

Supplementary information

Supplementary methods

LepMap3 protocol ([Rastas 2017](#)) was used to construct the linkage map. The module ParentCall calls informative parental markers and uses genotype likelihood information from the offspring to impute missing or erroneous parental markers. This module was run with default values, except that non-informative markers were removed and $zLimit=2$, which was applied to detect markers segregating as sex chromosomes. Markers mapping to the W-chromosome, mitochondria or to repeats were removed with BEDTools ([Quinlan and Hall 2010](#)). The markers were assigned to linkage groups using SeparateChromosomes2 with $lodDifference=2$ and $distortionLod=1$. The LOD-limit was estimated empirically by testing a range of LODscores (1-30) and finally set to 24, which resulted in the expected number (31) of linkage groups. To assign additional unlinked markers to the linkage groups, JoinSingles was run with $lodLimits = 18$.

OrderMarkers was run over 50 iterations for each linkage group to determine the most likely distance between the markers and the maps with the highest likelihood were selected for further refinement. Since butterflies have female achiasmy, we limited the analysis to markers that were informative only in males ($informativeMask=1$) or in both sexes ($informativeMask=13$). To account for partial interference the Kosambi distance method was applied. The trimmed map was reevaluated with OrderMarkers with the options $evaluateOrder$ and $improveOrder=1$. The maps were thinned so that only SNPs > 300 bp apart were retained (i.e. at least one SNP per RAD-tag). Any remaining unlinked markers at linkage group ends were manually removed after visual inspection and the final maps were once more reevaluated with OrderMarkers. Collinearity between physical and genetic positions were tested with Pearson's product moment correlation as implemented in $cor.test$ in R ([R Core Team \(2021\)](#)).

Supplementary figures

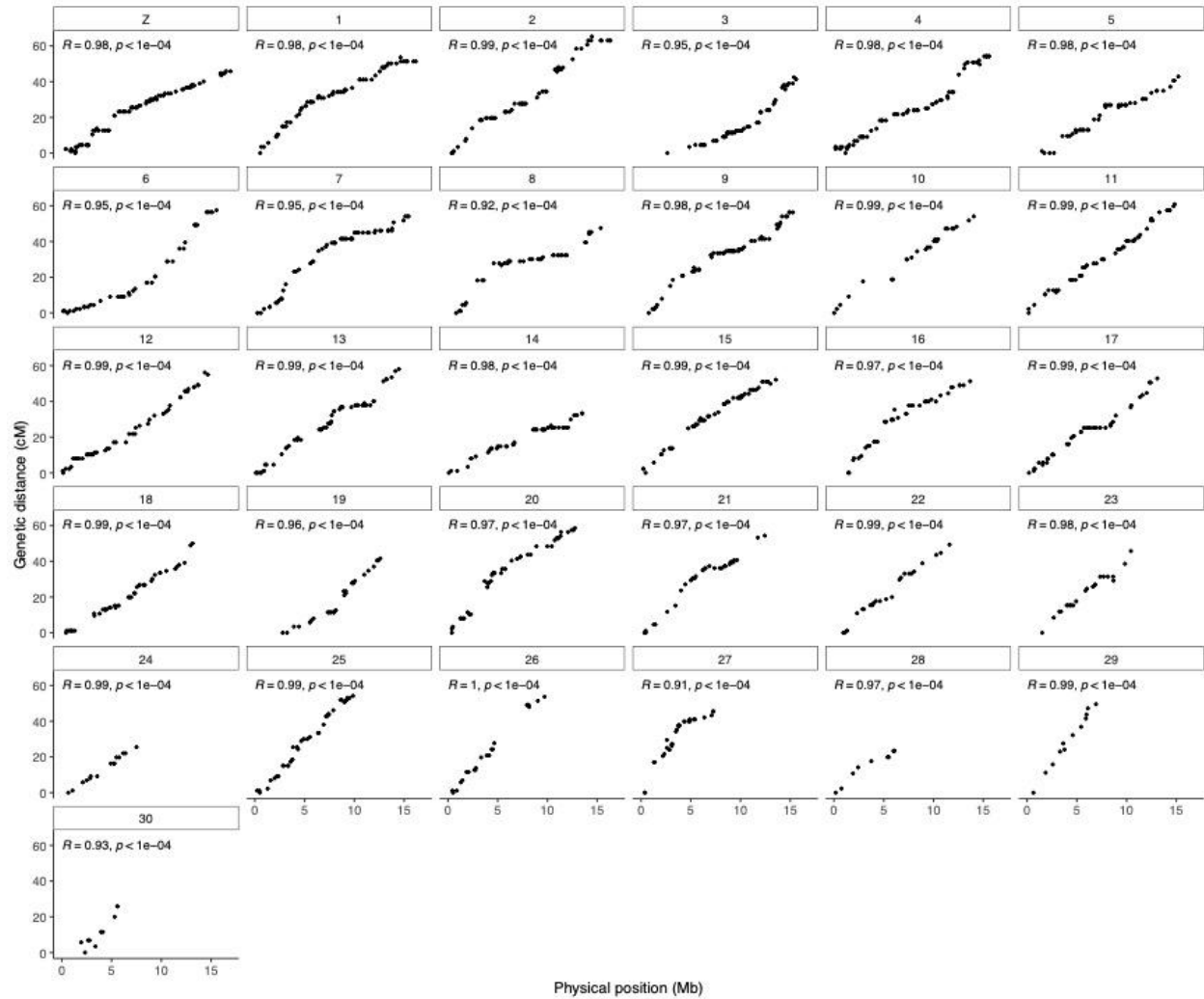


Figure S1.

Marey maps for each chromosome showing recombination distance in centiMorgan on the y-axis and physical position on the chromosome in megabases on the x-axis. Pearson's correlation coefficient (R) and p-values (p) are included.

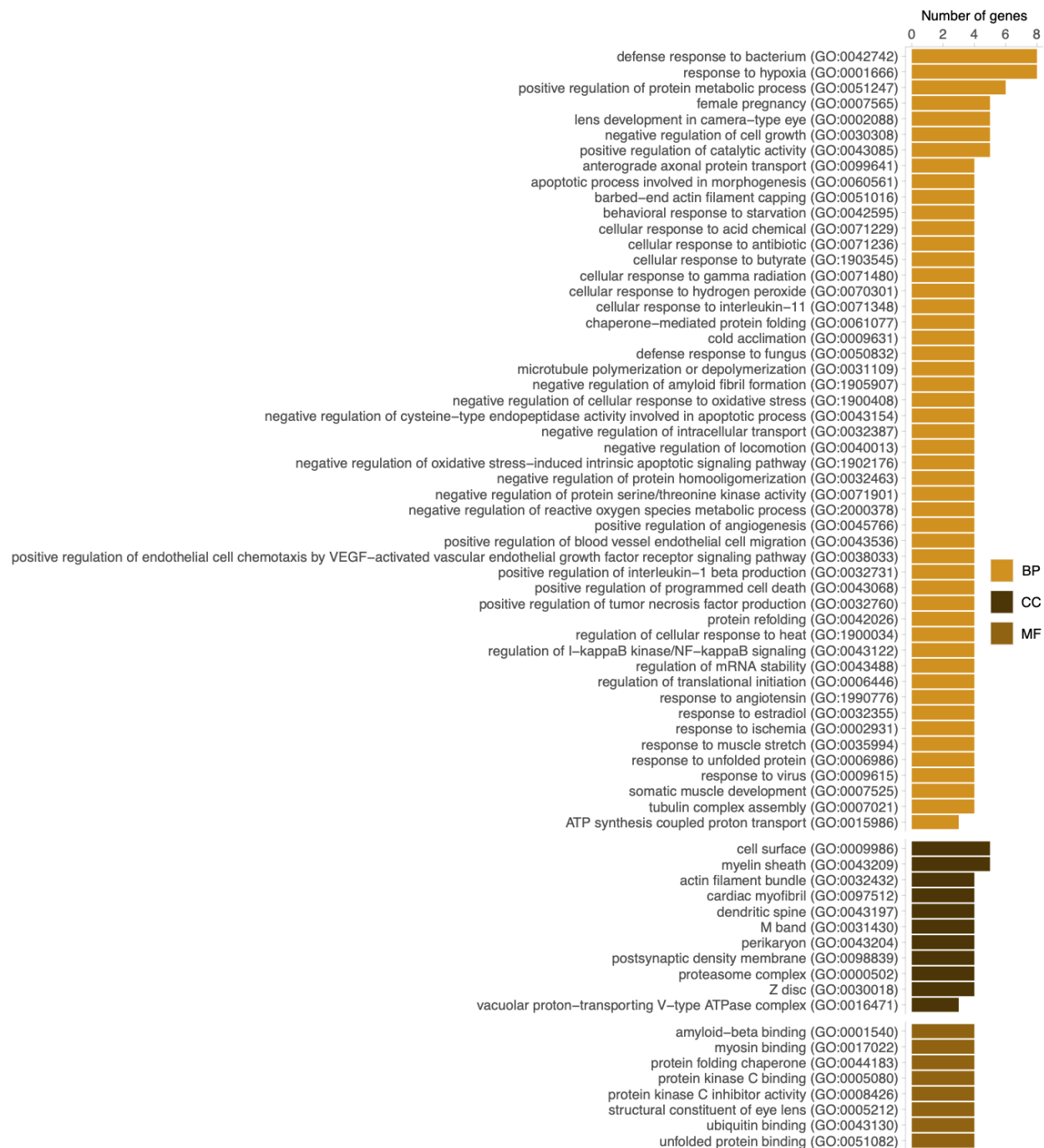


Figure S2.

Significantly (p -value < 0.05 after FDR-correction) enriched gene ontology (GO) terms associated with expanded gene families in both the painted lady and monarch. The bars show the number of genes associated with each GO-term. The different GO-categories are biological process (BP), cellular compartment (CC) and molecular function (MF).

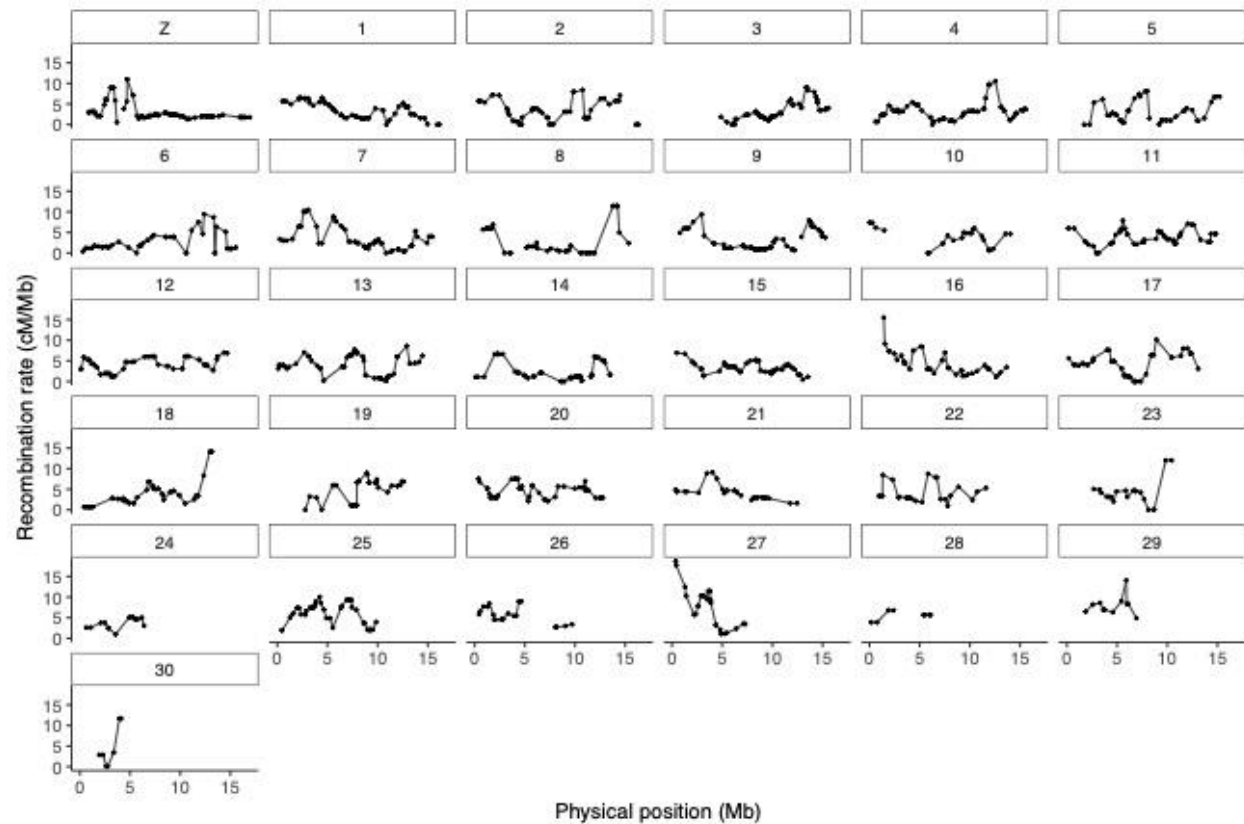


Figure S3.

Physical position (Mb, x-axis) and the estimated recombination rate (cM / Mb; y-axis) for individual markers along each chromosome. Chromosome identities are given in the boxes above each chromosome plot. The W-chromosome has been excluded.

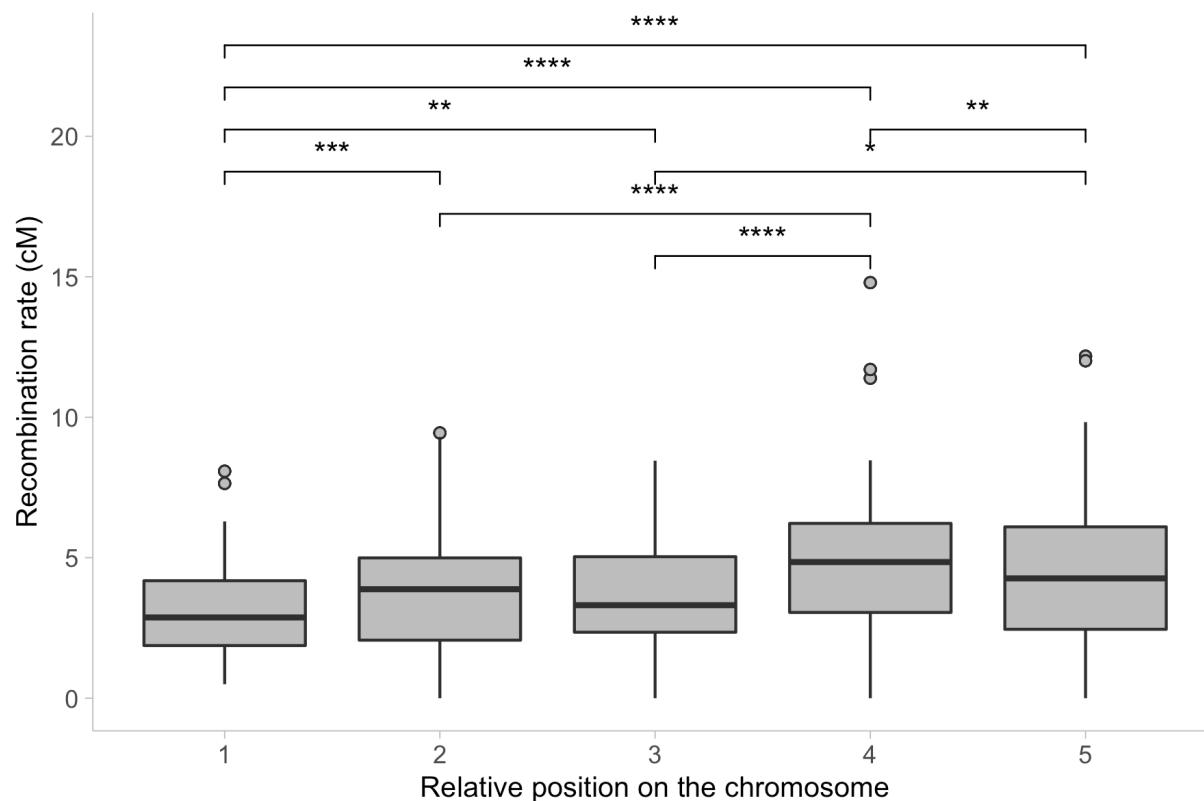


Figure S4.

The mean recombination rate for markers in bins representing non-overlapping windows spanning 10% of the physical distance from the center of the chromosome. Stars indicate pairwise Wilcoxon Rank-test (p-values) comparing the average recombination rate in different relative positions along the chromosome (* = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001, **** = p-value < 0.0001). Boxes hinges represent the 25th and 75th percentiles, whiskers extend to values within 1.5 times the distance between the 25th and 75th percentiles.

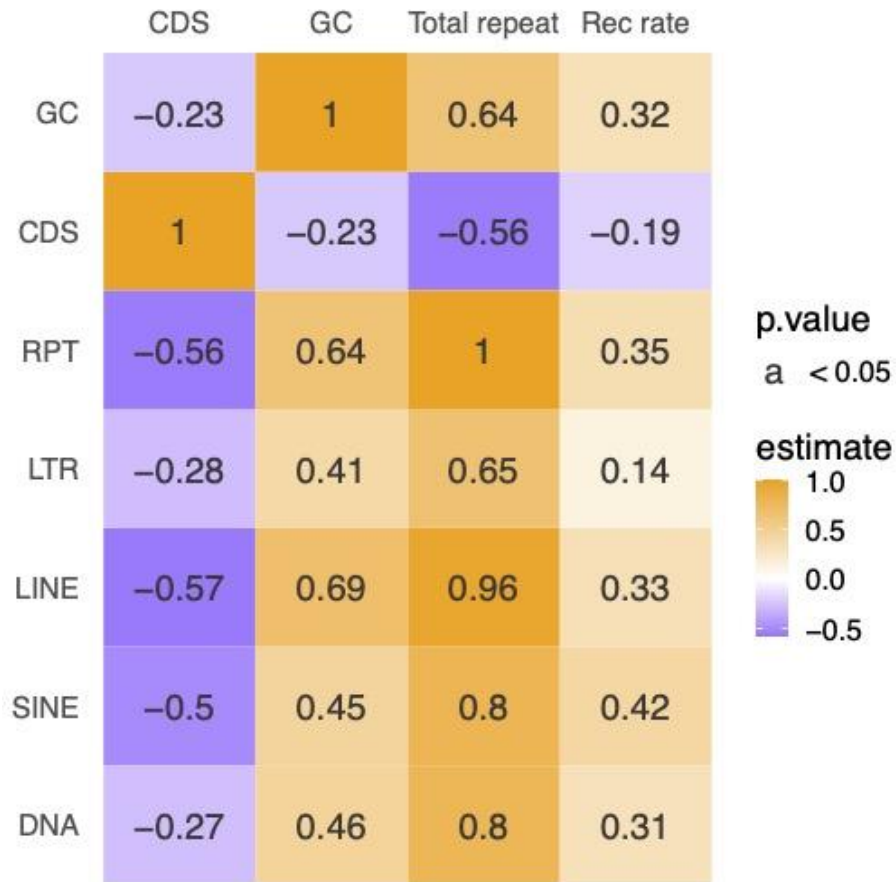


Figure S5.

Pair-wise Spearman's rank correlation statistics between the proportion of genes (CDS), the proportion of total number of repeats (Total repeats, RPT), the recombination rate (Rec rate) and GC-content (GC) and the specific proportion of different TE classes. The values in the squares represent the correlation coefficient (ρ), values with p-value < 0.05 are displayed.

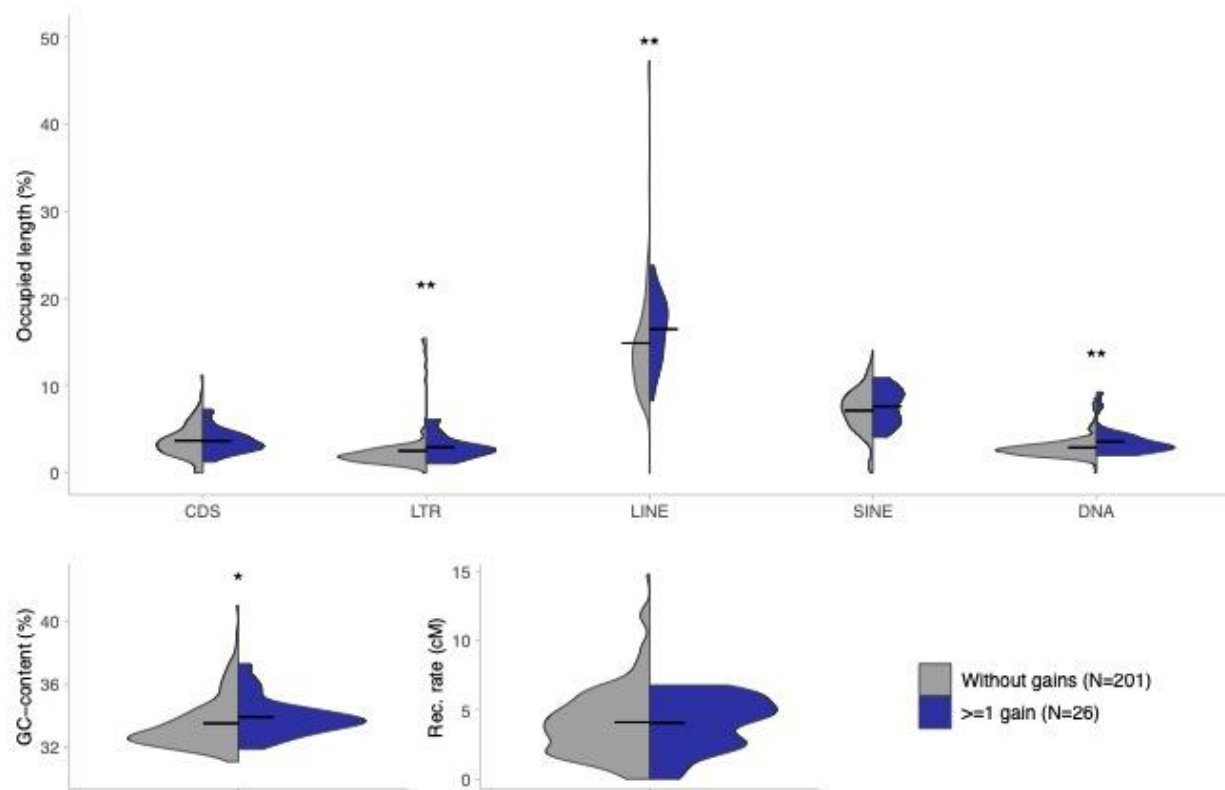


Figure S6.

Upper panel: Mean (horizontal bar) and overall distributions (violin plots) of the proportion of genes (CDS) and different TE-classes in non-overlapping windows (2 Mb), with (blue) and without (grey) gene gains, respectively. DNA = DNA-transposons. Lower panel: Mean (horizontal bar) and overall distributions (violin plots) of the nucleotide composition (GC-content; left) and recombination rate (right panel). Stars indicate Wilcoxon's rank sum test p-values (* = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001).

Supplementary tables

Table S1.

Result per species from OrthoFinder with included species, annotation version, total number of genes, number of genes assigned to orthogroups, number of unassigned genes, species-specific orthogroups and genes in species-specific orthogroups.

Species	Annotation version	Nr of genes	Genes in orthogroups	Genes in orthogroups (%)	Unassigned genes (%)	Orthogroups containing species	Species-specific orthogroups	Genes in species-specific orthogroups
<i>Bicyclus anynana</i>	Bicyclus_anynana_BaGv2.proteins.faa.gz	18188	15764	86.7	13.3	11668	172	667
<i>Danaus plexippus</i>	Danaus_plexippus_v3_-_proteins.faa.gz	15130	13853	91.6	8.4	11567	53	169
<i>Heliconius erato lativitta</i>	Heliconius_erato_lativitta_v1_-_proteins.faa.gz	14612	13189	90.3	9.7	9841	104	278
<i>Heliconius melpomene melpomene</i>	Heliconius_melpomene_melpomene_Hmel2.5.proteins.faa.gz	17701	15990	90.3	9.7	11188	149	1072
<i>Junonia coenia</i>	Junonia_coenia_Jc_v2.proteins.faa.gz	17244	15444	89.6	10.4	11137	116	424
<i>Maniola hyperantus</i>	GCF_902806685.1_iAphHyp1.1_protein.faa.gz	13936	13769	98.8	1.2	10847	35	115
<i>Pararge aegeria</i>	GCF_905163445.1_ilParAegt1.1_protein.faa.gz	15064	14923	99.1	0.9	10997	61	457
<i>Vanessa cardui</i>		13161	12692	96.4	3.6	10361	19	63
<i>Vanessa tameamea</i>	GCF_002938995.1_ASM293899v1_protein.faa.gz	13266	13209	99.6	0.4	10752	19	56

Table S2.

Chromosome length (Chr length, Mb), chromosome map length (Map length, cM), recombination rate (cM / Mb), mean recombination rate in 2 Mb windows (Wind rec rate, cM / Mb) and standard deviation (Wind rec rate sd), number of markers for the linkage group / chromosome (Markers) and the marker density of each linkage group / chromosome (Marker density).

Chromosome	Chr length (Mb)	Map length (cM)	Rec rate (cM/Mb)	Wind rec rate (cM/Mb)	Wind rec rate sd	Markers	Marker density
Z	17.04	45.79	2.69	3.09	2.19	84.00	4.93
1	16.61	53.60	3.23	3.55	1.99	68.00	4.09
2	16.36	65.27	3.99	3.10	2.56	60.00	3.67
3	16.09	42.40	2.64	3.15	2.26	64.00	3.98
4	16.00	54.19	3.39	3.25	2.15	71.00	4.44
5	15.95	42.86	2.69	3.13	2.64	44.00	2.76
6	15.72	57.56	3.66	2.64	2.30	45.00	2.86
7	15.57	54.07	3.47	3.80	2.81	61.00	3.92
8	15.43	47.54	3.08	2.97	3.84	41.00	2.66
9	15.30	56.32	3.68	3.09	2.44	59.00	3.86
10	14.95	54.07	3.62	3.71	2.23	26.00	1.74
11	14.87	60.89	4.09	3.95	1.81	55.00	3.70
12	14.77	56.15	3.80	4.11	1.62	50.00	3.38
13	14.62	58.10	3.98	4.00	2.26	56.00	3.83
14	14.61	33.32	2.28	2.22	2.11	52.00	3.56
15	13.92	52.16	3.75	3.38	1.33	55.00	3.95
16	13.77	51.32	3.73	4.39	2.85	42.00	3.05
17	13.55	52.86	3.90	3.98	2.62	47.00	3.47
18	13.24	50.05	3.78	3.65	2.94	45.00	3.40
19	12.93	41.59	3.22	4.84	2.70	31.00	2.40
20	12.86	58.60	4.56	4.77	1.78	47.00	3.65
21	12.59	54.36	4.32	4.08	1.73	33.00	2.62
22	11.70	49.35	4.22	4.14	2.28	24.00	2.05
23	11.33	45.74	4.04	4.07	3.08	22.00	1.94

Chromosome	Chr length (Mb)	Map length (cM)	Rec rate (cM/Mb)	Wind rec rate (cM/Mb)	Wind rec rate sd	Markers	Marker density
24	11.20	25.48	2.27	3.65	1.33	15.00	1.34
25	9.96	54.15	5.44	6.02	2.65	41.00	4.12
26	9.84	53.66	5.45	5.58	2.12	23.00	2.34
27	8.26	45.51	5.51	7.43	4.85	29.00	3.51
28	8.18	23.35	2.85	5.56	0.99	11.00	1.34
29	7.38	49.50	6.71	8.00	2.37	12.00	1.63
30	6.17	25.84	4.19	4.67	4.98	10.00	1.62
Mean (sd)		48.89	3.81 (±0.98)	4.13 (±1.29)		42.68	
Total	410.76	1515.63					

Table S3.

Summary of the multiple linear regression analysis with recombination rate as the dependent variable and chromosome length (chr_length), proportion of genes (CDS), total repeats (Total repeats), LTRs, SINEs, LINEs and DNA-transposons, and GC-content as explanatory variables, respectively.

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	3.94	0.20	19.43	$<2.00 \times 10^{-16}$
Chromosome length	-0.48	0.24	-1.98	4.88×10^{-02}
CDS	0.06	0.22	0.28	7.79×10^{-01}
SINE	1.59	0.48	3.35	9.75×10^{-04}
DNA	-0.18	0.34	-0.51	6.09×10^{-01}
LINE	-0.43	1.38	-0.31	7.57×10^{-01}
LTR	0.51	0.67	0.76	4.47×10^{-01}
Total repeats	-0.76	1.40	-0.54	5.87×10^{-01}
GC-content	0.68	0.35	1.94	5.43×10^{-02}

Residual standard error: 2.182 on 197 df

Multiple R-squared: 0.2389, Adjusted R-squared: 0.2079

F-statistic: 7.728 on 8 and 197 DF, p-value: 5.36×10^{-09}