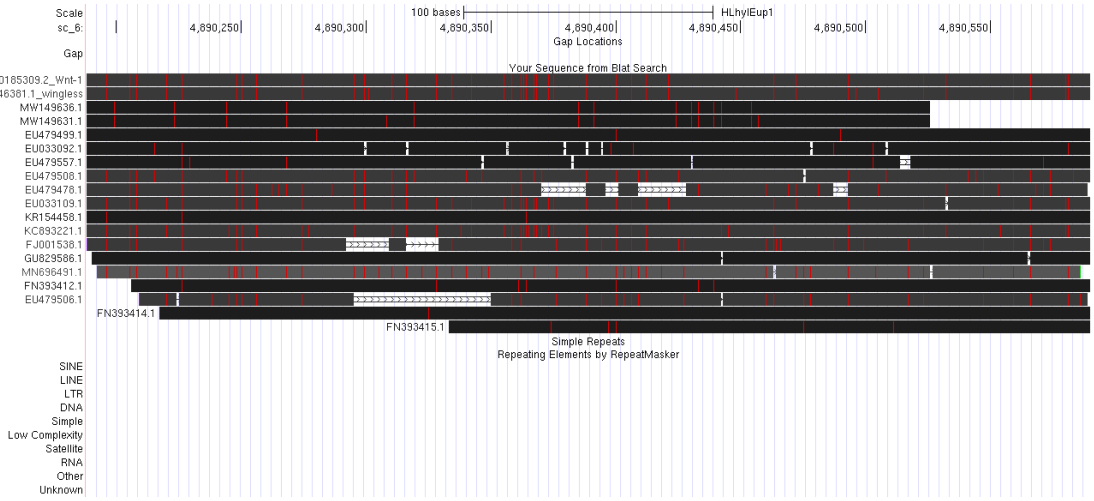
Fig. 6. Genomic divergence in 2 Kb windows surrounding wing pattern genes **a)** *optix,* **b)** *Wnt-A*/*wingless* and **c)** *cortex*. The arrows point to the position of the respective gene.

The public Lepidoptera (Sphingidae) *wingless* gene sequences (Tables S1, S2) mapped to the *Hyles* genomes returned a positioning on chromosome 4 in *H. euphorbiae* (HeChr4, "sc_6:4,890,189-4,890,590" size: 402 bp; Fig. S2a) and *H. vespertilio* (HvChr4, "ScWp86a_21_HRSCAF_27:6,363,295-6,363,697" size: 403 bp; Fig. S2b; Table 3). This region represents the first part of the second exon as annotated in the reference *M. sexta* complete *wingless* sequence (XM_037446381.1; sc_6:4,888,977-4,892,533).

Table S2: Accession numbers and species of additional *wingless* and *wnt-1** sequences of Sphingidae included for comparison.

Accession Number	Genus	species
XM_030185309.2*	*Manduca*	*sexta*
EU033109.1	*Manduca*	*sexta*
FN393412.1	*Deilephila*	*elpenor*
FJ001538.1	*Proserpinus*	*proserpina*
EU479478.1	*Daphnis*	*nerii*
EU479506.1	*Laothoe*	*populi*
EU479508.1	*Macroglossum*	*stellatarum*
EU479557.1	*Xylophanes*	*porcus*
KC893221.1	*Sphinx*	*pinastri*
KR154458.1	*Theretra*	*oldenlandiae*
MN696491.1	*Smerinthus*	*ocellata*
EU033092.1	*Hyles*	*lineata*
EU479499.1	*Hyles*	*hippophaes*
GU829586.1	*Hyles*	*gallii*
MW149631.1	*Hyles*	*calida*
MW149636.1	*Hyles*	*perkinsi*
FN393415.1	*Hyles*	*livornicoides*
FN393414.1	*Hyles*	*euphorbiae*

a)



b)

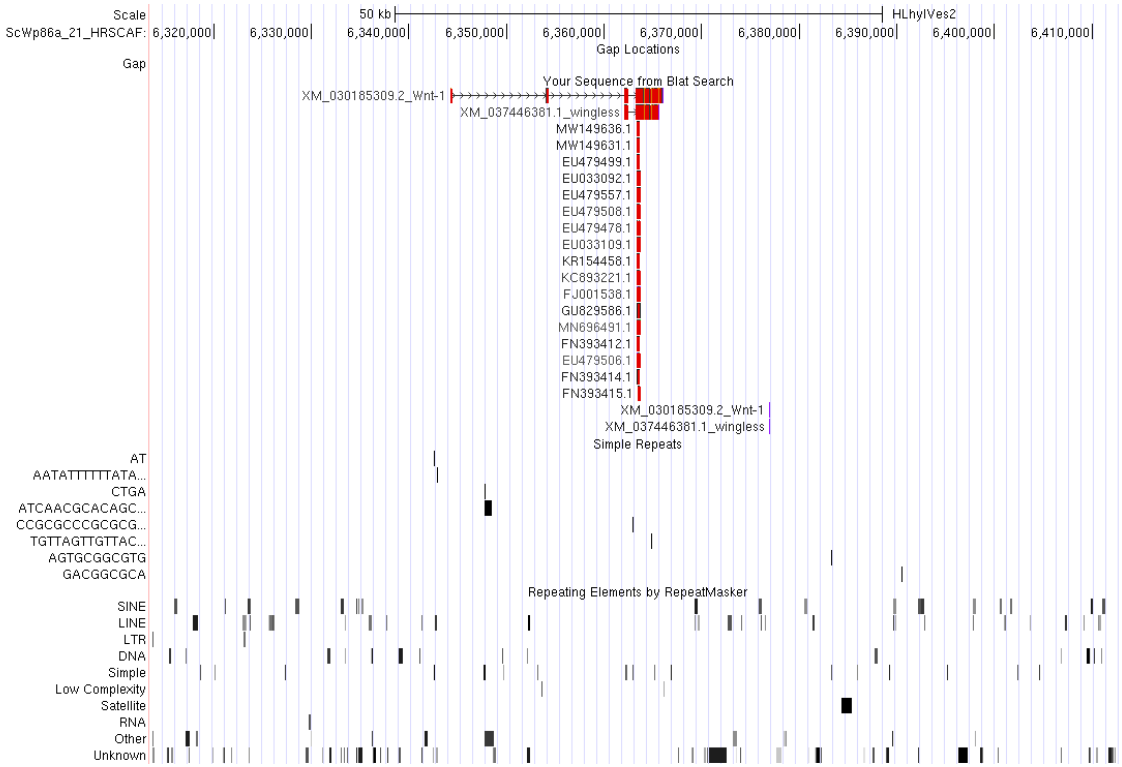


Fig. S2. **View of publicly available *wingless/wnt-1* sequences** (first part of exon 2). **a**) aligned to the spurge hawkmoth genome (looks identical in *H. vespertilio*) and **b**) in a 100 Kb view aligned to the new *H. vespertilio* assembly presented here, showing the repeat occurrences in the vicinity.

The public Lepidoptera (Sphingidae) *wingless* gene sequences (Table S2), the corresponding ~400 bp spurge hawkmoth sequence “sc_6:4,890,189-4,890,590” and the slightly longer fragments of the genomic sequences of *H. euphorbiae* and *H. vespertilio* corresponding to the first and second exon were included to the sequence alignment used to estimate a phylogenetic tree (Fig. S3). The RaxML tree shows monophyletic *Hyles* sequences with *Theretra* and *Deilephila* as sister group.

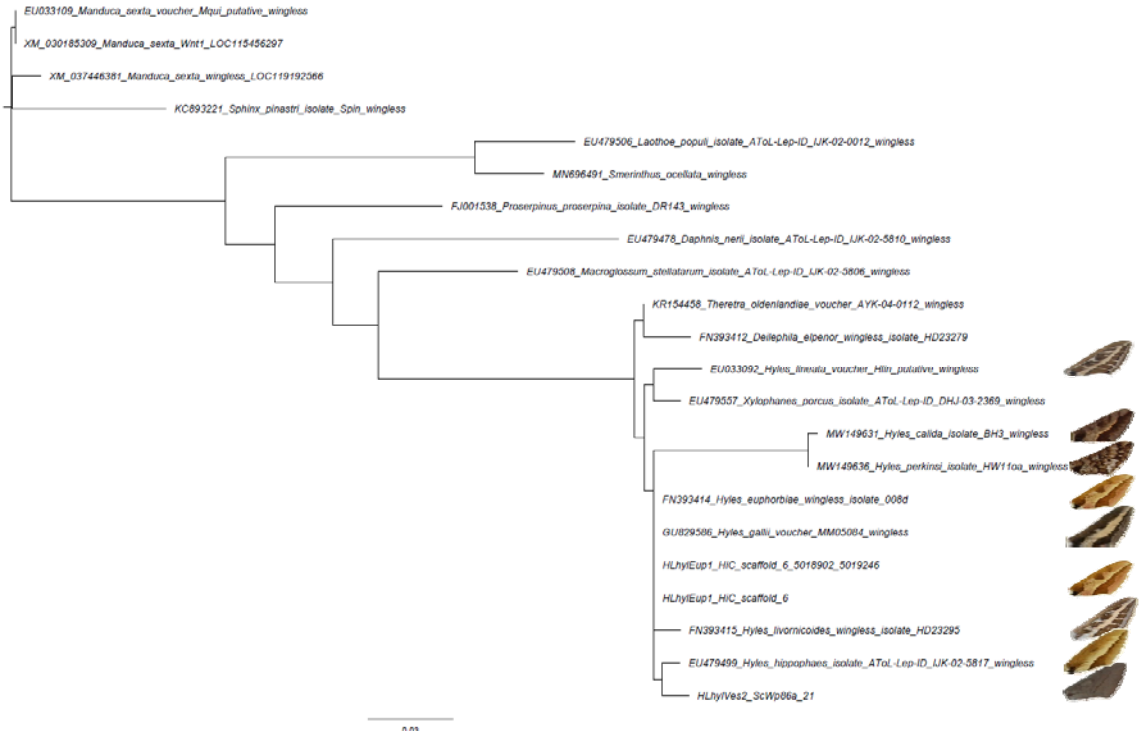


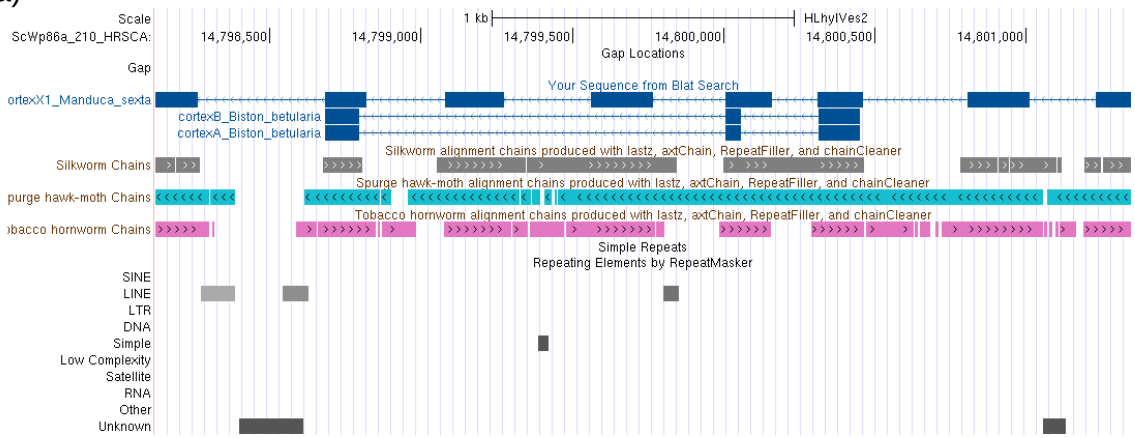
Fig. S3. Phylogenetic hypothesis of a subset of publicly available *Wnt*-A/*wingless* sequences of the family Sphingidae, including the homologous sequences from the two *Hyles* genomes ("HLhylEup1_HiC_scaffold_6_5018902_5019246" represents the same short fragment as the other genebank sequences only, whereas "HLhylEup1_HiC_scaffold_6" and "HLhylVes2_ScWp86a_21" correspond to the length of "XM_037446381_Manduca_sexta_wingless_LOC119192566" and represent both exon 1 and 2, intron insertions excluded).

The divergence plot of the 150 Kb surrounding the *wingless/wnt-1* gene exons (Fig. 6b) shows a pattern with three peaks (at 1-2000 bp, 16.4%; 64001-66000 bp, 16.3%; 76001-78000 bp, 14.2%), and very low diversity at and in the vicinity of exon1 (56627-59348 bp in the alignment; window 56001-58000 bp, 0.95%).

BLAT search of *M. sexta cortex* protein sequence resulted in a match corresponding to 8 exons on chromosome 17 in both species (Fig. S4; *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 position, quality and quantity of repeats in the introns differs between the two species, e.g. the intron

between the first two exons has no repeats in *H. euphorbiae* but two LINEs plus a large stretch of unknown repeats in *H. vespertilio*. The 100 Kb view reveals a high number of repeats in the vicinity of the stretch of gene exons in both species (Fig. S5).

a)



b)

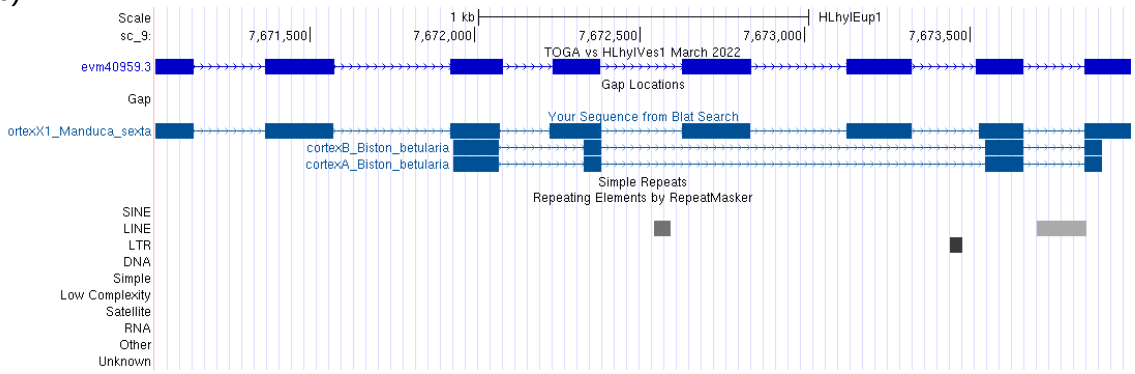
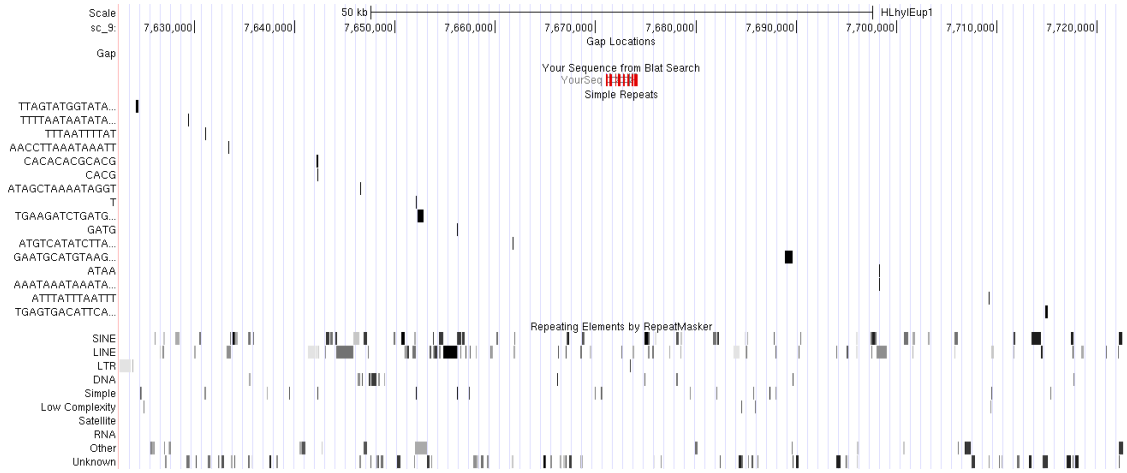


Fig. S4. BLAT result of the *M. sexta* and *B. betularia cortex* protein (Table S1) on chromosome 17 in **a**) *H. euphorbiae* (above) and **b**) *H. vespertilio* (below) showing the exon/intron structures. Both exon and intron sizes appear to differ between genera and species (scales are similar).

a)



b)

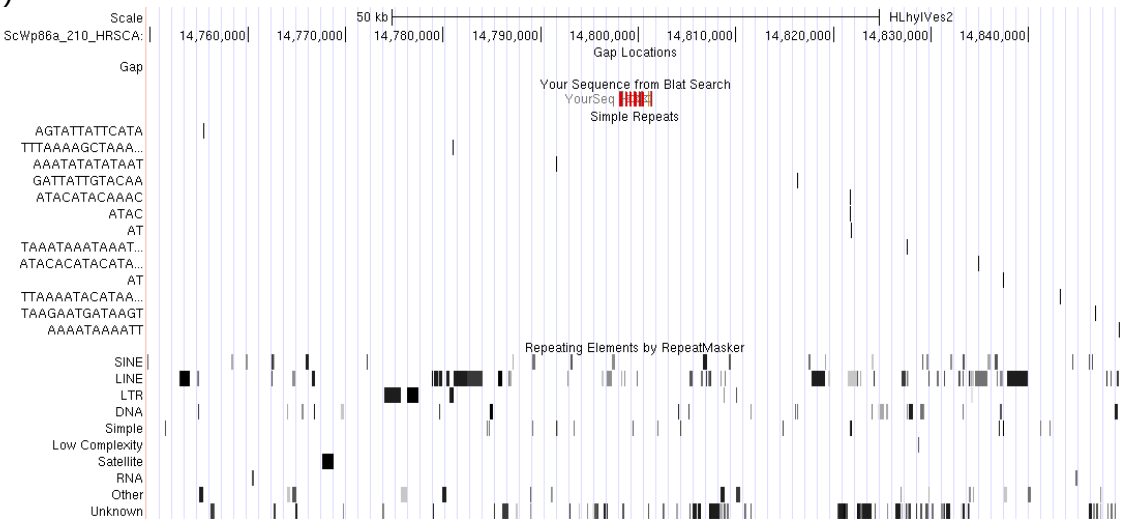


Fig. S5. BLAT result of the *M. sexta* mRNA (Table S1) on chromosome 17 in **a**) *H. euphorbiae* and **b**) *H. vespertilio* (below) in a 100 Kb views showing large numbers of repeats in the vicinity, position, length and type differing between species.

The divergence plot of the 150 Kb surrounding the *cortex* gene exons (Fig. 6c) shows a pattern with three very high peaks over 30% divergence (at window 4001-6000 bp, 38.9%; 22001-24001 bp, 41.9; 24001-26000 bp, 33.5%).

Discussion

Karyotype

The genome alignment of *B. mori* and *H. vespertilio* assemblies (Fig. 5a) allowed homology of chromosomes and thus a well-grounded chromosome taxonomy for the two *Hyles* species studied here (Fig. 5c), but also for the entire genus *Hyles* in all future work. In fact, we strongly emphasize the value of this chromosome taxonomy for all Lepidoptera based on homology reasoning. The new hawkmoth assemblies [13, 32], including that of *M. sexta* [26], only use arbitrary numbers (see Table S3) according to chromosome size, which prevents well-founded future comparisons for specific research questions, such as for an example the wing pattern gene homologies in the study at hand.

Table S3: Chromosome and scaffold name conversion table of *B. mori*, *M. sexta*, *H. euphorbiae* and *H. vespertilio*.

In lepidopteran karyotype evolution, BmChr11, BmChr23 and BmChr24 are often split up in other lepidopteran species, which increases chromosome numbers compared to *B. mori*, e.g. [33-35]. The ancestral number of chromosomes is considered to be *n* = 31 in the haploid Lepidoptera genome [71-73]. *Bombyx mori* (n = 28) and *Hyles* (n = 29) would appear to have undergone karyotype reductions from more basal taxa, independently of e.g. *Heliconius* (21 chromosomes). The well-known chromosome architecture, e.g. [36] in that genus was studied with linkage maps, e.g. [37-39].

The *H. euphorbiae* karyotype chromosome images obtained had a sufficient clarity to be annotated with a size estimate by napari-karyotype (Tab. 3) [25]. However, it should be noted that it was more difficult than expected, as the chromosomes were touching one another 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 in the pachytene stage. To be able 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 fluorescence-labelled probes following Yasukochi et al. [33] (Figure 2) in the future.

Assembly Quality

Contiguity measured through scaffold N50 of *H. euphorbiae* and *H. vespertilio* is higher compared to the related species *M. sexta* and *B. mori*. The percentages of assembly length in *Hyles euphorbiae* contained in the longest scaffolds is the largest (Table 2). Both measures underline the very high quality of the two *Hyles* assemblies. The difference between scaffold and contig N50 is highest in *M. sexta*, suggesting a more fragmented underlying assembly in comparison.

BUSCO completeness of all compared assemblies is similarly high. The only noteworthy exception might be that the duplication rate is highest in *M. sexta* with 4.6% (Table 2).

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 flow cytometry estimation, is unexpected. *Hyles lineata* is one of the oldest species of the genus [2] and it has an even smaller flow cytometry genome size estimate of 450 Mb (0.46 pg) [28]. The lower genome size estimate of this third *Hyles* species allows us to postulate that the larger genome of *H. vespertilio* is more derived.

Not directly comparable are the values of the genome sizes based on the assembly lengths. But the relative sizes are of 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.

The *H. euphorbiae* genome presented in this work has much fewer repetitive elements than its congener, especially LINEs (Long INterspersed Elements) and other repeats. The high content of repeats found in all four genomes, especially LINE is typical for lepidopteran genomes [40, 41]. However, the number of repeats varies among the genomes. Previous research has showed a correlation between the repetitive fraction of the genome, known as the repeatome, and genome size within and among species [42, 43]. Indeed, the genome of *H. vespertilio* is the largest in our comparison and the genome with the most extensive repeat content, whereas the other two sphingid genomes show decreasing amounts of repeats correlated with their genome size. As described in previous research, the repetitive elements found here are thus likely drivers for genome size expansion, possibly due to positive feedback that allows these elements to spread more easily in large genomes [44].

Furthermore, it is assumed that the repeatome plays a significant role in genome size evolution as well as genetic innovation and speciation [45-49]. As *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 [6, 50, 51].

As they belong to the same genus, a high synteny was expected between *H. vespertilio* and *H. euphorbiae* on the nuclear set of chromosomes (Fig. 5) and indeed the illustrated genome alignment (Fig. 5b) shows how similar the two *Hyles* genomes are. The mitochondrial genome of *H. euphorbiae* (Fig. S1) is also highly similar to that of *H. vespertilio* (see [12]).

Wing pattern genes

The locations of the three wing pattern genes have been well studied in *Heliconius*; chromosome 18 (o*ptix*) [14, 15], chromosome 10 (*wntA*) [14] and chromosome 17 (*cortex*) [21, 52] are unsurprisingly not the same in *Hyles* (*optix* on chromosome 23, *wnt-1* on chromosome 4 and *cortex* on chromosome 17) given *Heliconius* has only 21 chromosomes, eight less than *Hyles*.

optix

The invariance found at and close to *optix* was expected, as *optix* and surrounding genes are highly conserved within Lepidoptera [52]. In contrast, the high percentage of InDels found in the alignment downstream of *optix* between the genomes of *H. vespertilio* and *H. euphorbiae* could provide support for the hypothesis proposed by

[53], that the wing patterns are actually controlled by cis-regulatory elements close to the position of *optix*. Zhang et al. [15] showed that *optix* knock-outs show complete replacement of color pigments with melanins, resulting in black and grey butterflies. Although there are no particular divergence peaks within the area surrounding *optix* between *H. euphorbiae* and *H. vespertilio*, the stretches containing indels had to be deleted and could thus not be taken into account (n = 2). The question as to whether this gene could cause the grey wings in *H. vespertilio* thus cannot be answered until more sequence data is available from more individuals to calculate F_{ST} plots.

wingless/wnt-1

The variability between the two species *H. euphorbiae* and *H. vespertilio* in the genomic stretches surrounding *wingless/wnt-1* suggests the hypothesis that regulation of this gene influences the wing pattern determination in *Hyles*. The two species have very dissimilar forewings and the areas of peak p-distances between them could determine one or the other. Sequence data from more individuals is needed for a robust correlation.

cortex

The potential influence of *cortex* on forewing pattern development is strongly suggested by the large number of high p-distance peaks found between *H. euphorbiae* and *H. vespertilio*. The large high p-distance of over 40% in species that are only around 4% apart on neutral phylogenetic markers [6] strongly suggests an influence on determination of the highly dissimilar forewings.

Another striking difference between the two species in the vicinity of the stretch of the 8 *cortex* exons is an insertion of 229 bp marked as unknown repeat by repeatmasker.

Using BLAT, mapping onto the genome of *H. euphorbiae*, it yields 200 hits between 89.8-96.7% identity on every chromosome, and 203 hits on the new genome of *H. vespertilio* (data not shown). An NCBI blast yielded exactly one hit on every chromosome of the very closely related macroglossine species *Deilephila porcellus* [32] and *Hemaris fuciformis* (under review), as well as nine hits on the genome of the smerinthine *Mimas tiliae* [13], a species phylogenetically somewhat more distant. These three genomes are yet to be officially published. Analyzes and comparison with the genomes presented here is expected to yield a better understanding of wing pattern evolution within the family Sphingidae.

Conclusions

Earlier studies had already demonstrated that wing pattern similarity does not correlate with phylogenetic relatedness in the genus *Hyles* [2, 6, 7]. Wing patterns do not even reliably correlate with species either as currently defined or as reflected by molecular phylogeny. Morphologists have long argued that without knowledge of molecular data they have to rely on phenotypic characters and often include striking differences in wing patterns in their species descriptions. However, in *Hyles*, the evolution of the wing pattern characters do not reflect the evolution of the species. Of course, gene trees are not species trees [54], which is why traditional genetic analyzes, e.g., [2, 6, 7] also 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 thus opening possibilities for studying wing pattern evolution based on a numerically analyzable,

objective source of data. The two genes *wingless*/*wnt*-A and *cortex* promise utility in *Hyles*, since the genomic areas surrounding these two gene regions show peaks of high divergence between *H. euphorbiae* and *H. vespertilio*, which have very different wing patterns. The genome data at chromosome level provided in this study for these two species represent reliable references in the family Sphingidae for future studies involving as many species as possible to clarify the evolution of forewing patterns in this group of Lepidoptera.

Material and Methods

Material

For the karyotype, *H. euphorbiae* from Greece (leg. P. Mazzei, Serifos) was bred in the lab (summer 2019). Several larvae and young pupae were used to prepare the tissue slides (see below).

For the genome, one specimen of *H. euphorbiae* (Fig. 7) was collected near Berbisdorf (Germany) on 27.7.2021 and a second, similar moth from the same locality was placed as a voucher in the SNSD collection (Senckenberg Naturhistorische Sammlungen Dresden).

Fig. 7. *Hyles euphorbiae* male spurge hawkmoth from near Berbisdorf.

Karyotype

Spread chromosome preparations were made as described by Yoshido et al. [55]. 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 a saline solution, swollen either for 5 min (ovaries) or 15 min (testes and wing imaginal discs) in a hypotonic solution (75 mM KCl) and then fixed for 10 min in Carnoy's fixative (ethanol, chloroform, acetic acid, 6:3:1). 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 carried out following the procedure described by Yoshido et al. [56]. $(TTAGG)_n$ telomeric sequences were generated by non-template PCR according to the protocol of Sahara et al. [57]. Male and female

genomic DNAs (gDNAs) of *H. euphorbiae* were obtained separately from last instar larvae by standard phenol-chloroform extraction. DNA probes were labelled 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-labelled female gDNA, 100 ng of Cy3-labelled telomeric probe, 3 µg of unlabelled 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. 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 karyotype image was preprocessed with the image processing software GIMP (version 2.10) [58] to manually cut out individual chromosomes. The processed picture was loaded into the tool napari-karyotype (version c41103e) [25]. Image segmentation threshold, blur factor and genome size were set to 0.13, 0.5 and 504 Mb respectively.

Genome size estimation

The two hawkmoth genome sizes were estimated following the flow cytometry protocol with propidium iodide-stained nuclei described by Hare and Johnston [59]. Neural tissue of frozen (−80°C) adult samples of *H. vespertilio* and *H. euphorbiae* and neural tissue of 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 with a final concentration of 25 µg/ml. The mean red PI fluorescence of 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. We used the CytExpert 2.3 software for histogram analyzes. 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.

The genome size is reported as 1C, the mean amount of DNA in Mb in a haploid nucleus.

Additionally, genome size was estimated by mapping coverage using ModEst [60]. Estimations were calculated with backmap.pl 0.5 (https://github.com/schellt/backmap), in combination with bwa mem 0.7.17 [61], minimap 2.24 [62], samtools 1.15 [63], qualimap 2.2.1 [64], bedtools 2.30.0 [65], R 4.0.3 [66] and multiqc 1.12 [67]. Briefly, the reads used for assembly were mapped back to the assembly itself. Subsequently, the number of mapped nucleotides was divided by the mode of the mapping coverage distribution.

PacBio Genome DNA and sequencing

Head tissue (38 mg) of *Hyles euphorbiae* was used for high molecular weight DNA extraction using an adaptation of the protocol of Miller et al. [68]. Final DNA purity and concentrations were measured using NanoPhotometer® (Implen GmbH, Munich, Germany) and 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 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 high quality CCS reads (HiFi reads) with a N50 of 11.74 Kb. To further increase the read coverage we applied the tool DeepConsensus (v0.2 with default settings) [69] and gained an overall yield of 8.8 Gb (N50: 11.83 Kb). We ran HiFiasm (version 0. 16.1-r375) [70] to create the contig assembly. Remaining haplotypic duplications in the primary contig set were removed using purge-dups (v.1.2.3) [71]. 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 span the gap regions with 1000 anchor bases at 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 by at least 5x coverage; and 2) the coverage profile of the closed gaps fully supports the consensus sequence (i.e. no alignment breaks or huge repeat alignment piles occur). Two rounds of error polishing were performed by aligning the DeepConsensus reads to the assembly with pbmm2, calling variants with DeepVariant (version 1.3.0) [72] and correcting homozygous errors with bcftools consensus. The assembly was checked for contaminations with blobtoolkit (version 1.1) and an in-house pipeline which screens several blast databases. BUSCO (version 5.2.2) [73] scores and merqury (version 1.3) [74]. QV values were created for the final scaffolds (QV=58.1).

The mitochondrial genome was created with the mitoHifi pipeline (version 2) [75] based on CCS reads and the closely related reference mitochondrial genome of *Theretra oldenlandiae* (NCBI accession: MN885801.1).

Hi-C sequence data

The Dovetail Hi-C libraries for *H. vespertilio* and *H. euphorbiae* were prepared from head tissue (52.2 mg and 40.6 mg) using the Dovetail Hi-C kit (Dovetail Genomics, Scotts Valley, CA, USA) following the manufacturer's protocol version 1.4 for insect samples. Briefly, the chromatin was fixed with formaldehyde then extracted. Fixed chromatin was digested with DpnII, the 5' overhangs filled in 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. The DNA was then sheared to ~350 bp mean fragment size 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 from Pippel et al. [11] and *H. vespertilio* Dovetail Hi-C library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies [76]. 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 Hi-C read pairs mapped within draft scaffolds were analyzed by HiRise version 2.1.7 to produce a likelihood model for genomic distance between read pairs. The model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

Annotation

Repeat annotation for *H. vespertilio*, *H. euphorbiae*, *M. sexta* and *B. mori* was conducted using RepeatModeler 2.0.2a [77] and RepeatMasker 4.1.2-p1 [78] in combination with rmblastn 2.11.0+. For RepeatModeler, the additional option "-LTRstruct" and for RepeatMasker "-s -xsmall -e ncbi" were used. The repeat library for RepeatMasker contains a combination of all Repbase entries (release 26.07) from Lepidoptera and all repeat families identified from RepeatModeler.

Structural annotation of protein coding genes was conducted using TOGA [79], a method that uses pairwise genome alignment chains between an annotated reference genome (here *Hyles vespertilio* assembly) and other query species (here *Hyles euphorbiae*). Briefly, TOGA uses machine learning to infer orthologous loci for each reference transcript, utilizing 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.

Comparison to other species

The assemblies of *H. euphorbiae* and *H. vespertilio* are compared to those of the model species *Manduca sexta* (GCF_014839805.1) [26] and *Bombyx mori* (GCF_014905235.1) [27]. The *Mimas tiliae* [13] and *Deilephila porcellus* [32] genomes were still under review and was not published in time to include in a detailed comparison in this work. To compare contiguity between *H. euphorbiae*, *H. vespertilio*, *M. sexta* and *B. mori*, Quast 5.0.2 [84] was utilized. Assessment of completeness regarding single copy orthologs was conducted via BUSCO 4.1.4 (Manni et al., 2021) together with the lepidoptera_odb10 set and the options "`--long --offline`".

Genomes were aligned using LASTZ 1.04.03 [85] with parameters (K = 2400, L = 3000, Y = 9400, H = 2000 and the lastz default scoring matrix). Then, we used axtChain [30] (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 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 S4 (supplementary file).

Table S4: Chromosome proportion values of the *B. mori* – *H. vespertilio* alignment.

Wing pattern genes

The positions of wing pattern genes *optix, wingless/wnt-1* and *cortex* (from *M. sexta*, accession numbers see Table S1) were identified in the two *Hyles* genomes by using the BLAT tool [30, 31] with default options as implemented in the Senckenberg Genome browser. Blat results are presented sorted by alignment length and the longest was chosen for every gene. The resulting *Hyles* alignment in an interval of +/- 70 Kb around the exons was downloaded from the genome browser for a plot illustrating the divergence using the proportion (p) of nucleotide sites at which the two genome sequences compared are different. Genomic divergence based on p-distance values (in percent) is plotted for 2 Kb windows.

Additional public *wingless/wnt-1* sequences of further individuals of the genus *Hyles* and the family Sphingidae (Table S3) from different sources (e.g. AToL, own) were

mapped to the *H. euphorbiae* genome using BLAT [30, 31]. The alignment together with the *M. sexta* (XM_037446381) BLAT sequence results from both *Hyles* genomes was used to produce a small phylogenetic tree using the online RaxML [88] BlackBox portal (https://raxml-ng.vital-it.ch/#/) to illustrate orthology and variability.

For the gene *cortex*, we additionally compared *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 [89] caused by a transposon in a *cortex* intron [22, 90].

Acknowledgements

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). It benefitted from the sharing of expertise within the DFG priority program SPP 1991 Taxon-Omics. MP was partially funded by the BMBF (grant 01IS18026C). AY and FM acknowledge support from grant 20-13784S of the Czech Science Foundation. Hi-C sequence data were obtained in cooperation with LOEWE-TBG (Frankfurt a.M.). We thank Ian J. Kitching for language correction and helpful suggestions.

References

1. Hundsdoerfer AK, Tshibangu JN, Wetterauer B, Wink M: **Sequestration of phorbol esters by aposematic larvae of *Hyles euphorbiae* (Lepidoptera: Sphingidae)?** *Chemoecology* 2005, **15:**261-267.
2. 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.
3. 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:**35-47.
4. 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:**2136–2150.
5. 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
6. 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.
7. 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:**552-570.
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:**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:**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:**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 under review [version 1; peer review: 2 approved] **6**.
14. 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:**52.
15. 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:**10707-10712.
16. 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*. *Volume* 54: Elsevier; 2018: 85-115

17. Sharma RP, Chopra VL: **Effect of the *Wingless* (*wg1*) mutation on wing and haltere development in *Drosophila melanogaster*.** *Dev Biol* 1976, **48:**461-465.
18. Nusse R, Brown A, Papkoff J, Scambler P, Shackleford G, McMahon A, Moon R, Varmus H: **A new nomenclature for *int-1* and related genes: the *Wnt* gene family.** *Cell* 1991, **64:**231.
19. Rubinoff D, San Jose M, Hundsdoerfer AK: **Cryptic diversity in a vagile Hawaiian moth group suggests complex factors drive diversification.** *Mol Phylogenet Evol* 2021, **155:**107002.
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**.
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:**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:**102-105.
23. 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:**2864-2878.
24. Stevens M, Merilaita S: **Animal camouflage: current issues and new perspectives.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2009, **364:**423-427.
25. napari, contributors: **napari: a multi-dimensional image viewer for python.** 2019.
26. 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:**1-9.
27. 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.
28. Hanrahan SJ, Johnston JS: **New genome size estimates of 134 species of arthropods.** *Chromosome Research* 2011, **19:**809-823.
29. Rasch EM: **The DNA content of sperm and hemocyte nuclei of the silkworm, *Bombyx mori* L.** *Chromosoma* 1974, **45:**1-26.
30. 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:**11484-11489.
31. 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:**D1115-D1122.
32. 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 under review [version 1; peer review: 1 approved], **7:**80.
33. 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:**e7465.
34. 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:**644-653.

35. 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:**283-295.
36. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, Garcia-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al: **Genomic architecture and introgression shape a butterfly radiation.** *Science* 2019, **366:**594-599.
37. 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:**138-154.
38. 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:**557-570.
39. 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:**417-426.
40. 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:**7680-7685.
41. Lavoie CA, Platt RN, Novick PA, Counterman BA, Ray DA: **Transposable element evolution in *Heliconius* suggests genome diversity within Lepidoptera.** *Mob DNA* 2013, **4:**21.
42. Lynch M, Walsh B: *The origins of genome architecture* Sunderland, MA: Sinauer Associates; 2007.
43. 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:**1180-1193.
44. 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:**1419-1428.
45. Charlesworth B, Sniegowski P, Stephan W: **The evolutionary dynamics of repetitive DNA in eukaryotes.** *Nature* 1994, **371:**215-220.
46. 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:**2491-2505.
47. 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:**756-760.
48. 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:**228-239.
49. Maumus F, Fiston-Lavier AS, Quesneville H: **Impact of transposable elements on insect genomes and biology.** *Curr Opin Insect Sci* 2015, **7:**30-36.
50. Pittaway AR: *The hawkmoths of the Western Palaearctic.* Colchester: Harley Books; 1993.

51. 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.
52. 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
53. 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:**1137-1141.
54. Maddison WP: **Gene trees in species trees.** *Systematic Biology* 1997, **46:**523-536.
55. 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
56. 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:**193-202.
57. Sahara K, Marec F, Traut W: **TTAGG telomeric repeats in chromosomes of some insects and other arthropods.** *Chromosome Research* 1999, **7:**449-460.
58. Team TGD: **GIMP. Retrieved from https://www.gimp.org.** 2019.
59. 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
60. Pfenninger M, Schönnenbeck P, Schell T: **ModEst: Accurate estimation of genome size from next generation sequencing data.** *Molecular Ecology Resources* 2021.
61. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** *arXiv preprint arXiv:13033997* 2013.
62. Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics* 2018, **34:**3094-3100.
63. 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:**2078-2079.
64. Okonechnikov K, Conesa A, Garcia-Alcalde F: **Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data.** *Bioinformatics* 2016, **32:**292-294.
65. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:**841-842.
66. R Core Team R: **A Language and Environment for Statistical Computing.** 2021.
67. Ewels P, Magnusson M, Lundin S, Kaller M: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics* 2016, **32:**3047-3048.
68. Miller S, Dykes D, Polesky H: **A simple salting out procedure for extracting DNA from human nucleated cells.** *Nucleic Acids Research* 1988, **16:**1215.
69. 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.
70. 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:**170-175.

71. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics* 2020, **36:**2896-2898.
72. 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:**983-987.
73. 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.
74. Rhie A, Walenz BP, Koren S, Phillippy AM: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biology* 2020, **21:**1-27.
75. 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:**892-905.
76. 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:**342-350.
77. Smit A, Hubley R, Green P: **RepeatModeler Open-1.0. 2008–2015.** *Institute for Systems Biology, Seattle, USA Available from: http://wwwrepeatmaskerorg* 2015.
78. Smit A, Hubley R, Green P: **RepeatMasker Open-4.0. 2013–2015. Available from: http://www.repeatmasker.org.** 2015.
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:**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:**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:**313-319.
83. Yu Y, Ouyang Y, Yao W: **shinyCircos: an R/Shiny application for interactive creation of Circos plot.** *Bioinformatics* 2018, **34:**1229-1231.
84. Gurevich A, Saveliev V, Vyahhi N, Tesler G: **QUAST: quality assessment tool for genome assemblies.** *Bioinformatics* 2013, **29:**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:**1596-1603.
88. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A: **RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.** *Bioinformatics* 2019, **35:**4453-4455.
89. Cook LM: **The rise and fall of the Carbonaria form of the peppered moth.** *Q Rev Biol* 2003, **78:**399-417.

90. 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:**958-960.