Supplemental Note 1. Genome assembly and BUSCO analysis

The *P. napi* initial assembly used Illumina paired-end (PE) reads and mate-pair (MP) reads to generate scaffolds directly and was composed of 7,829 scaffolds with an N50-length of 300 kb and total length of 347 Mb. A second round of scaffolding was done using an added Illumina long mate-pair (LMP) library generated from a sibling of each individual used in the initial sequencing and assembly and SSPACE. This resulted in a *P. napi* assembly of 5860 scaffolds with an N50-length of 477 kb and total length of 350 Mb. The final scaffolding step used a third sibling of each species and a Chicago Illumina variable insert size library which brought the *P. napi* assembly to 3005 scaffolds with an N50-length of 4.2 Mb and a total length of 350 Mb.

We assessed our genome assembly and annotation completeness using Benchmarking Universal Single-Copy Orthologs (BUSCO) version is: 3.0.2, run in mode genome with the lineage dataset: insecta_odb9 (Creation date: 2016-02-13, number of species: 42, number of BUSCOs: 1658). Below are shown the output for the full genome assembly, followed by the results for the AllPaths scaffolds only (i.e. before scaffolding). This demonstrates that the full assembly has a couple more complete SCOs, which were brought together by our scaffold joins during chromosome level assembly.

C:94.5%[S:92.3%,D:2.2%],F:2.7%,M:2.8%,n:1658

1567 Complete BUSCOs (C)
1531 Complete and single-copy BUSCOs (S)
36 Complete and duplicated BUSCOs (D)
44 Fragmented BUSCOs (F)
47 Missing BUSCOs (M)
1658 Total BUSCO groups searched

C:94.3%[S:92.3%,D:2.0%],F:2.8%,M:2.9%,n:1658

1563 Complete BUSCOs (C)
1530 Complete and single-copy BUSCOs (S)
33 Complete and duplicated BUSCOs (D)
46 Fragmented BUSCOs (F)
49 Missing BUSCOs (M)
1658 Total BUSCO groups searched

Supplemental Note 2 (HiRise Scaffolding and misassembly correction)

The HiRise[1] scaffolding step provided a method to detect and break low confidence input scaffolds from SSPACE[2] and Allpaths-LG[3] using an independent set of reads. 90 scaffold mis-joins in 44 scaffolds were detected and broken before assembly continued.

Supplemental Note 3. Linkage group correction of misassemblies

113 scaffolds of the genome assembly contained multiple segregating markers which allowed for the manual identification of 60 misassemblies within scaffolds and merging of corrected scaffolds into chromosomes. K-mer estimation of genome size and heterozygosity rate with GenomeScope[4] was as follows:

k = 15

property	min	max
Heterozygosity	1.99681%	2.01325%
Genome Haploid Length	261,121,754 bp	261,578,934 bp
Genome Repeat Length	192,745,333 bp	193,082,797 bp
Genome Unique Length	68,376,421 bp	68,496,136 bp
Model Fit	81.9164%	95.0115%
Read Error Rate	0.0419424%	0.0419424%

Supplemental Note 4 (Functional Annotation Summary)

Annotation predicted 13,622 gene models (119,909 exons) derived from 20,325 mRNA transcripts respectively, with 9,346 genes annotated using a combination of KEGG[4], PFAM[5], InterPro[6], GO[7], MetaCyc[8], UniPathway[9], and Reactome[10] (AnnotationTable1-4.xls).

Information about Coding Genes:

Number of genes 13622

Number of mrnas 20325
Number of mrnas with utr both sides 9467
Number of mrnas with at least one utr 16893
Number of cdss 20325
Number of five_prime_utrs 13690
Number of three_prime_utrs 12670
Number of exons 119909
Number of introns 99584
Number of exon of cds 111587
Number of intron of cds 91262
Number of exon of five_prime_utr 17985
Number of intron of five_prime_utr 4295
Number of exon of three_prime_utr 16525
Number of intron of three_prime_utr 3855
mean mrnas per gene 1.5
mean cdss per mrna 1.0
mean five_prime_utrs per mrna 0.7
mean three_prime_utrs per mrna 0.6
mean exons per mrna 5.9
Total gene length 81632528
Total mrna length 122624718
Total cds length 20172833
Total five_prime_utr length 2661035
Total three_prime_utr length 7454905
Total exon length 30288280
mean gene length 5992
mean mrna length 6033
mean cds length 992
mean five_prime_utr length 194
mean three_prime_utr length 588
mean exon length 252
Longest genes 230888
Longest mrnas 230888
Longest cdss 38541

Longest five_prime_utrs 3377
Longest three_prime_utrs 8137
Longest exons 13013
Shortest genes 24
Shortest mrnas 24
Shortest cdss 6
Shortest five_prime_utrs 1
Shortest three_prime_utrs 1
Shortest exons 1

Functional inference for genes and transcripts was performed using the translated CDS features of each coding transcript. Each predicted protein sequences was blasted against the Uniprot/Swissprot[11] reference data set (downloaded 2014-05-15) in order to retrieve the gene name and the protein function as well as run against InterProscan version 5.7-48[12] in order to retrieve Interpro[6], PFAM[5], and GO[7] data. Outputs from both analyses have been parsed using the Annie annotation tool[13] to extract and reconcile relevant meta data into predictions for canonical protein names and functional predictions.

Database origin	Total Term Number	mRNA number referred by a term	Gene number referred by a term
PFAM	22931	17471	10105
Interpro	30268	17922	10337
GO	30590	12583	7243

Moreover, functional information was retrieved for 18413 transcripts (10624 genes), and 8060 transcripts (5283 genes) don’t have any functional information.

Supplemental Note 5. RBH of orthologs

The *B. mori* Geneset A from KAIKOBASE v3.2.2[14] was filtered for single copy orthologs (SCOs) as defined in orthoDB v9.1[15]. 3101 SCOs were then compared to the annotated protein set of *P. napi* and 2743 genes were called as orthologs by reciprocal best hit (RBH) in blastp between *B. mori* and *P. napi*. To achieve a higher resolution map of the synteny relationship between species the full gene set of *B. mori* was compared to *P. napi* and 8176 genes were identified by RBH as orthologs.

Supplemental Note 6. WGS and RNA based linkage map

A second linkage map for *P. napi* was constructed from 16 full-sib offspring and the parents of a cross between a mother from Abisko, Sweden and father from Barcelona, Spain. None of these individuals were used in either the original genome assembly or the first linkage map based on RAD-tag sequencing. Illumina RNAseq reads for 12 individual offspring and Illumina whole genome sequencing reads for 4 individual offspring and 2 parents were used to construct a WGS linkage map of 106,362 SNP markers. Maternally inherited markers confirmed linkage of scaffolds within chromosomes due to achiasmatic recombination.

See Supplemental Figure 2 for these results per chromosome.

Code used to prepare data:

samtools mpileup -r Chromosome_"$i" -A -gu -Q 15 -t DP -f Pieris_napi_repmask_fullAsm.fasta -b list_of_bamfiles.txt | bcftools call -cv --skip-variants indels - | vcftools --vcf - --maf .1 --max-missing .85 --min-meanDP 6 --minDP 3 --recode --out consensus_Chr"$i"

Supplemental Note 7. Mate Pair Spanning

35,492,386 read pairs from the 3 kb insert MP library, 39,072,074 read pairs from the 7 kb library, and 2,648,844 read pairs from the 40 kb library mapped to the chromosomes with a mapq > 20, in the proper orientation, and with the expected insert size. Every base pair position of the assembly was assigned a count of 0 for each mate pair library. Using a custom awk script, each read pair was taken in turn and the counter of every base pair of the assembly between the reads was incremented by one for that library. After all reads were processed each base pair position had a count of the number of reads that spanned that position for each library. The number of total mate pair spans used to diagnose potential misassemblies was the sum of the 3kb, 7kb, and 40kb library counts.

Mate pair spanning of a position is expected to fall towards zero if there is a misassembly or the region around that position contains low complexity or repeat rich sequence that decreases the mapq below the threshold of mapq <= 20.

See Supplemental Figure 2 for these results per chromosome.

Supplemental Note 8. Alignment between *P. napi* and *P. rapae*

The 75 largest scaffolds of *P. napi* and 101 largest scaffolds of *P. rapae* containing 90% of the content of each genome were aligned with each other using last aligner v. 714.

```
lastal -m10 -r5 -q97 -a0 -b97 -C2 -l100 PrapaeN90 " \
< Pieris_napi_chromosomes.fa > PnapiChr_PrapaeN90.maf
```

And visualized in dotplot format

```
/data/programs/last-714/scripts/last-dotplot -s 6 -x 5000 -y 5000 Pnapi_chm_N90.maf
Pnapi_chm_N90.png
```

Visual inspection of alignments between scaffolds showed 47 places where two *P. napi* scaffolds which were collinear on the linkage map were also collinearly aligned to a single *P. rapae* scaffold.

Supplemental Note 9. Syntenic block support of scaffold joins

Blocks of synteny were defined from 7109 of the genes identified as reciprocal best blast hits between *B. mori* geneset A and *P. napi* final annotation that occurred on the chromosomes of the *P. napi* assembly. Adjacent blocks were merged by recursively removing genes that shared no neighbors from the same *B. mori* chromosome, then blocks of size 2, 3, 4, and 5. This reduced the number of genes in syntenic blocks to 6,839, and the number of blocks from 499 to 99. While it is possible that a single gene or group of 5 genes translocated from a different chromosome or existed as a small fragment prior to the ancestral fusion events, it seemed more likely that misidentification in the reciprocal best blast hit of these non-single copy orthologs was responsible for the attribution of these small blocks of synteny.

See Supplemental Figure 2 for these results per chromosome.

Supplemental note 10: SNP generation from nextRAD library sequence.

Use trimmomatic to trim the adapters from the reads

```
system("java -jar Trimmomatic-0.32/trimmomatic-0.32.jar SE $file $outfile ILLUMINACLIP:Trimmomatic-0.32/adapters/NexteraPE-PE.fa:1:30:5 MINLEN:75 HEADCROP:$start_trim CROP:$trim");
```

Read in the sequence files with a Perl script and track the number of occurrences of each read. Keep all the reads with counts above a threshold. This is the first set of loci.
Call the reads with a count above a threshold “repetitive” and create a fasta file of the repeats. Align the first set of loci to the repeat set allowing many mismatches (we use bwa-mem or BBMap). Remove all loci that align to the repeats. The remaining loci are the second set of loci.

```
system("bwa mem -t 6 -B 2 -O 3 -k 13 -a $repeat_fasta_file $fastq_file > $repeat_align_file_mt");
```

Align the second set of loci to itself. Reads that map to each other are likely alleles at the same locus. Pick one and remove the others. The winnowed set of loci are the reference set.

```
system("java -ea -Xmx31g -cp align2.BBMapPacBioSkimmer in=" . $fastq_file . " out=" . $align_file . " minid=" . $align_threshold . " maxsites=150 ambig=all secondary=true ow sssr=" . $sssr . " slow idtag pambig=f maxindel=15 nodisk k=9 minhits=1 expectedsites=40");
```

The above is the “rad” specific set of steps. We now try to use a general pipeline for SNP calling.
Align all the reads to the reference set (bwa-mem or BBMap)

```
$t = "java -ea -Xmx31g -cp align2.BBMap in=" . $gf . " ref=" . $ref_fasta_file . " nodisk out=stdout minid=" . $align_threshold . " ow slow k=9 idtag maxindel=40 | samtools view -bSu - | samtools sort -@ 8 - " . $sort_file;
```

Sort the resulting bam files

```
samtools view -bSu - | samtools sort -@ 8 - " . $sort_file;
```

Use samtools and bcftools to create a vcf genotype table

```
samtools mpileup -gu -Q 15 -t DP -f ref_2014_txt.fasta -b ref_txt.align_samples | bcftools call -cv - > 2014.vcf
```

Then filter the vcf on minimum read depth and allele frequency

```
# vcftools --vcf 2014.vcf --maf .1 --max-missing .85 --min-meanDP 6 --minDP 3 --recode
```

Supplemental Table 1. Read data used in assembly and analysis.

Library	Mean insert size	Source	# raw read pairs	# clean read pairs	Coverage
Illumina 180 bp PE	-20bp overlap	Female pupae #1	350 M	348 M	180x
Illumina 3 kb MP	3600bp	Female pupae #1	242 M	61 M	34x
Illumina 7 kb MP	6500bp	Female pupae #1	264 M	70 M	40x
Illumina 40kb MP	40kb	Unsexed pupae #2	8.9 M	8.9 M	12x
Chicago	variable	Unsexed pupae #3	13 M	13 M	19x
Total				500 M	285 X

Supplemental Table 2. GO term enrichment in syntenic blocks

A GO term was considered enriched within a syntenic block if at least 3 genes were assigned that GO term and it was overrepresented compared to the rest of the genome by a fisher exact test with a p-value < 0.01. The GO.ID and Term identify the GO category analyzed, the number of total genes Annotated within that category, and the number of Significant genes within a particular syntenic block are tabulated along with the Fisher exact test significance (classicFisher) of that category's enrichment. The syntenic block id (sbid) and that blocks genomic coordinates indicate the region containing that significant GO term.

Table of results: GOenrichment.tsv

Supplemental Table 3: Lepidopteran phylogeny

Node ages in millions of years ago (MA). Node numbers correspond to Supplemental Fig. 8. Reference for each node age in that figure is provided below.

Node	Age (MA)	Reference
25	140.47	[19]
26	116.74	[19]
27	109.78	[19]
28	94.05	[19]
29	83.85	[19]
30	90.70	[19]
31	28.24	[19,21]
32	92.74	[19]
33	110.87	[16]
34	30.92	[17]
35	24.79	[17]
36		
37	109.46	[16]
38	104.73	[16]
39	86.18	[18]
40	75.39	[18]
41	15.79	[18]
42	101.41	[16]
43	89.79	[19]
44	85.35	[19]
45	78.55	[19]
46	10.75	[20]

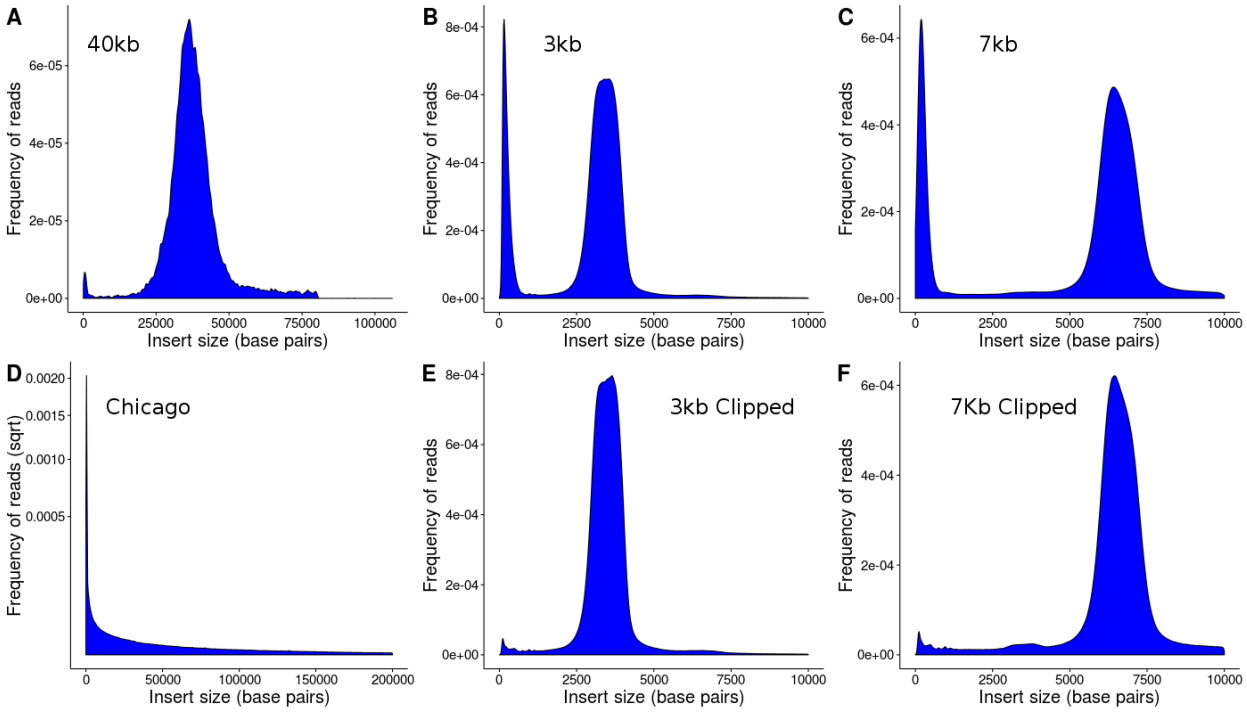
Notes:

1) **Node 31.** Neither *Helicoverpa armigera* nor *Spodoptera frugiperda*, both of which belong to the so-called noctuid pest clade, have been included in timing of divergence analyses. We provide an estimate for the age of their divergence by using their phylogenetic placement relative to other noctuids[21] together with the first divergence within the pest clade [19].
2) **Node 36.** Available timing of divergence analyses do not have the same topology as the one inferred here with respect to species of *Papilio*.[20] resolved the relationships as (*P. polytes* (*P. machaon*, *P. xuthus*)) and dated the split between *P. machaon* and *xuthus* at 19.56 MA.

Supplemental Table 4: Lepidopteran data sources

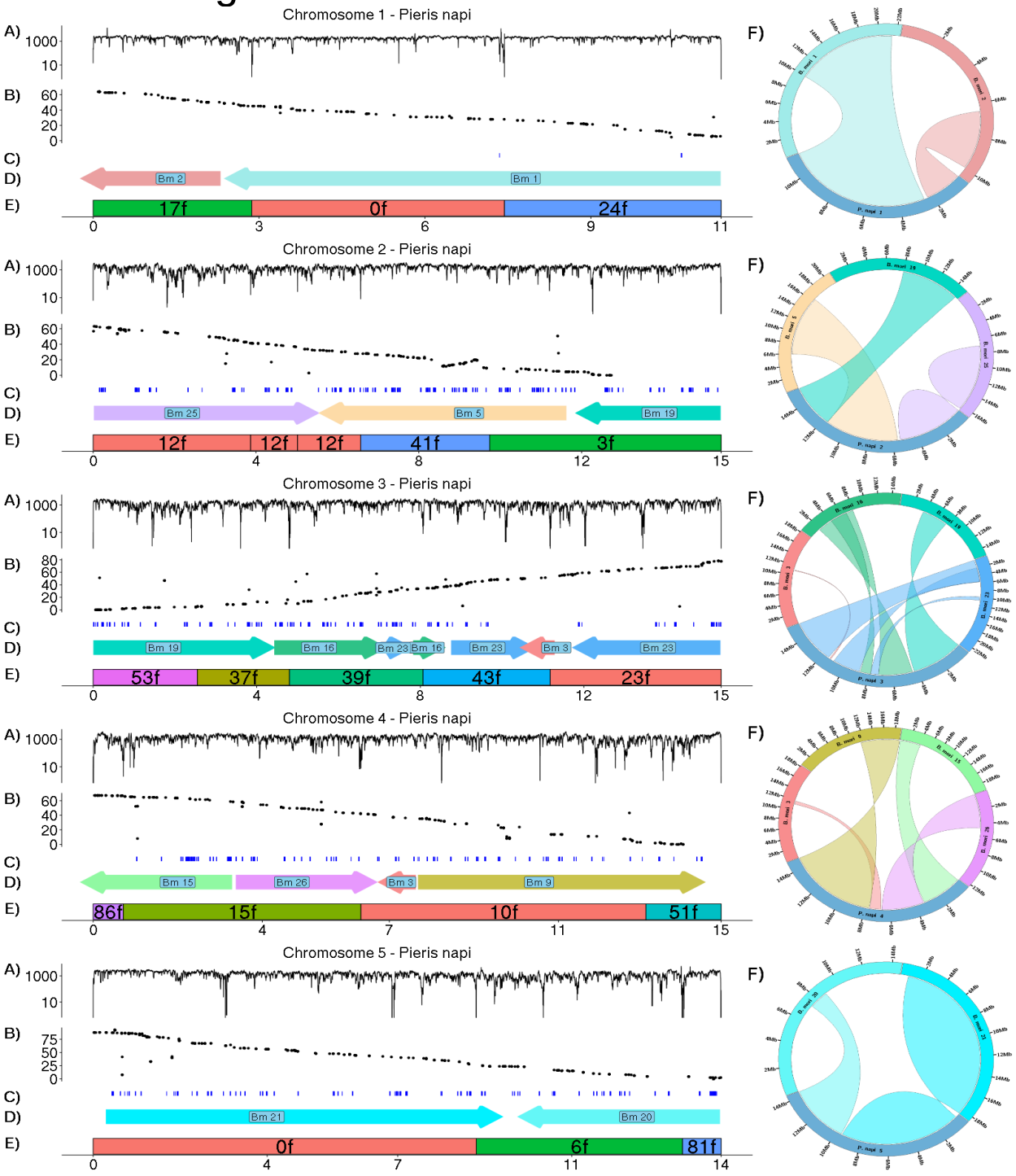
Genomes and gene annotations for 24 genome assemblies representing 23 species of Lepidoptera were retrieved from various locations and databases. In this table, their various sources and details are listed.

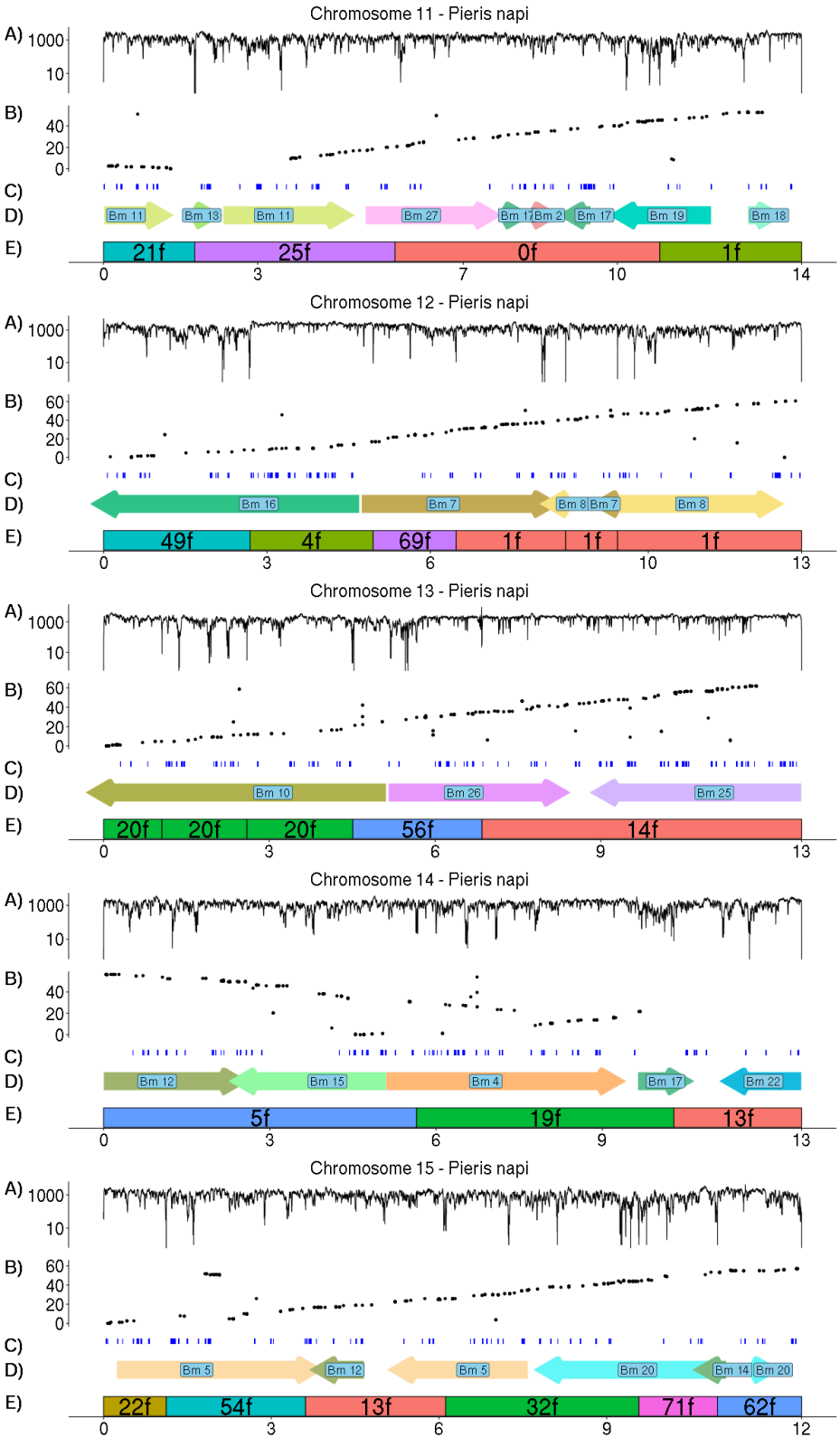
Supplementary Figure 1: Insert Size Distribution

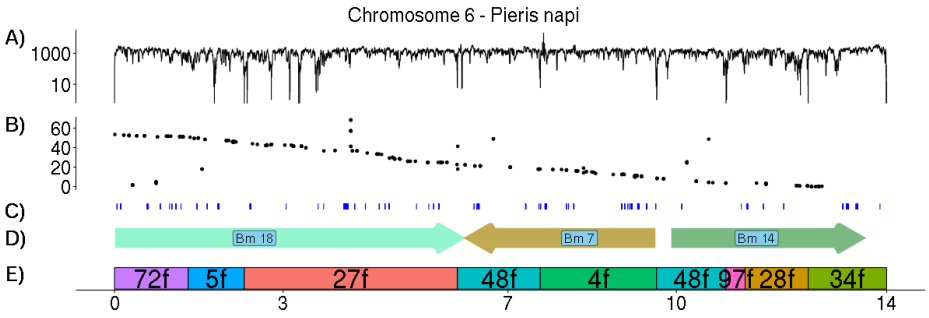


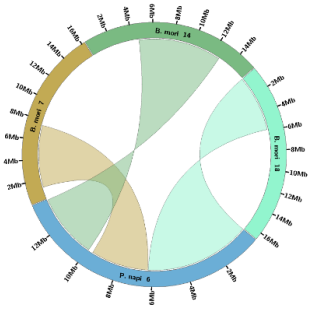
Insert size distribution of mate pair libraries used in genome construction and assessment. Insert sizes determined after mapping libraries back to final assembly with bbmap[22] after libraries were quality and adapter trimmed with bbduk. A) 40kb library, B) 3kb library, C) 7kb library, D) Chicago library, E) 3kb library after filtering out paired end reads that were included as an unwanted byproduct of library construction, and F) 7kb library after nextclip[23] filtering.

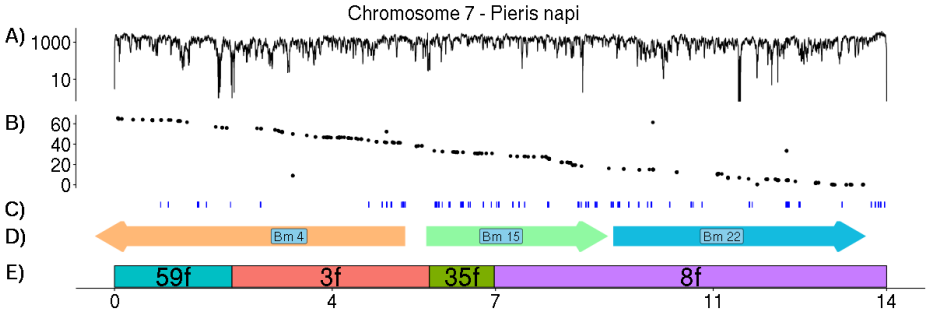
Supplementary Figure 2. Chromosomal assembly validation figures

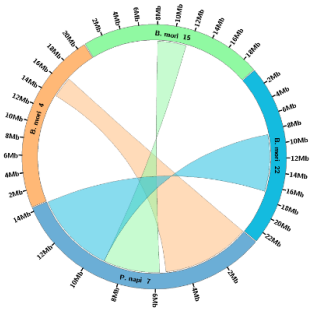


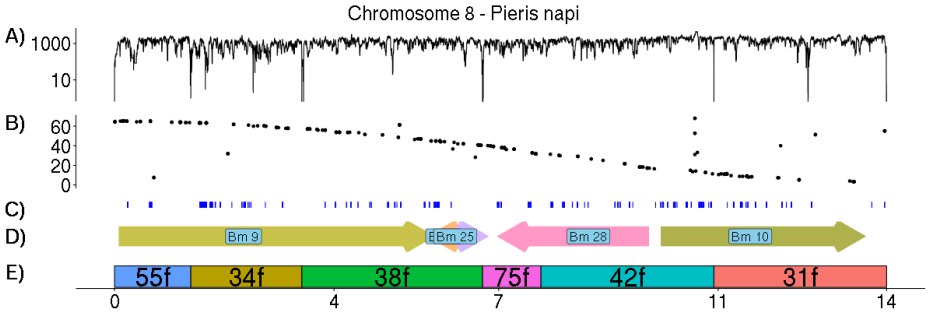

Chromosome 11 - Pieris napi
A)
1000
10
B)
40
20
0
C)
D)
Bm 11
Bm 13
Bm 11
Bm 27
Bm 17
Bm 2
Bm 17
Bm 19
Bm 18
E)
21f
25f
0f
1f
0
3
7
10
14
Chromosome 12 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 16
Bm 7
Bm 8
Bm 7
Bm 8
E)
49f
4f
69f
1f
1f
1f
0
3
6
10
13
Chromosome 13 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 10
Bm 26
Bm 25
E)
20f
20f
20f
56f
14f
0
3
6
9
13
Chromosome 14 - Pieris napi
A)
1000
10
B)
40
20
0
C)
D)
Bm 12
Bm 15
Bm 4
Bm 17
Bm 22
E)
5f
19f
13f
0
3
6
9
13
Chromosome 15 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 5
Bm 12
Bm 5
Bm 20
Bm 14
Bm 20
E)
22f
54f
13f
32f
71f
62f
0
3
6
9
12

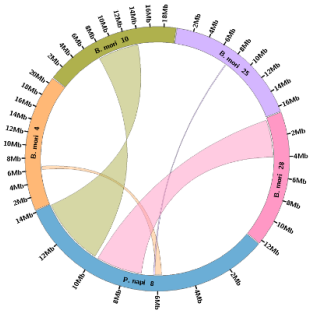

Chromosome 6 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 18
Bm 7
Bm 14
E)
72f
5f
27f
48f
4f
48f
97f
28f
34f
0
3
7
10
14

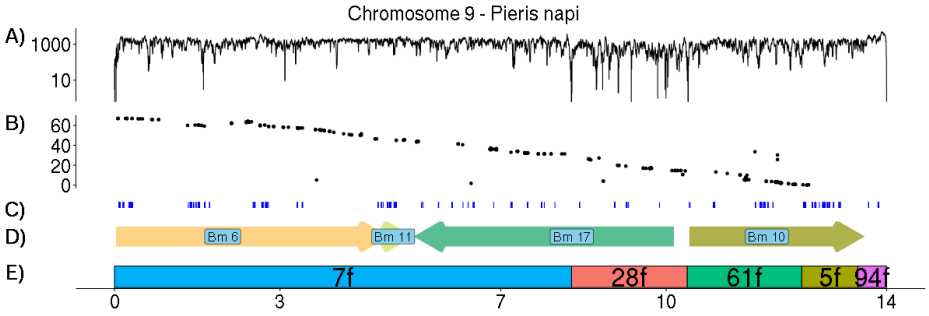


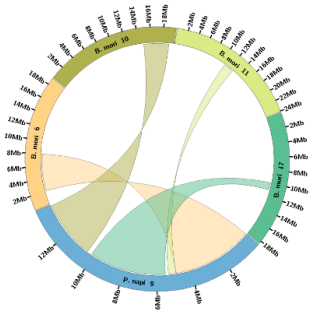

Chromosome 7 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 4
Bm 15
Bm 22
E)
59f
3f
35f
8f
0
4
7
11
14

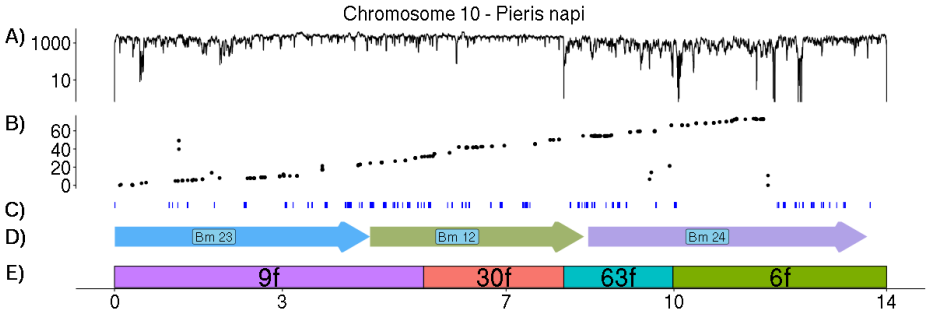


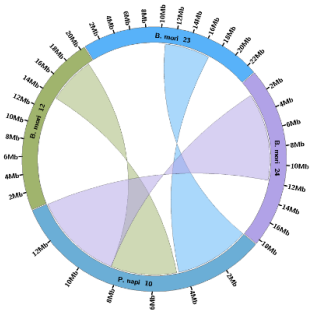

Chromosome 8 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 9
EBm 25
Bm 28
Bm 10
E)
55f
34f
38f
75f
42f
31f
0
4
7
11
14

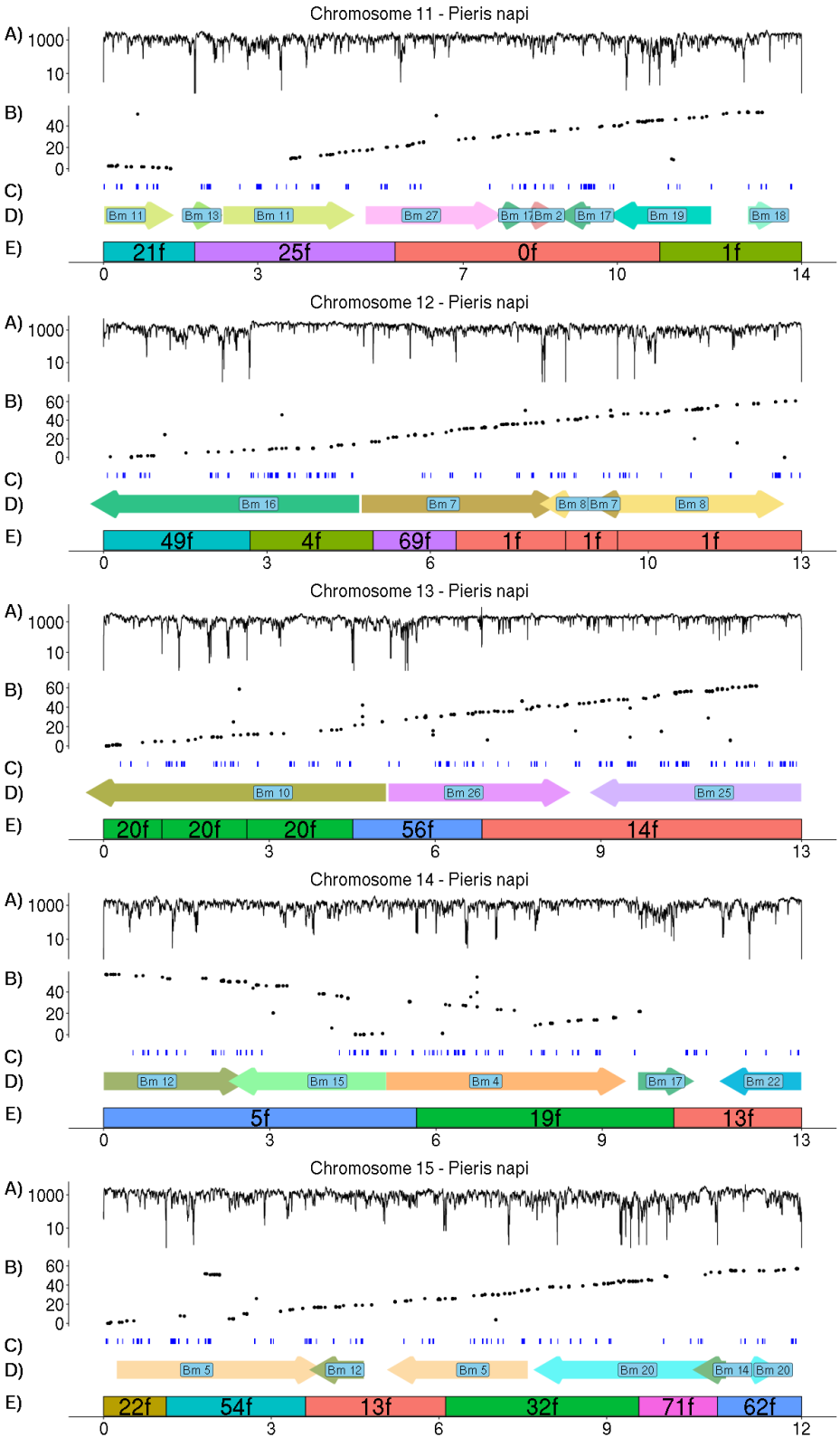


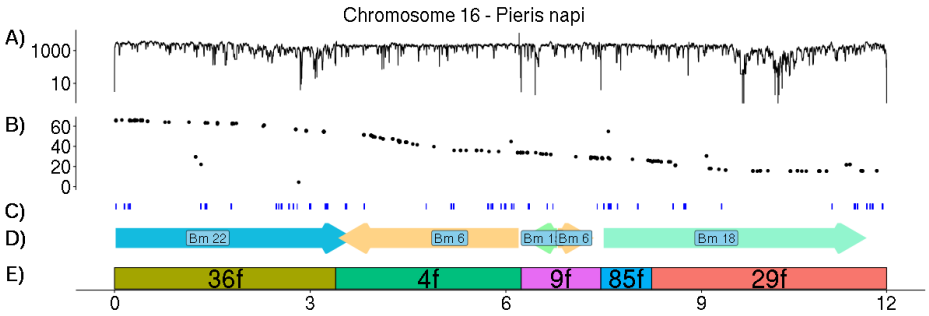

Chromosome 9 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 6
Bm 11
Bm 17
Bm 10
E)
7f
28f
61f
5f
94f
0
3
7
10
14

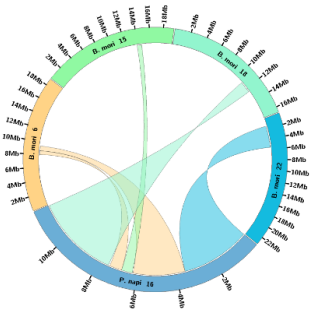


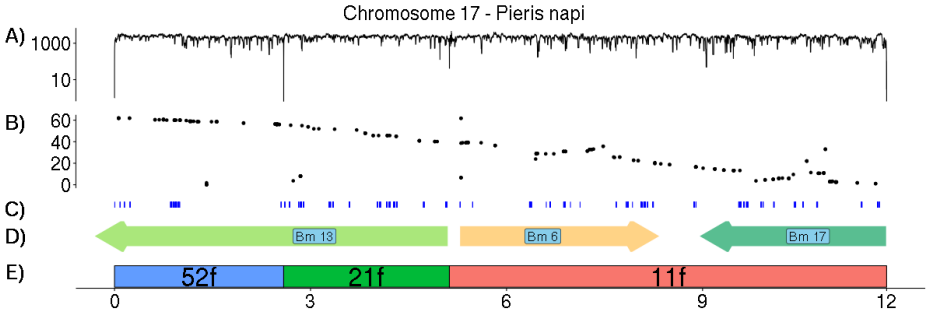

Chromosome 10 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 23
Bm 12
Bm 24
E)
9f
30f
63f
6f
0
3
7
10
14

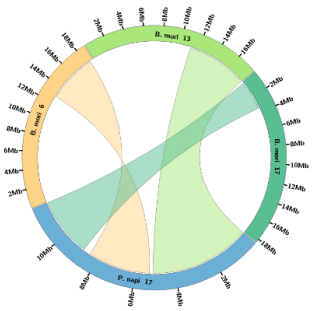


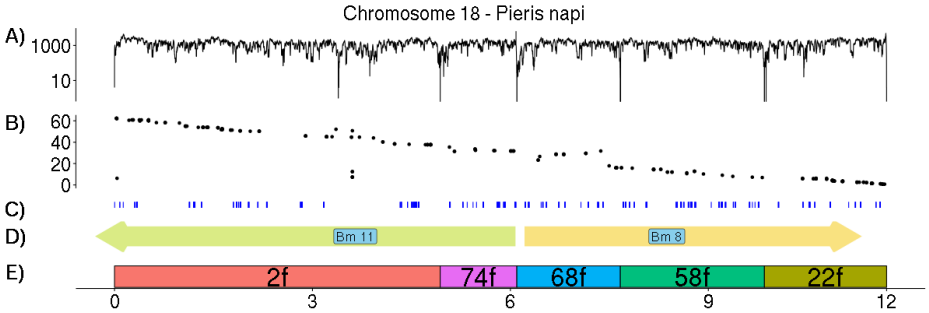

Chromosome 11 - Pieris napi
A)
1000
10
B)
40
20
0
C)
D)
Bm 11
Bm 13
Bm 11
Bm 27
Bm 17
Bm 2
Bm 17
Bm 19
Bm 18
E)
21f
25f
0f
1f
0
3
7
10
14
Chromosome 12 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 16
Bm 7
Bm 8
Bm 7
Bm 8
E)
49f
4f
69f
1f
1f
1f
0
3
6
10
13
Chromosome 13 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 10
Bm 26
Bm 25
E)
20f
20f
20f
56f
14f
0
3
6
9
13
Chromosome 14 - Pieris napi
A)
1000
10
B)
40
20
0
C)
D)
Bm 12
Bm 15
Bm 4
Bm 17
Bm 22
E)
5f
19f
13f
0
3
6
9
13
Chromosome 15 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 5
Bm 12
Bm 5
Bm 20
Bm 14
Bm 20
E)
22f
54f
13f
32f
71f
62f
0
3
6
9
12

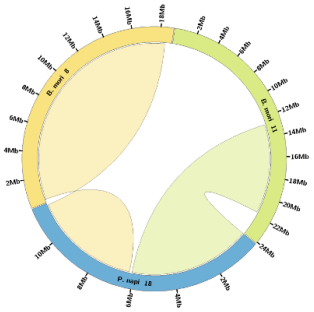

Chromosome 16 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 22
Bm 6
Bm 1 Bm 6
Bm 18
E)
36f
4f
9f
85f
29f
0
3
6
9
12

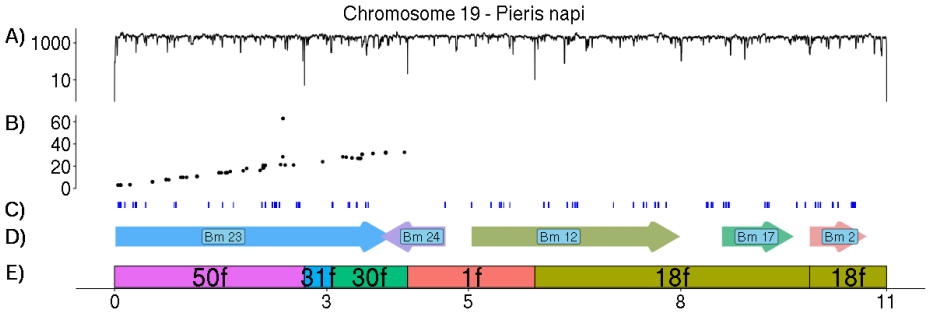


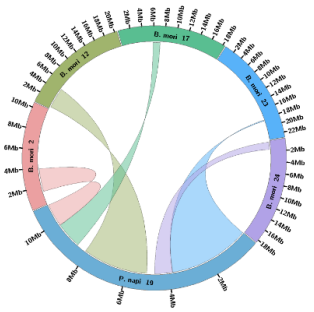

Chromosome 17 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 13
Bm 6
Bm 17
E)
52f
21f
11f
0
3
6
9
12

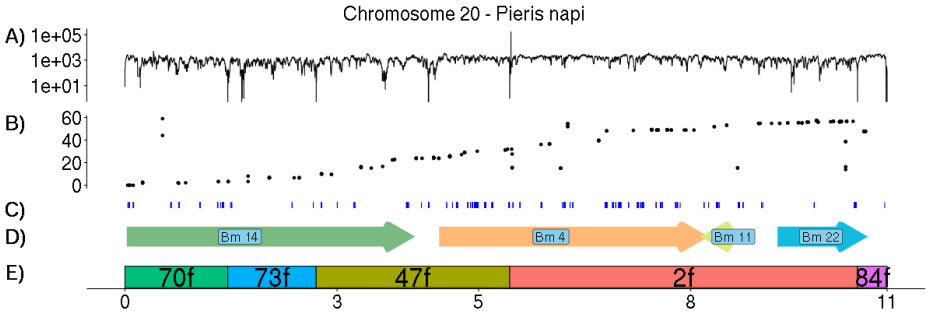


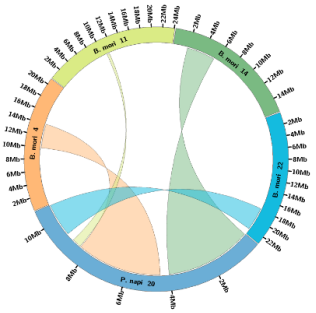

Chromosome 18 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 11
Bm 8
E)
2f
74f
68f
58f
22f
0
3
6
9
12

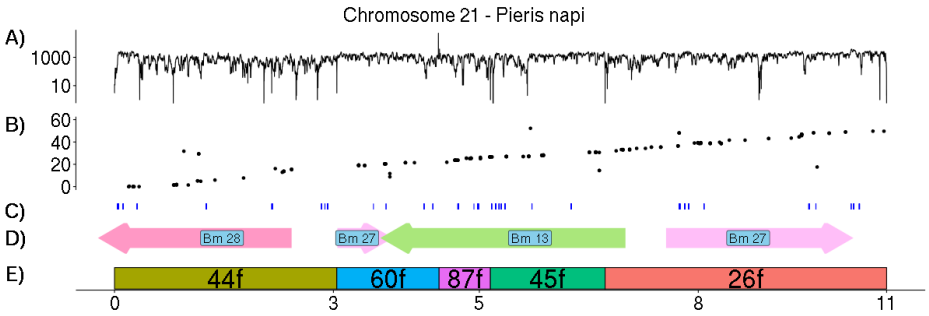


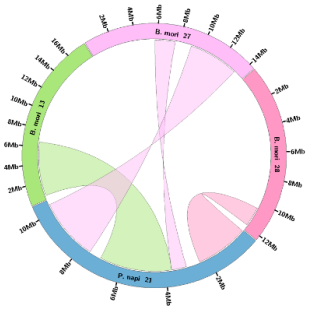

Chromosome 19 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 23
Bm 24
Bm 12
Bm 17
Bm 2
E)
50f
31f
30f
1f
18f
18f
0
3
5
8
11

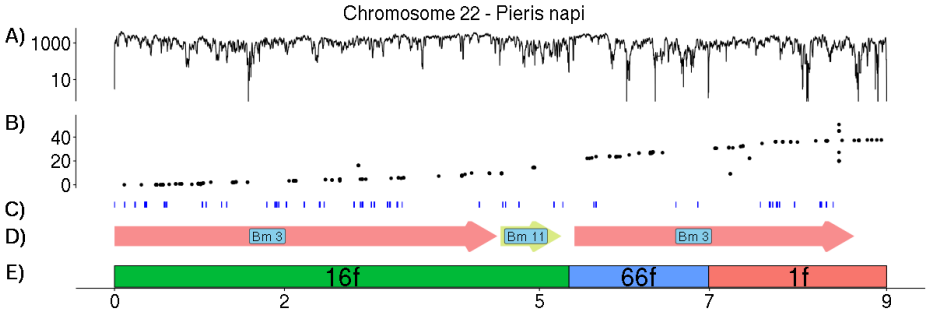


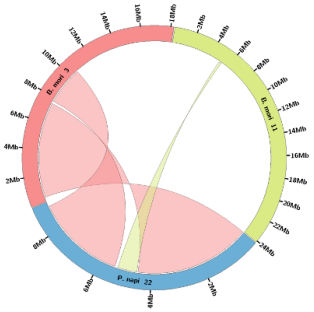

Chromosome 20 - Pieris napi
A)
1e+05
1e+03
1e+01
B)
60
40
20
0
C)
D)
Bm 14
Bm 4
Bm 11
Bm 22
E)
70f
73f
47f
2f
84f
0
3
5
8
11

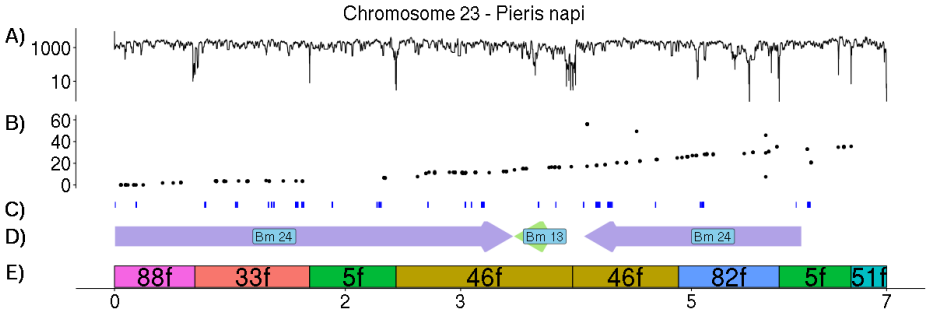


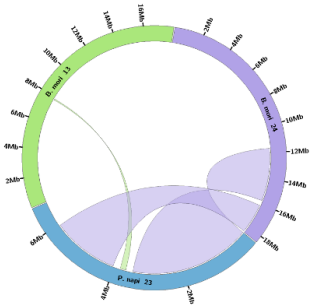

Chromosome 21 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 28
Bm 27
Bm 13
Bm 27
E)
44f
60f
87f
45f
26f
0
3
5
8
11

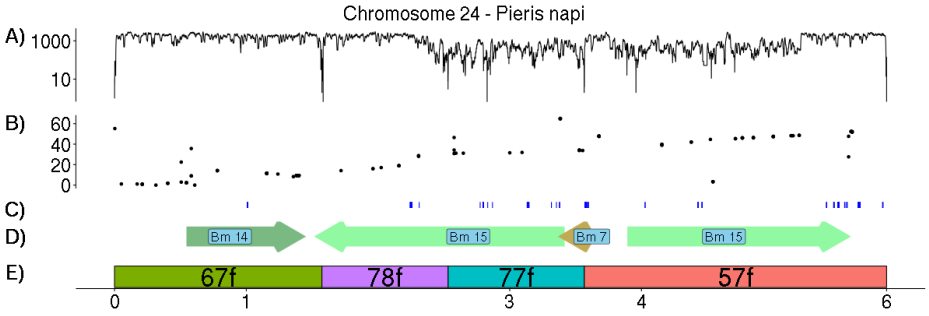


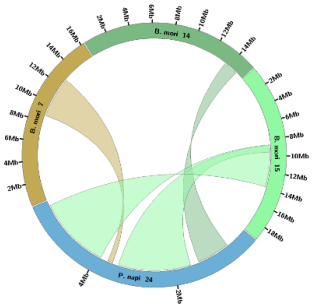

Chromosome 22 - Pieris napi
A)
1000
10
B)
40
20
0
C)
D)
Bm 3
Bm 11
Bm 3
E)
16f
66f
1f
0
2
5
7
9

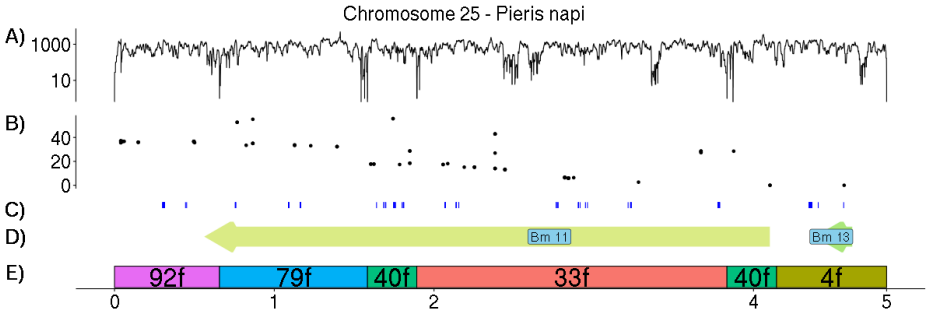


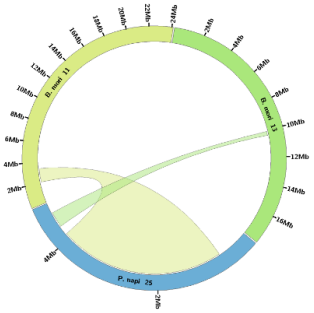

Chromosome 23 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 24
Bm 13
Bm 24
E)
88f
33f
5f
46f
46f
82f
5f
51f
0
2
3
5
7




Chromosome 24 - Pieris napi
A)
1000
10
B)
60
40
20
0
C)
D)
Bm 14
Bm 15
Bm 7
Bm 15
E)
67f
78f
77f
57f
0
1
3
4
6

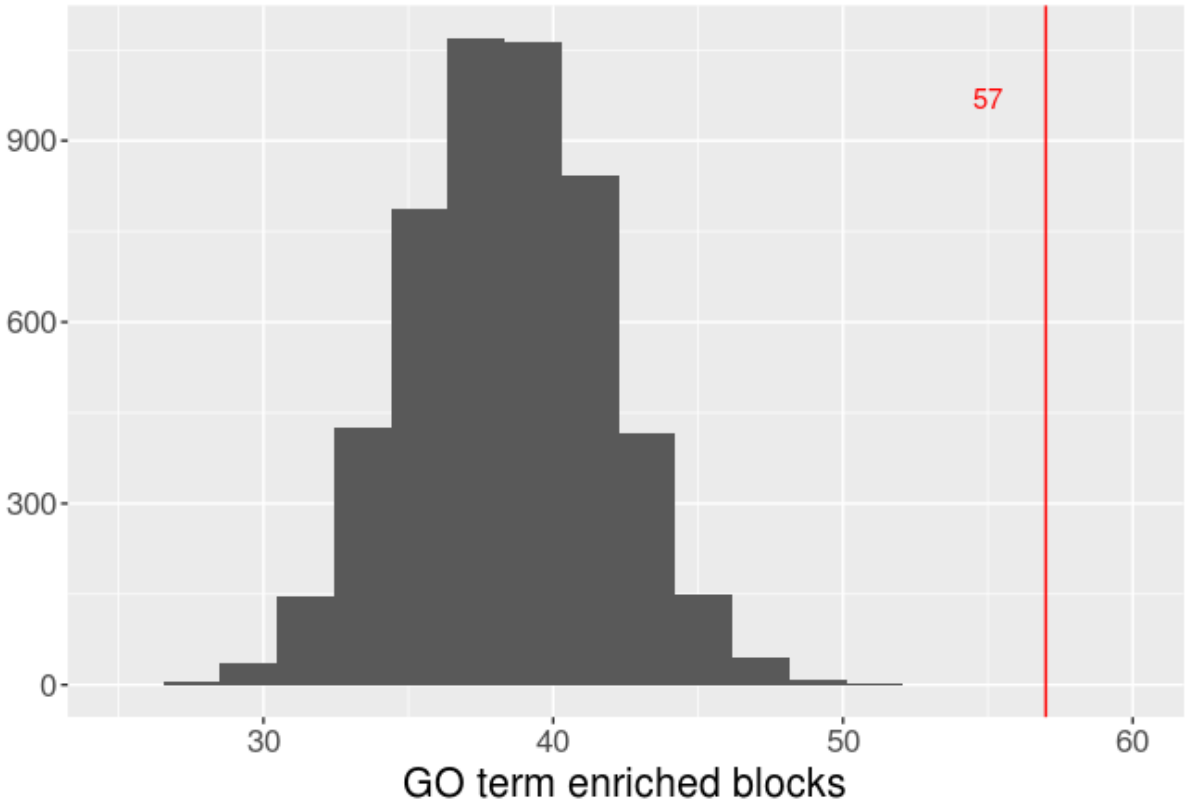



Chromosome 25 - Pieris napi
A)
1000
10
B)
40
20
0
C)
D)
Bm 11
Bm 13
E)
92f
79f
40f
33f
40f
4f
0
1
2
4
5



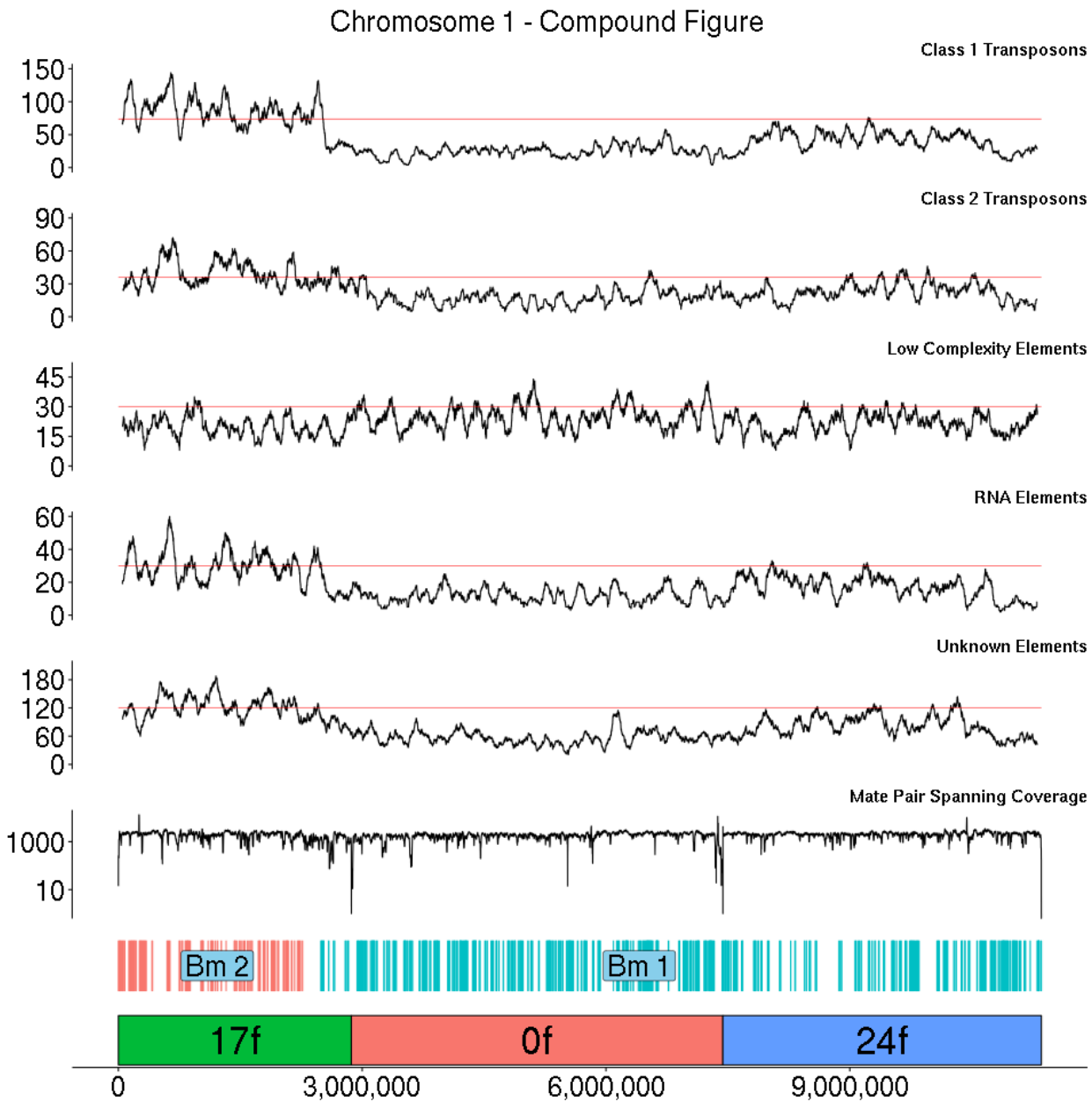
As in Fig. 2 of main text. Validation of the largest four *P. napi* chromosomes. Within each, **a)** mate pair spanning depth is shown across each chromosome, summed from the 3kb, 7kb, and 40kb libraries (genome averaged = 1356). Of the scaffold join positions 74 of 97 were spanned by > 50 properly paired reads (mean = 117.8, S.D. = 298.7), while the remaining 23 scaffold joins had 0 mate pair spans. **b)** black dots represent RAD-seq linkage markers and their recombination distance along chromosomes from the first linkage map **c)** Results from the second linkage map of maternally inherited markers (RNA-seq and whole genome data), where all markers within a chromosome are completely linked due to suppressed recombination in females (i.e. recombination distance is not shown on Y axis). **d)** *B. mori* collinear blocks, colored and labeled by their chromosomal origin, along with orientation by arrow, as in Fig. 1a. **e)** *P. napi* scaffolds comprising each chromosome, labeled to indicate scaffold number and orientation. **f)** To the right of each *P. napi* chromosome is a circos plot showing the location and orientation of the collinear blocks from each *B. mori* donor chromosome that comprise a given *P. napi* chromosome, colored as in Fig. 1a. A twist in the ribbon indicates a reversal of the 5' to 3' orientation of the *B. mori* relative to the *P. napi* chromsomes. Ribbon width on the *P. napi* chromosome is relative to the size of the collinear block. Remaining chromosomes shown in Supplementary Fig. 2.

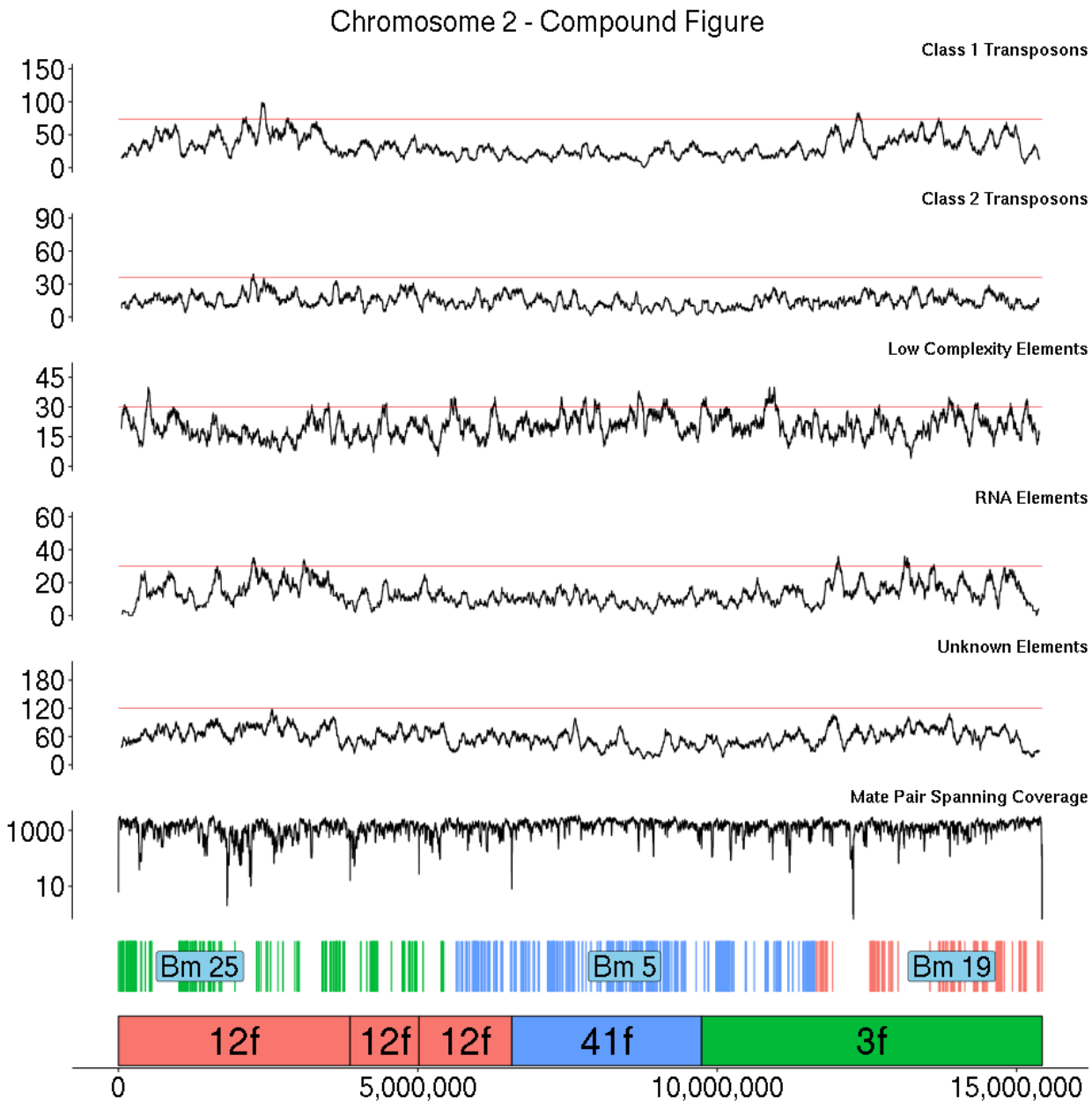
Supplementary Figure 3: GO term enrichment in syntenic blocks

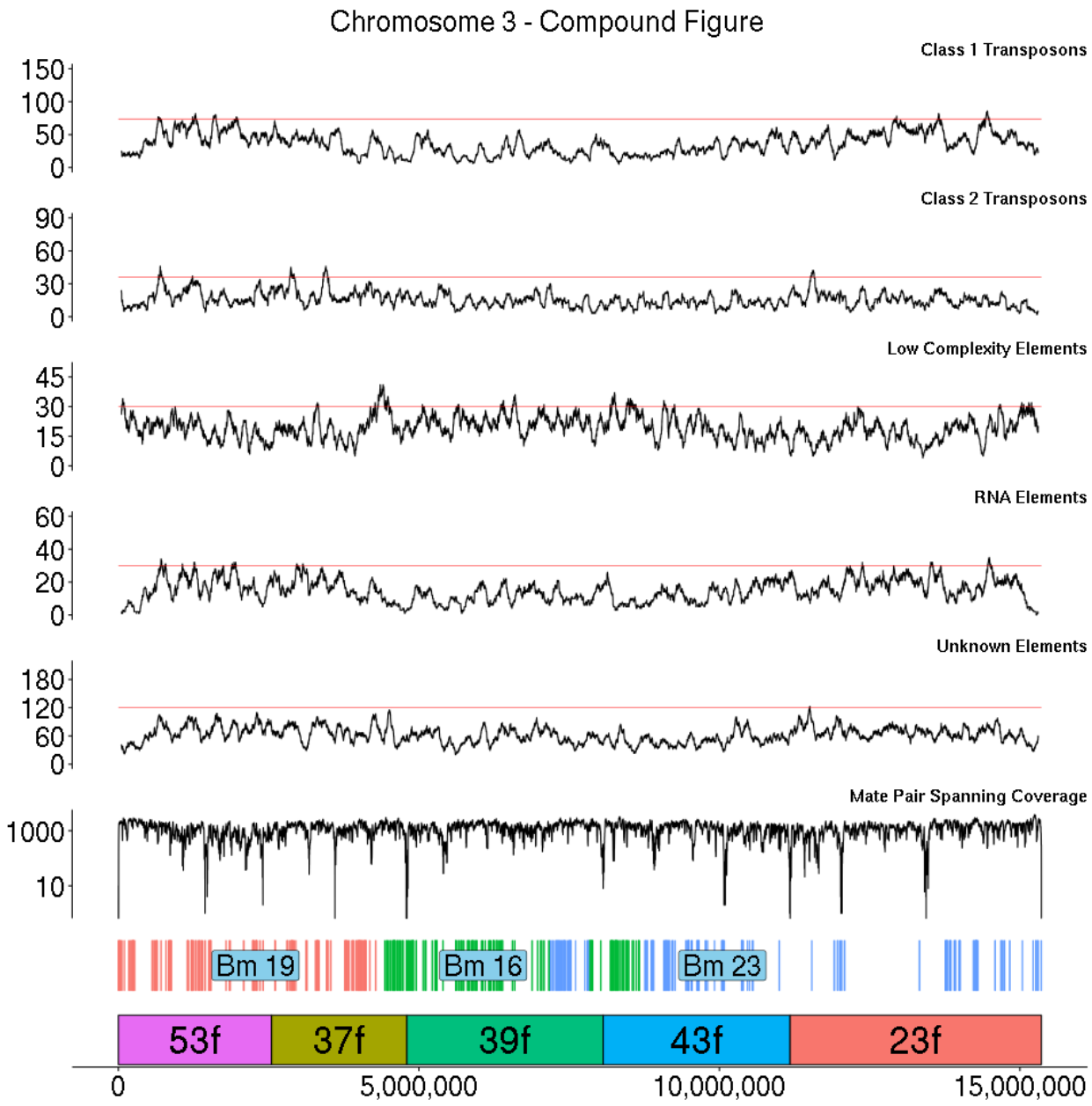


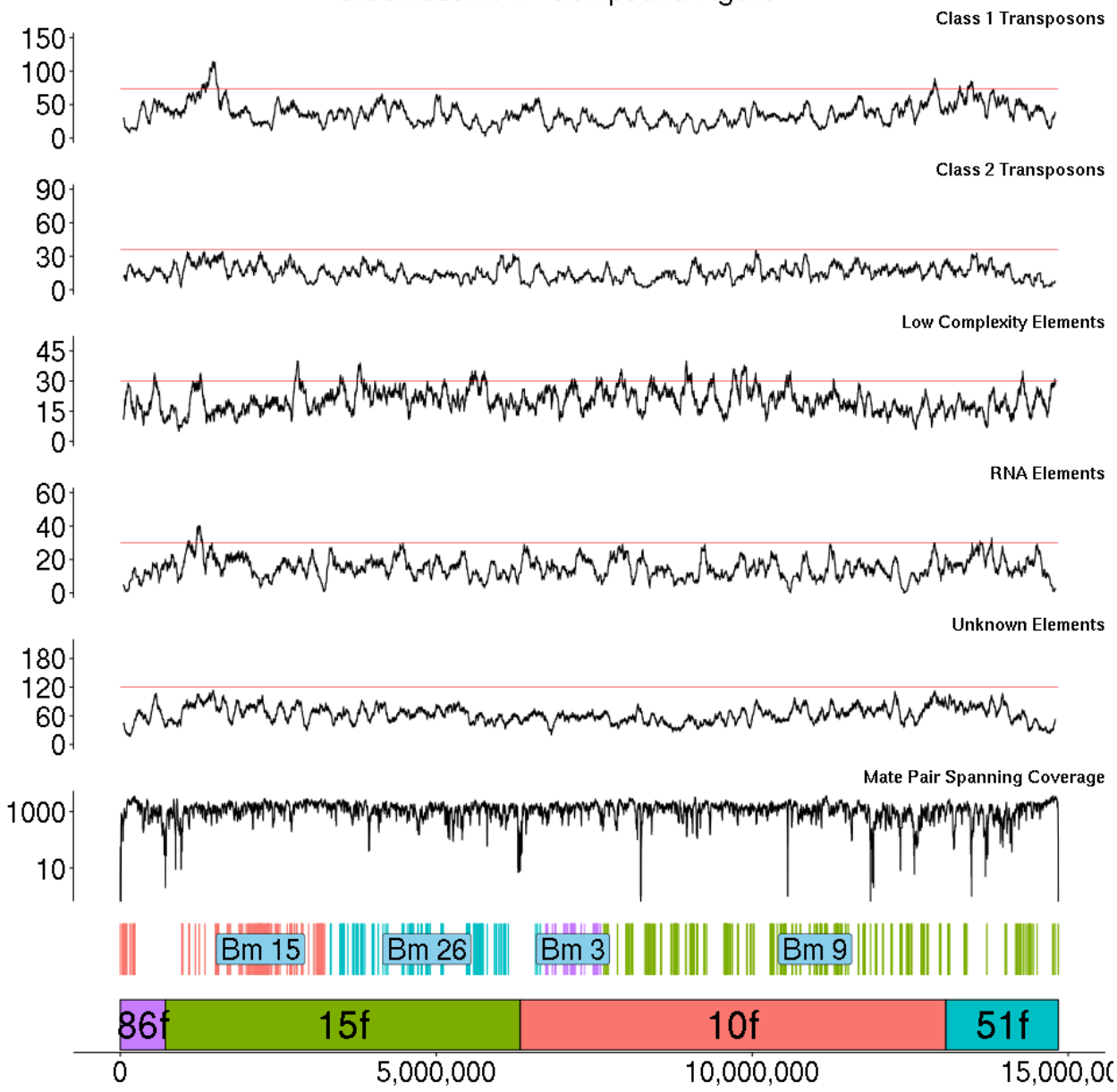
Distribution of number of syntenic blocks with GO term enrichment in 10,000 simulated genomes. Simulated syntenic blocks were constructed by breaking the *P. napi* genome into blocks of the same size as observed but in a random order. The mean number of GO enriched fragments in each of the simulated 10,000 genomes was 38.8 with a variance of 46.6 and maximum of 52. This is significantly lower than the observed 57 enriched regions is *P. napi* ($p < 0.0001$).

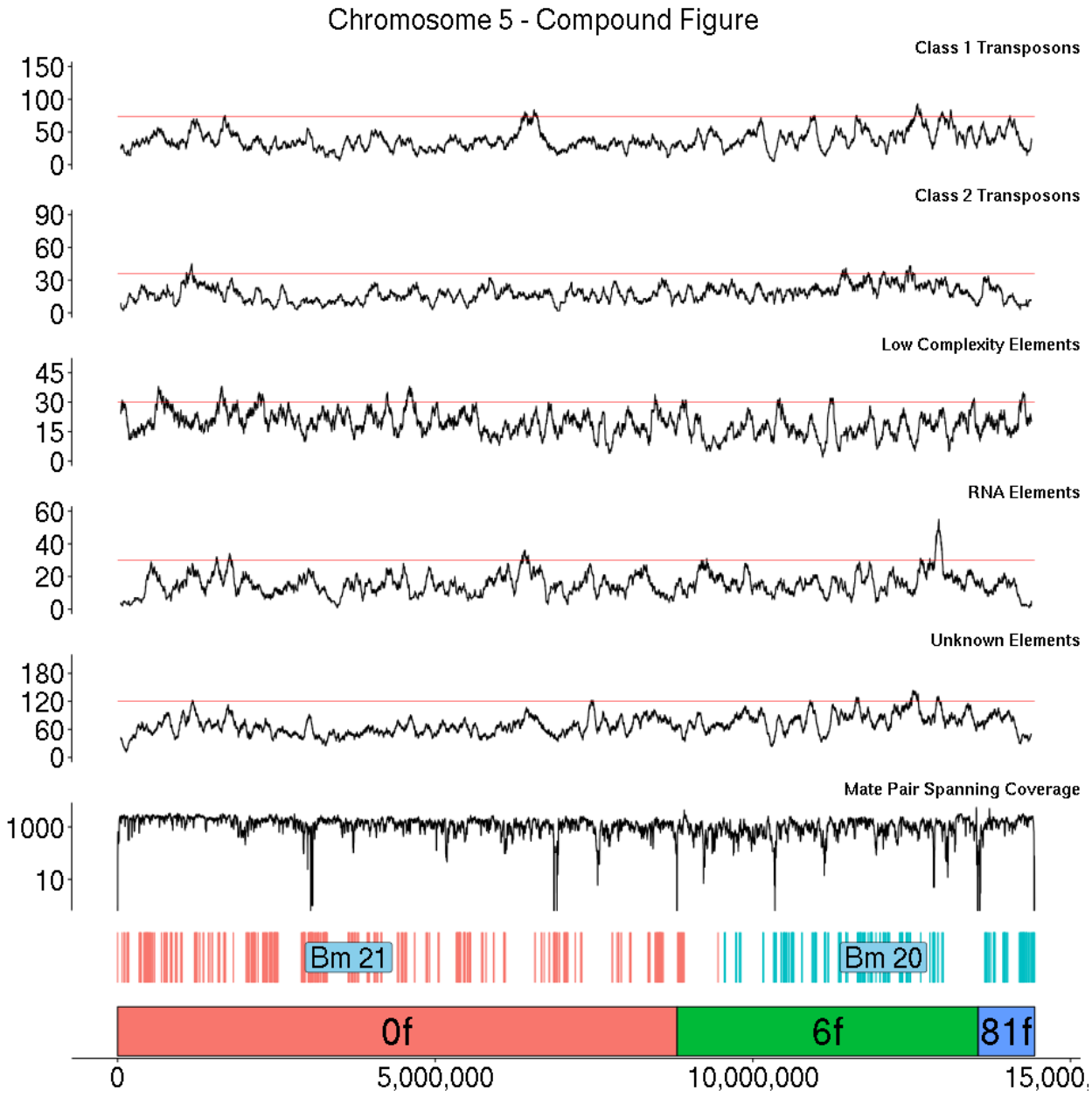
Supplementary Figure 4: Repetitive element distribution across chromosomes

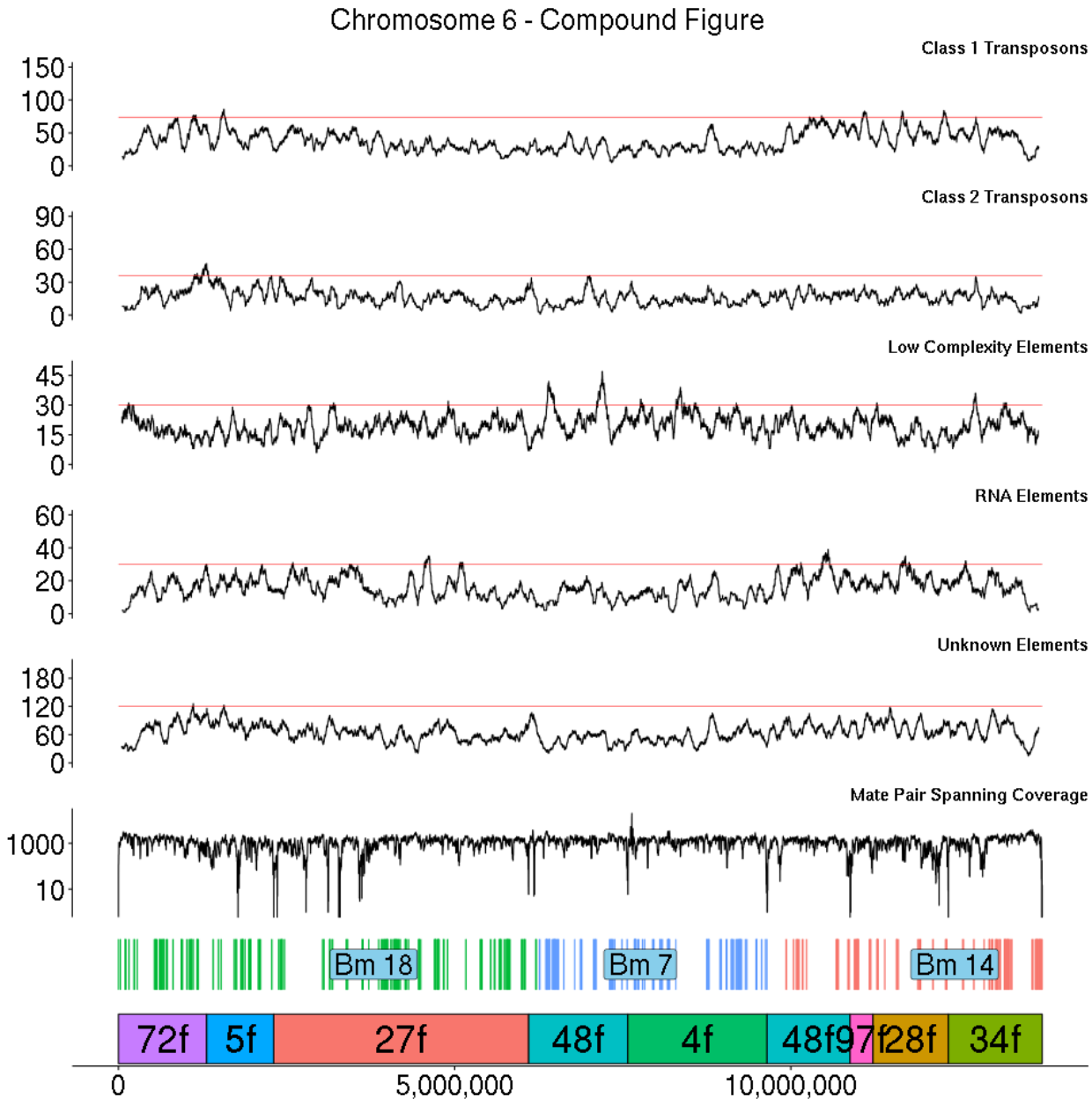


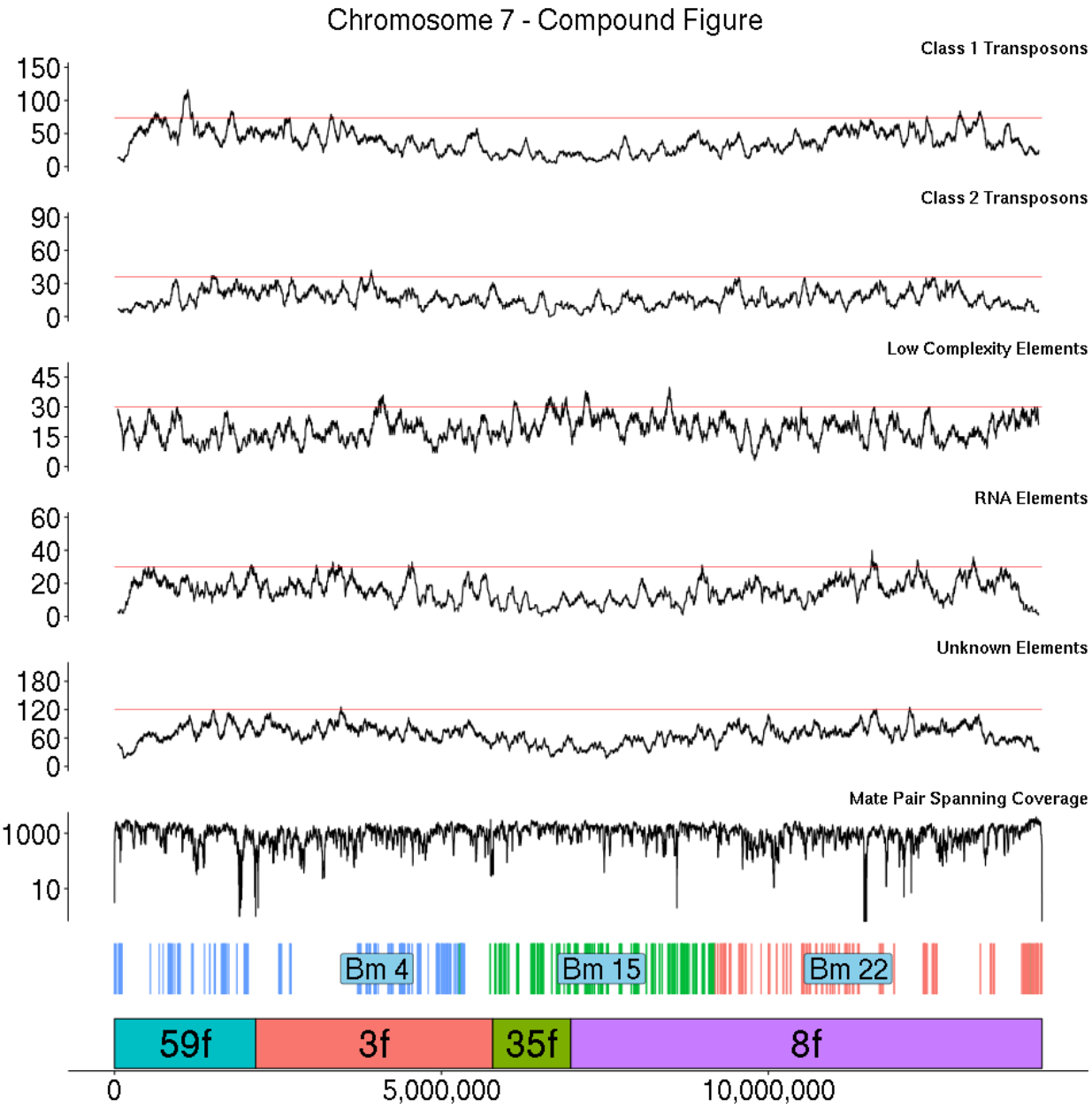

Chromosome 2 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 25
Bm 5
Bm 19
12f
12f
12f
41f
3f
0
5,000,000
10,000,000
15,000,000

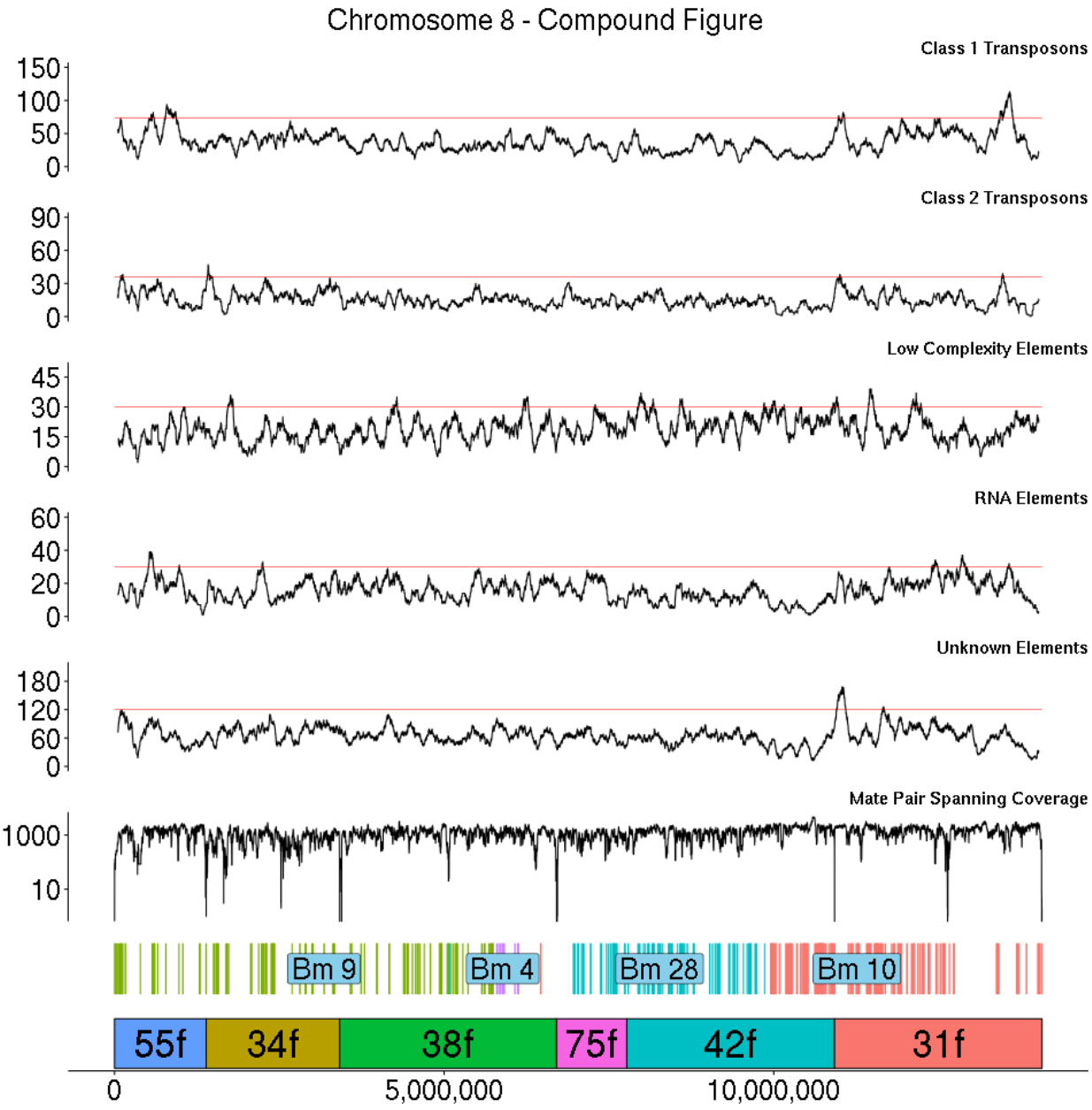

Chromosome 3 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 19
Bm 16
Bm 23
53f
37f
39f
43f
23f
0
5,000,000
10,000,000
15,000,000

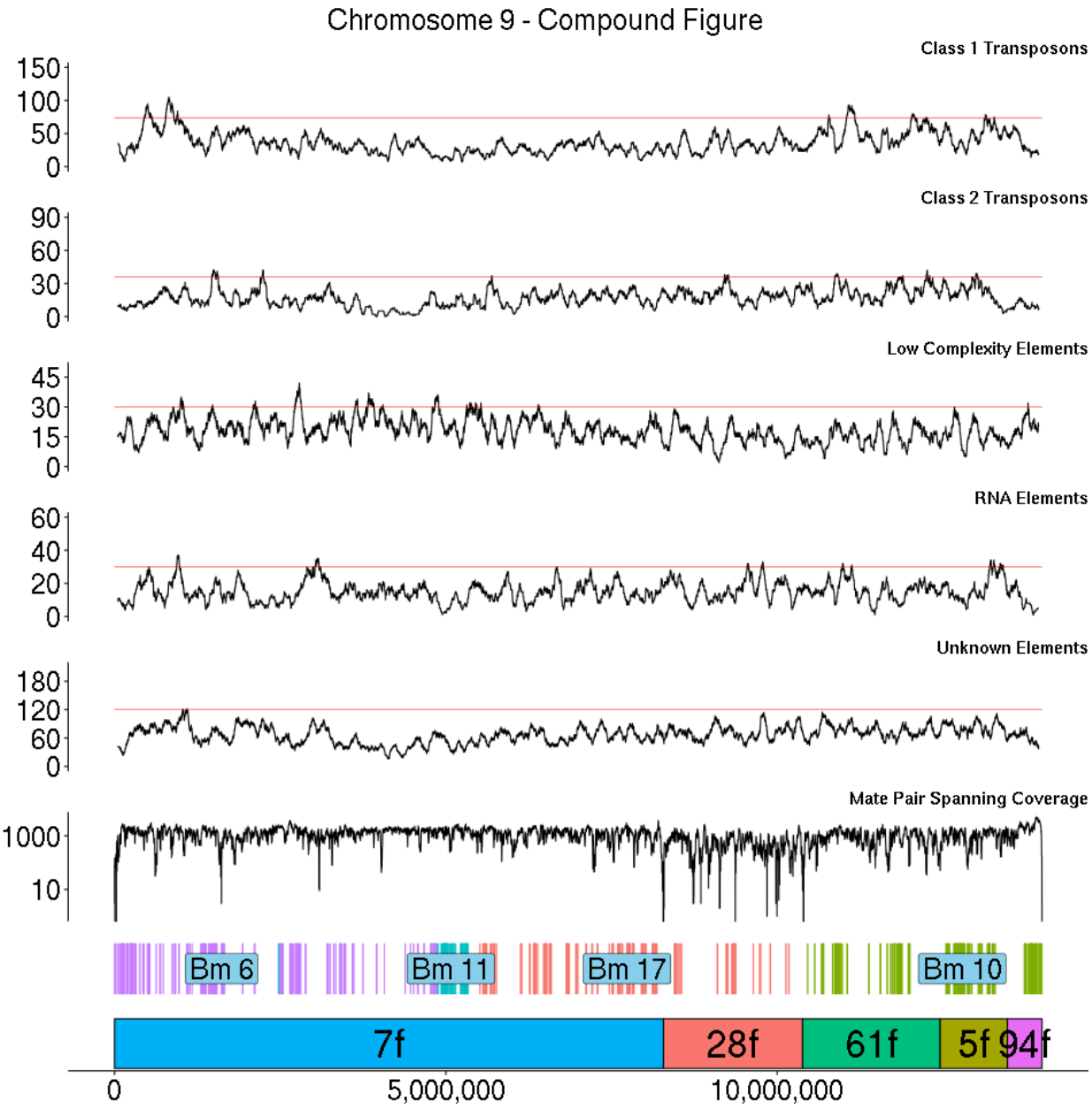

Chromosome 4 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 15
Bm 26
Bm 3
Bm 9
86f
15f
10f
51f
0
5,000,000
10,000,000
15,000,00

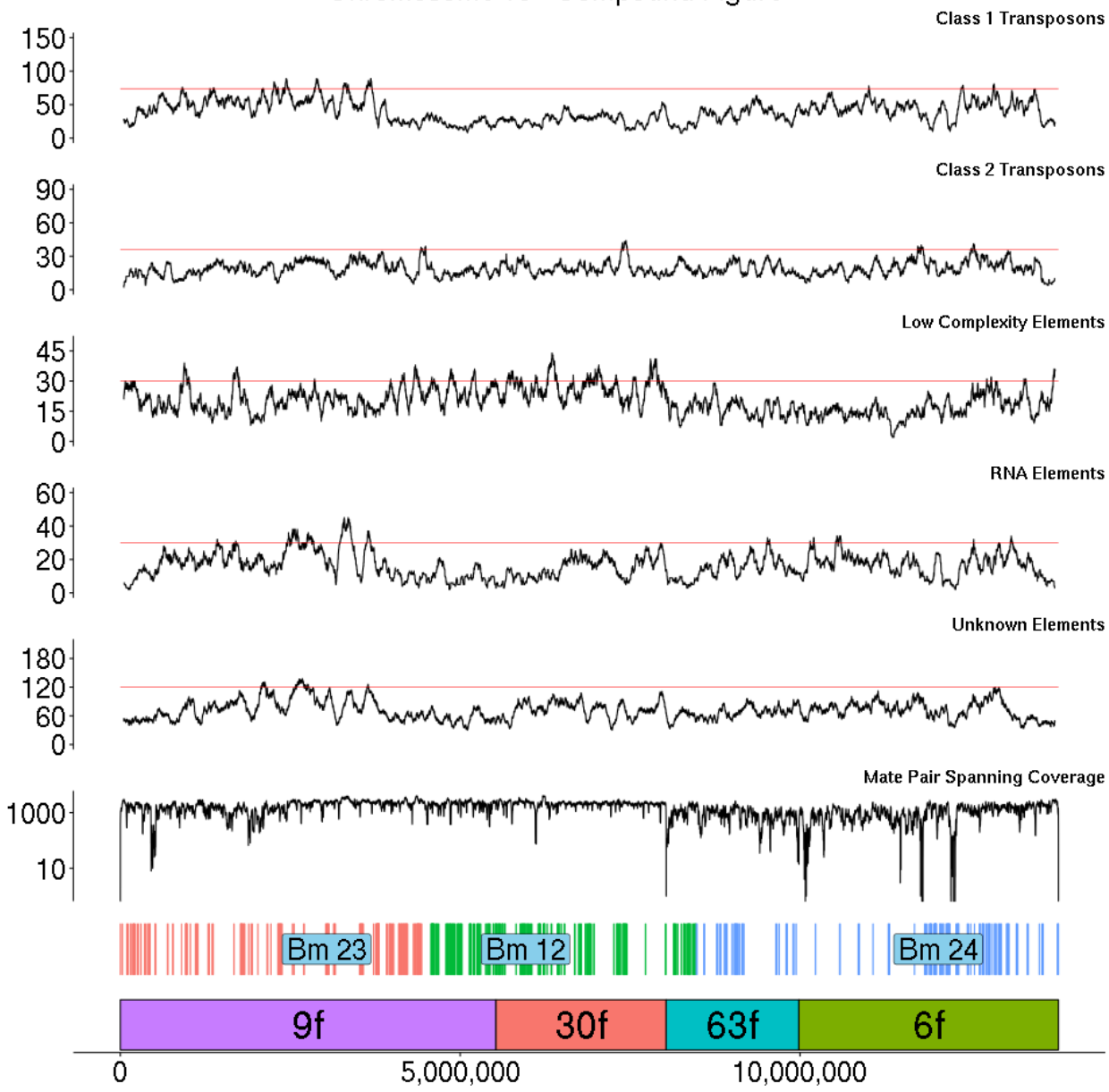

Chromosome 5 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 21
Bm 20
0f
6f
81f
0
5,000,000
10,000,000
15,000,

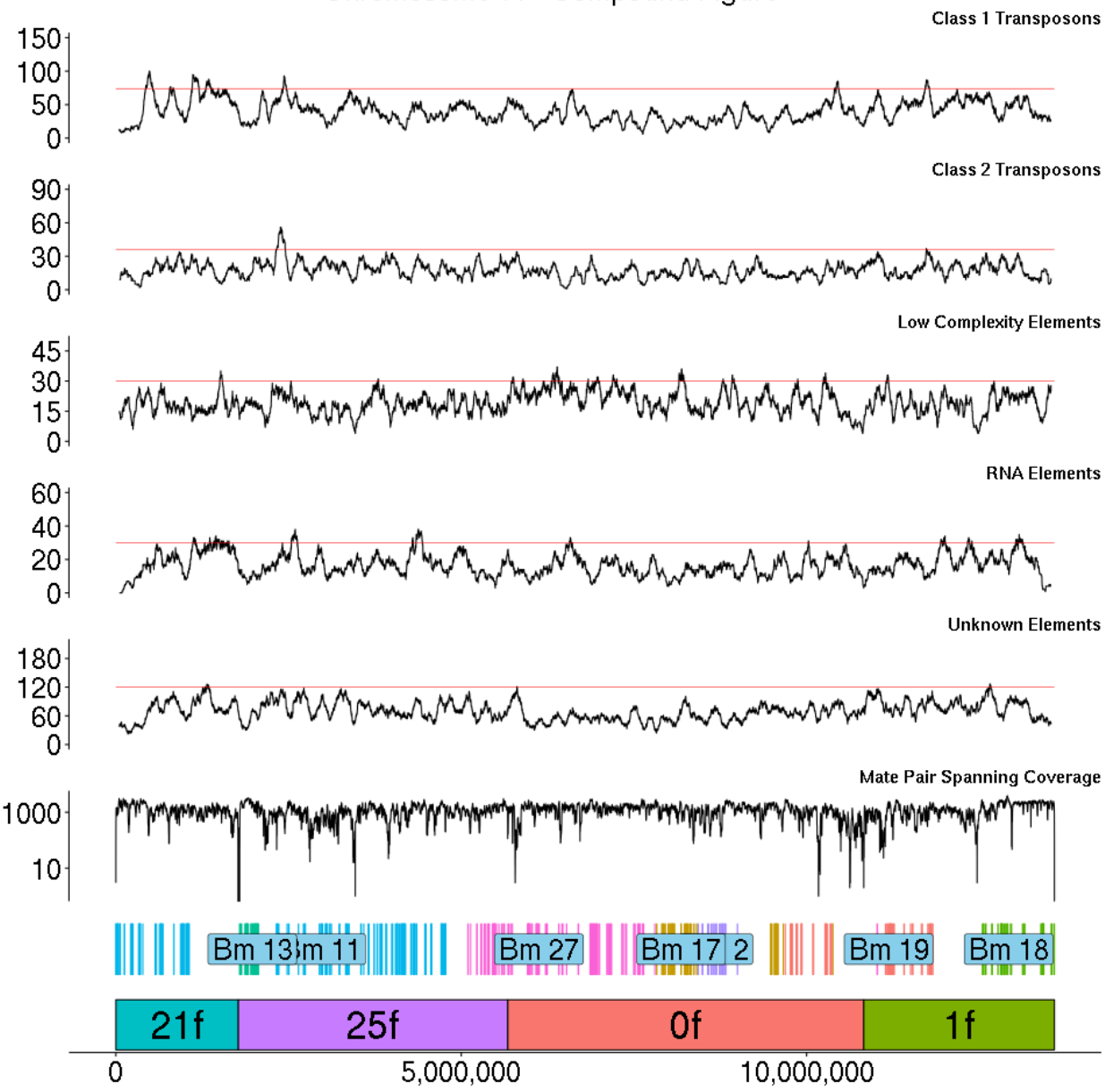

Chromosome 6 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 18
Bm 7
Bm 14
72f
5f
27f
48f
4f
48f
97f
28f
34f
0
5,000,000
10,000,000

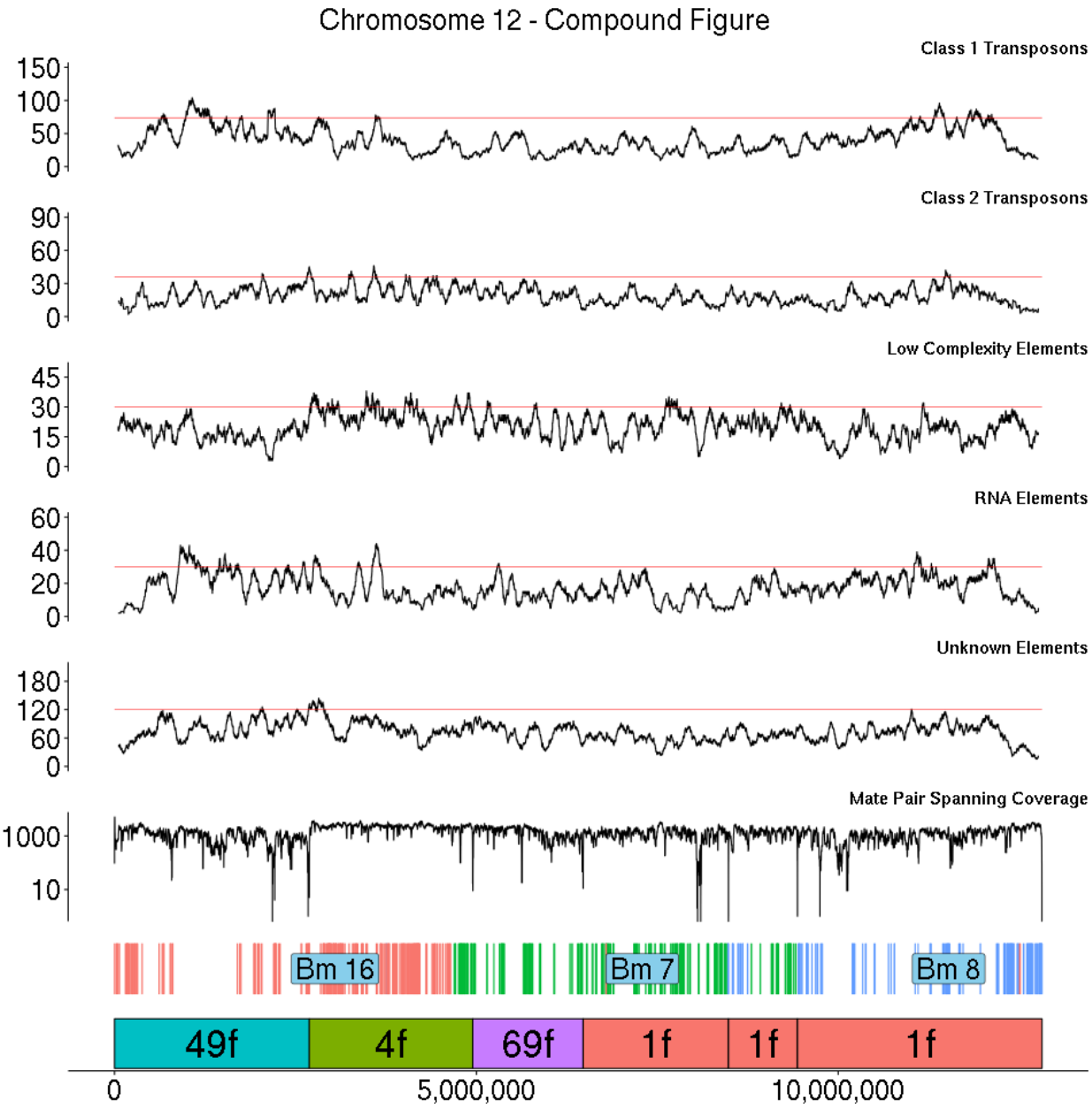

Chromosome 7 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 4
Bm 15
Bm 22
59f
3f
35f
8f
0
5,000,000
10,000,000

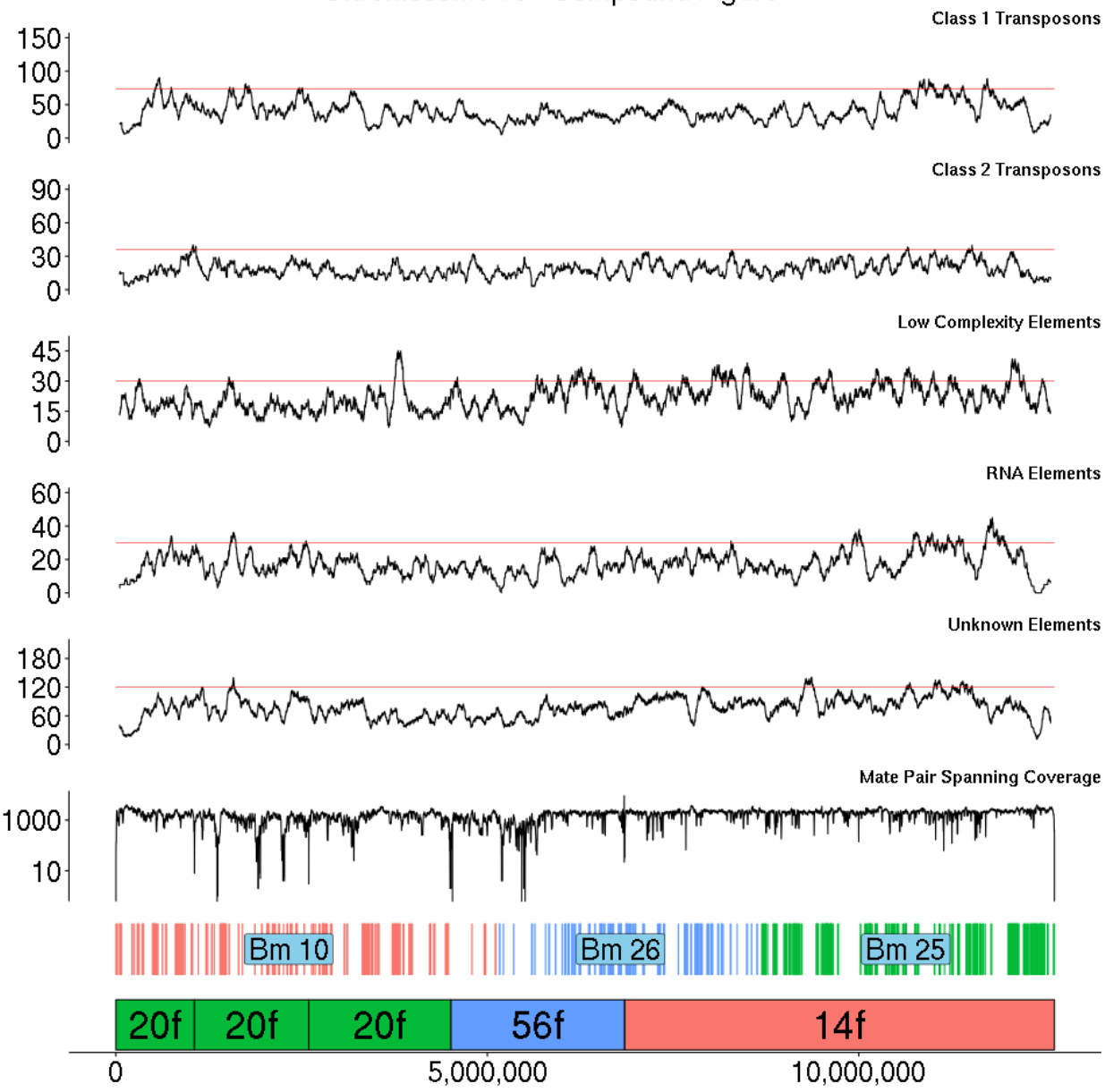

Chromosome 8 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 9
Bm 4
Bm 28
Bm 10
55f
34f
38f
75f
42f
31f
0
5,000,000
10,000,000

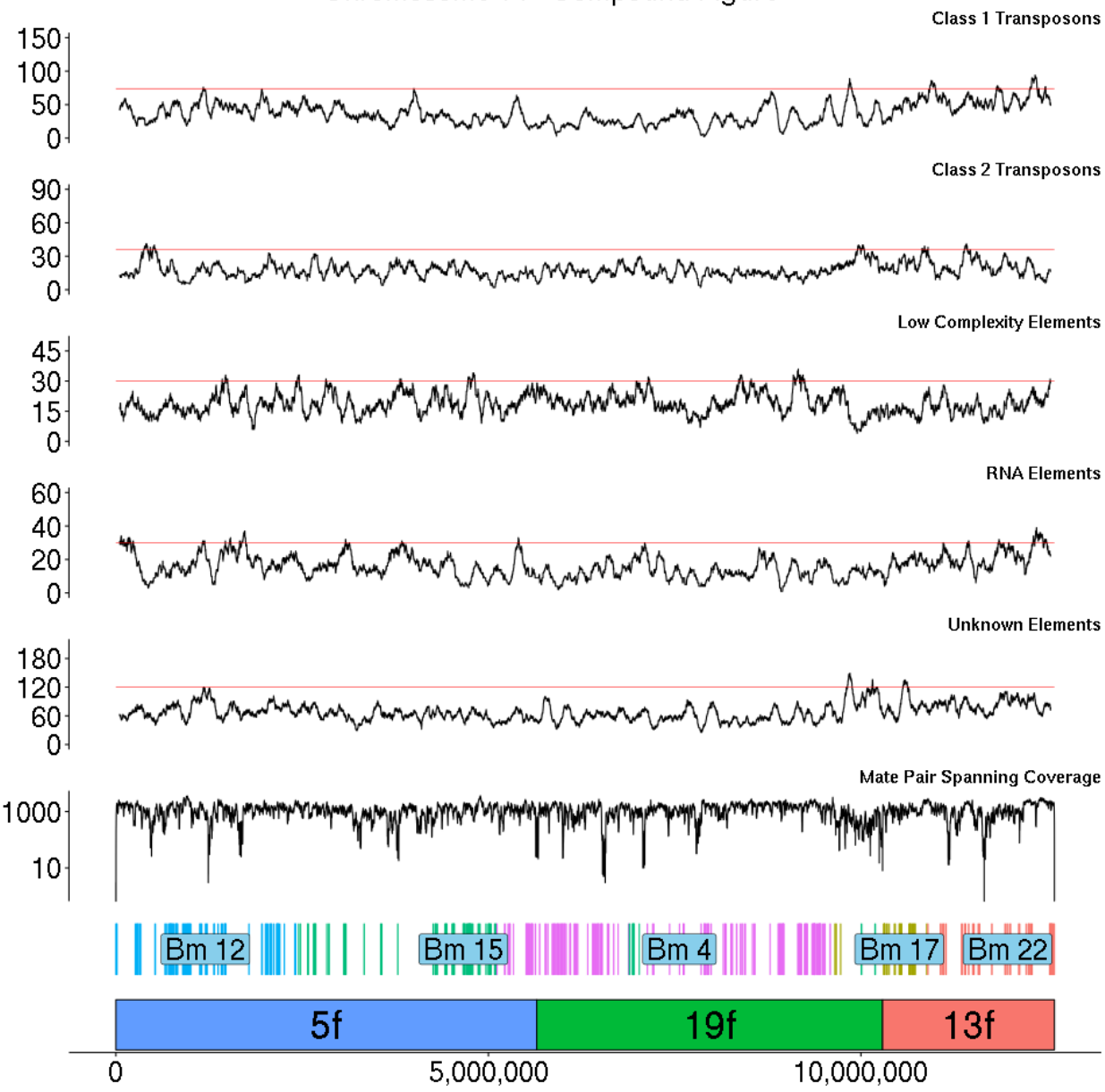

Chromosome 9 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 6
Bm 11
Bm 17
Bm 10
7f
28f
61f
5f
94f
0
5,000,000
10,000,000

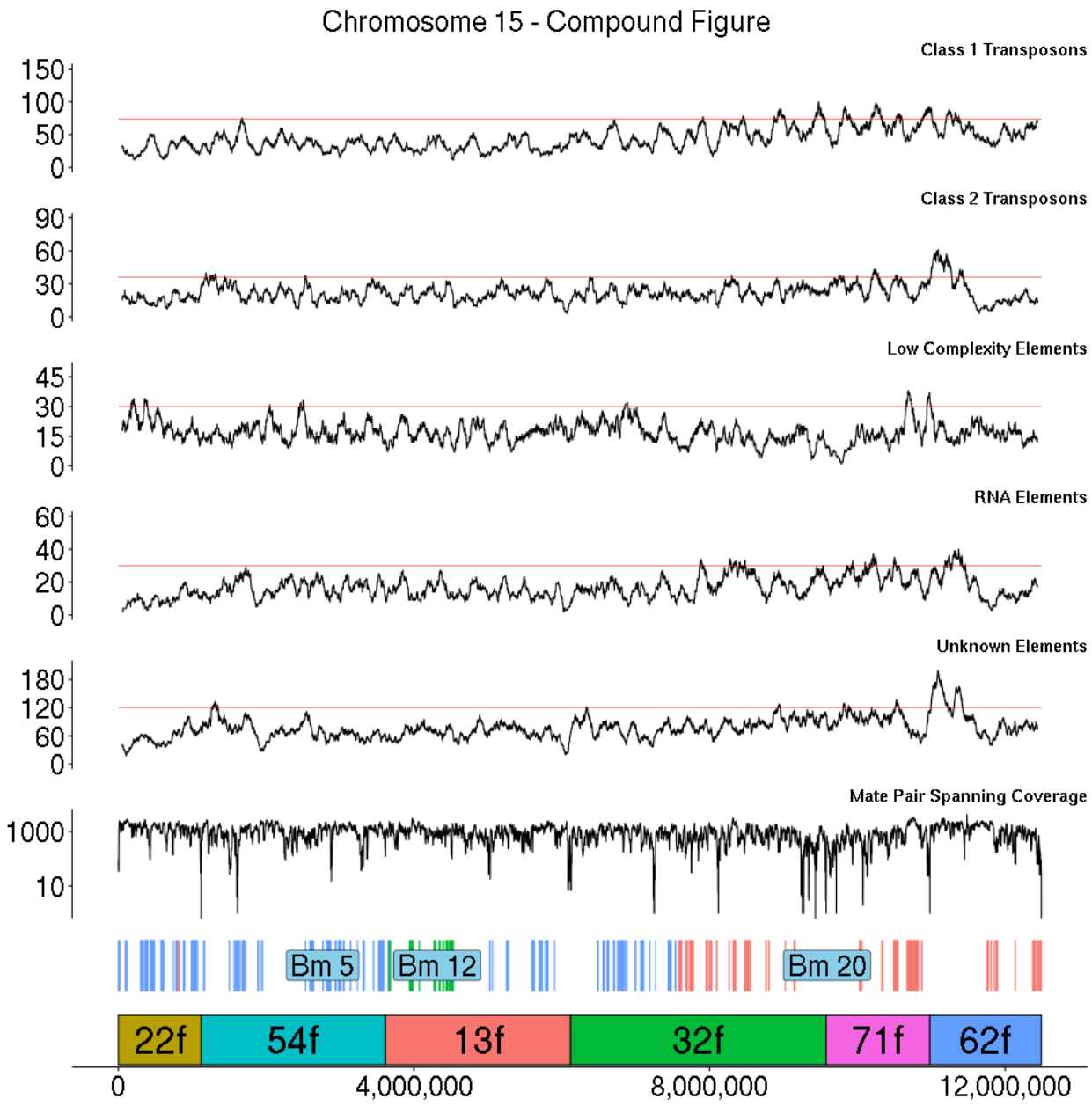

Chromosome 10 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 23
Bm 12
Bm 24
9f
30f
63f
6f
0
5,000,000
10,000,000

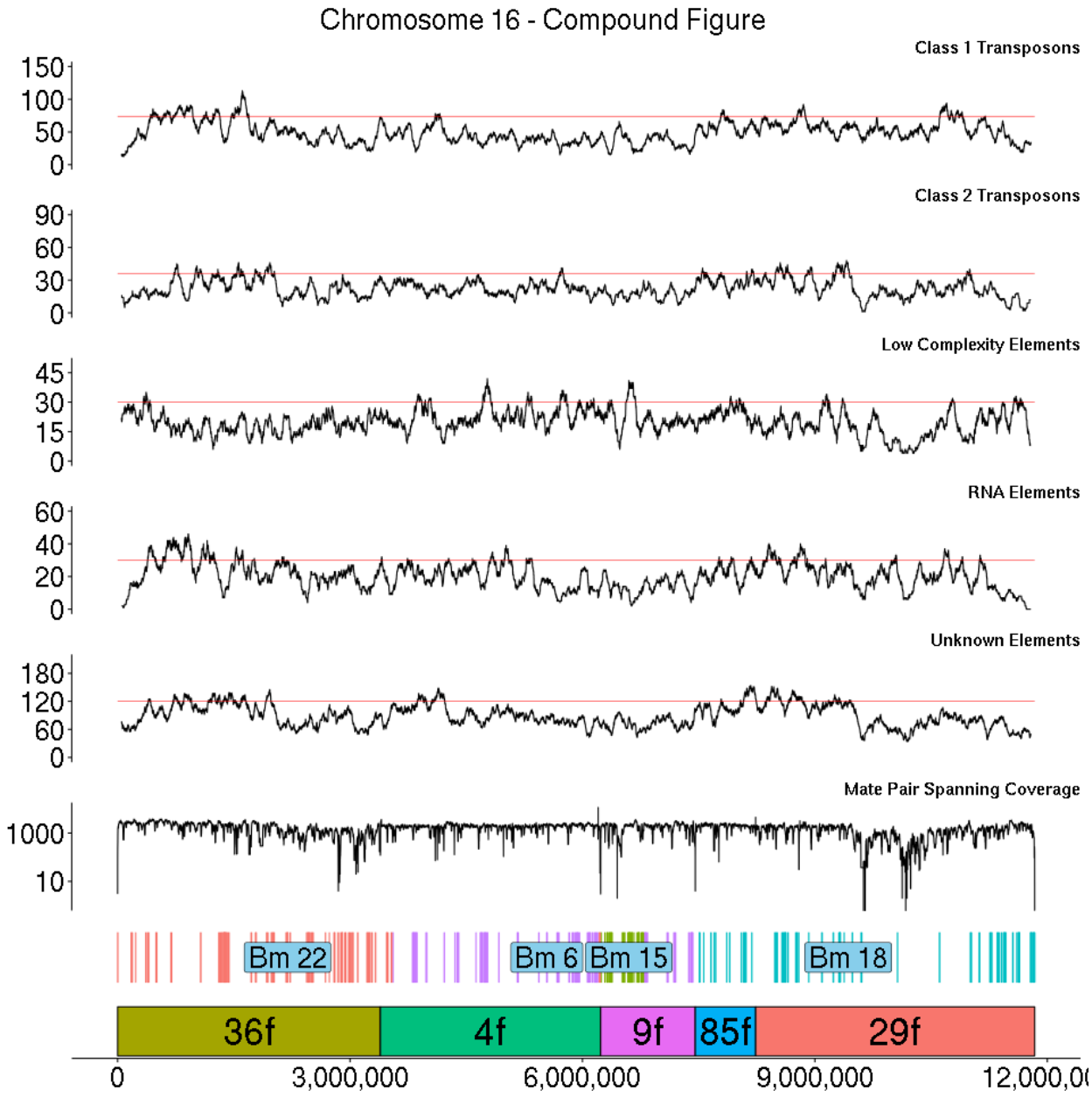

Chromosome 11 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 13
m 11
Bm 27
Bm 17
2
Bm 19
Bm 18
21f
25f
0f
1f
0
5,000,000
10,000,000

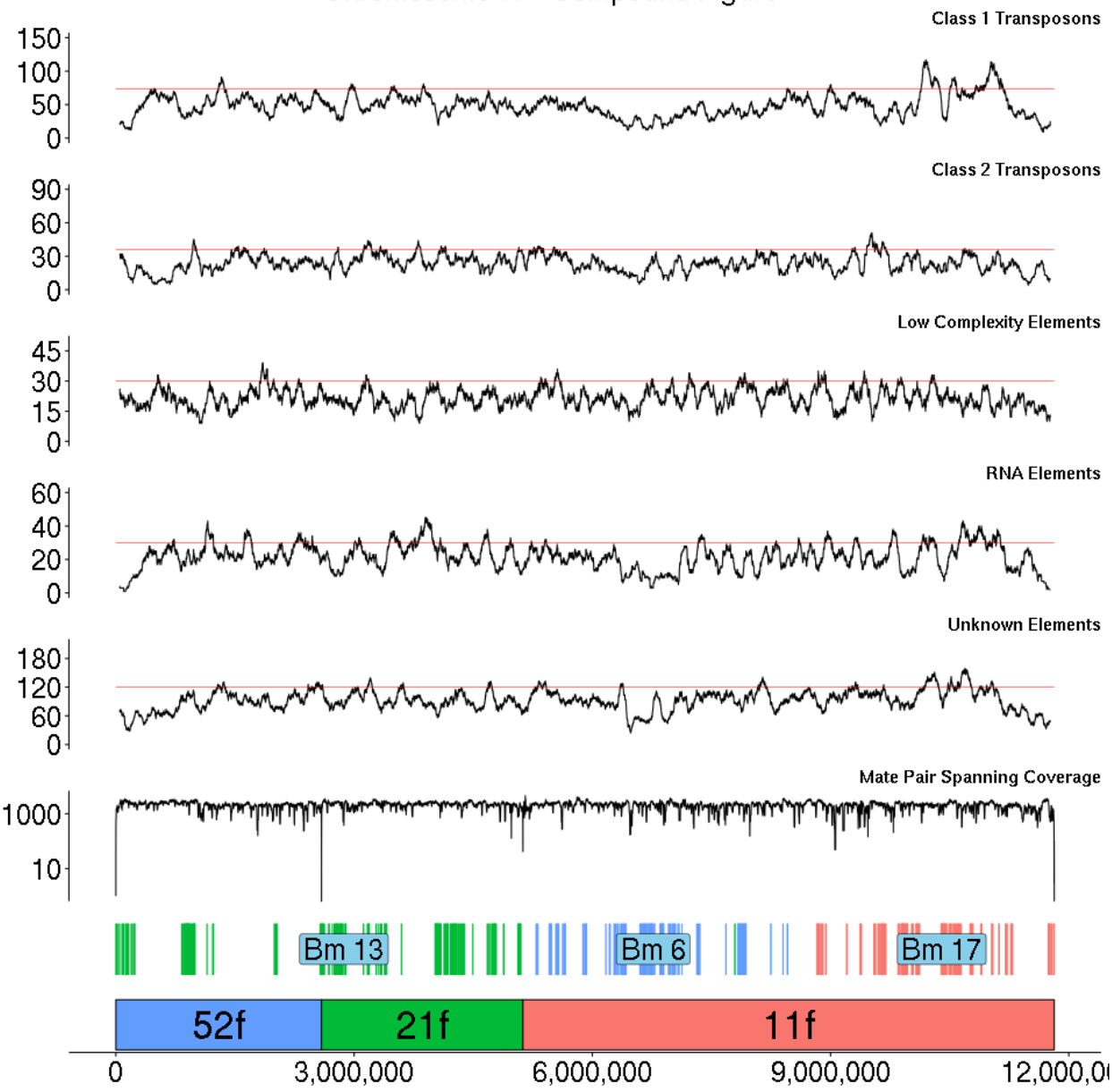

Chromosome 12 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 16
Bm 7
Bm 8
49f
4f
69f
1f
1f
1f
0
5,000,000
10,000,000

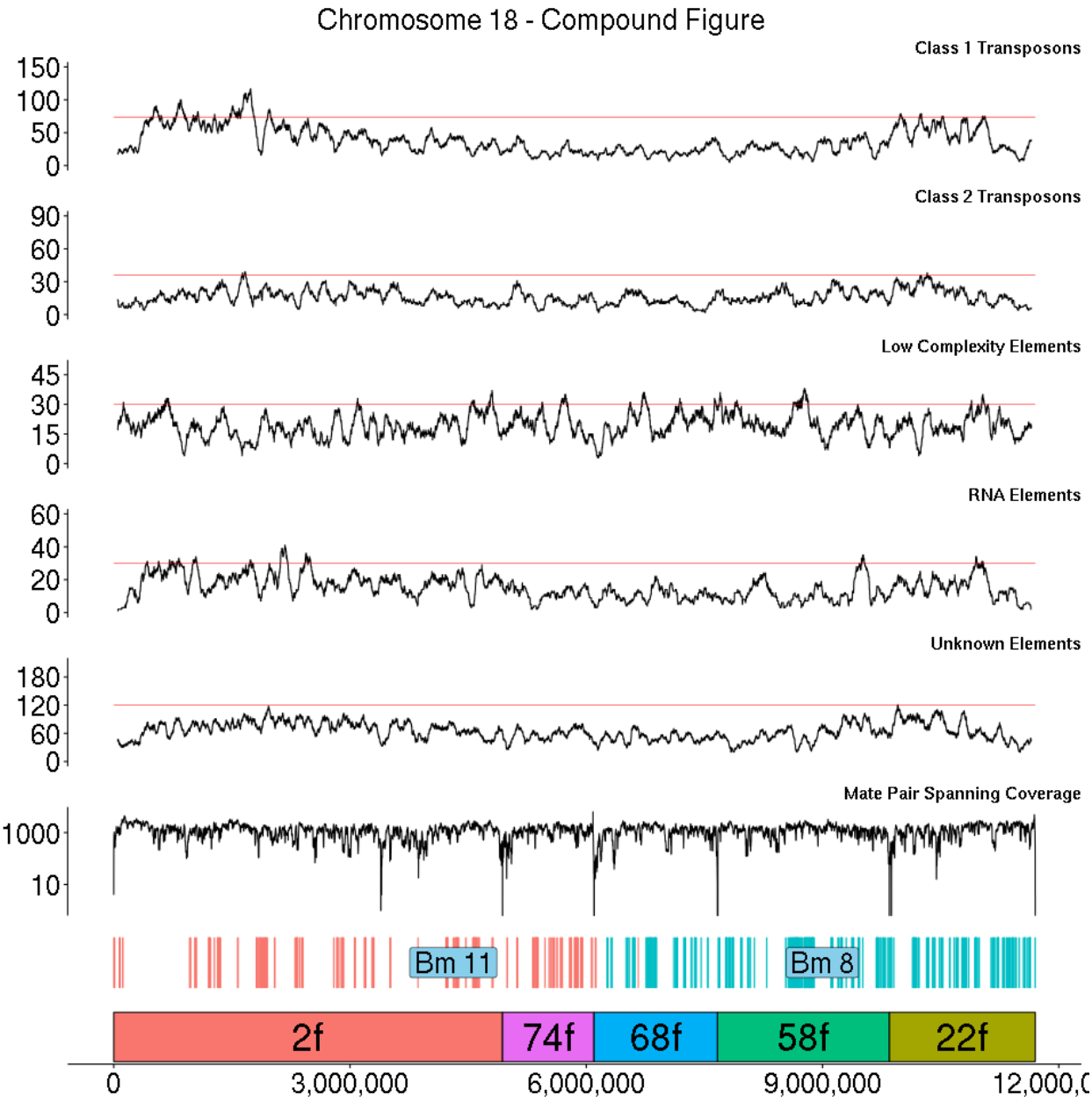

Chromosome 13 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 10
Bm 26
Bm 25
20f
20f
20f
56f
14f
0
5,000,000
10,000,000

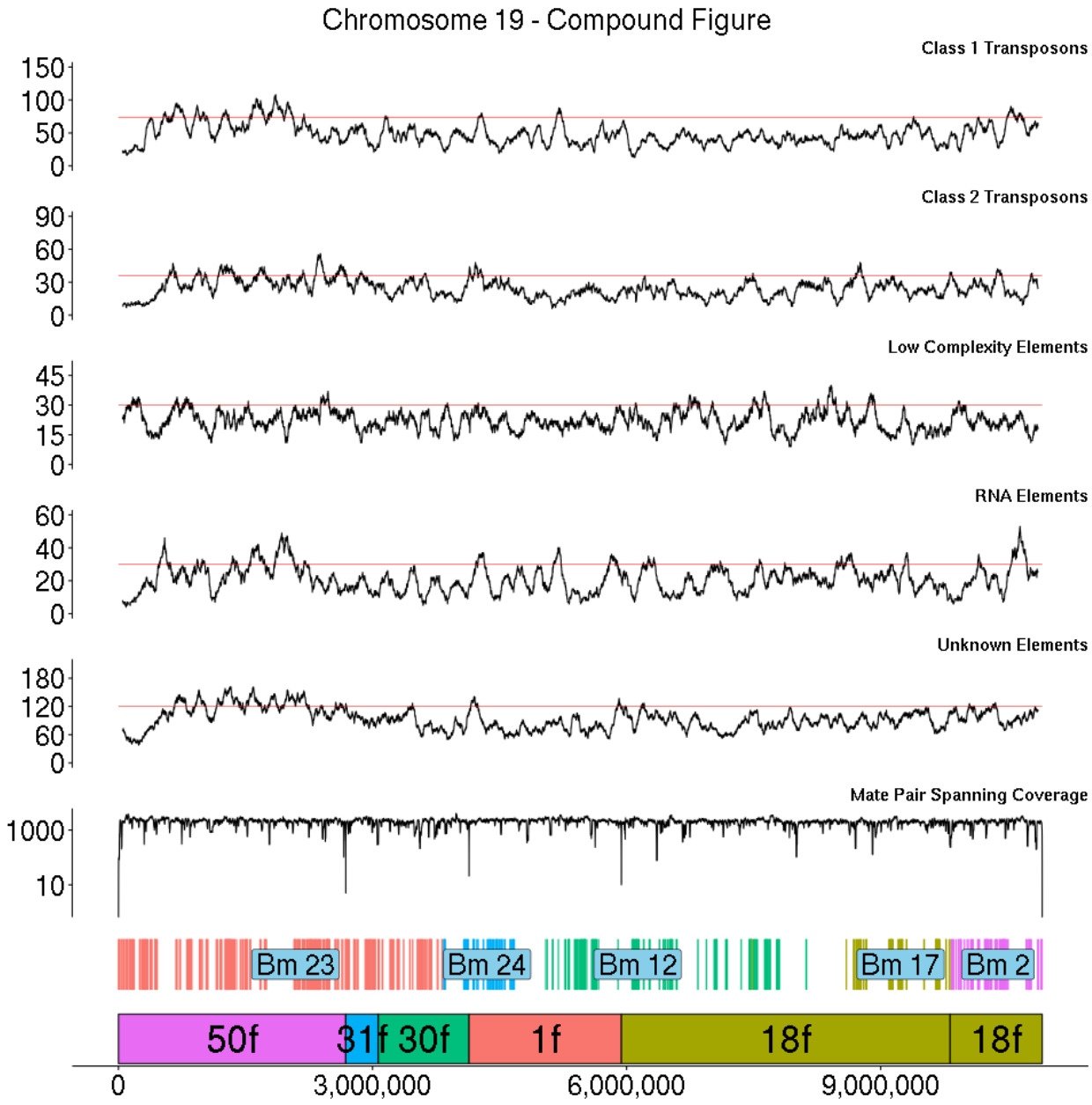

Chromosome 14 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 12
Bm 15
Bm 4
Bm 17
Bm 22
5f
19f
13f
0
5,000,000
10,000,000

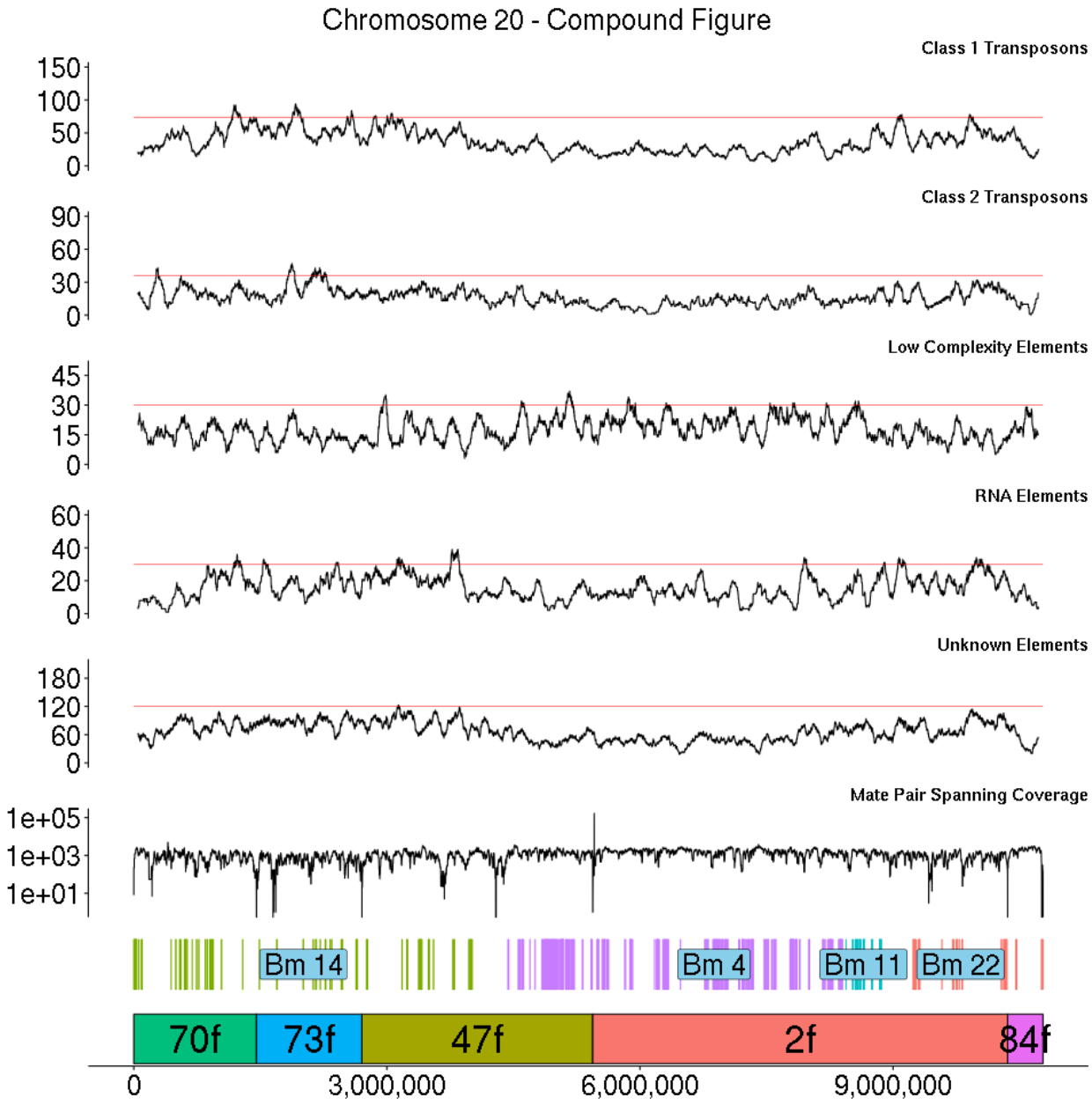

Chromosome 15 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 5
Bm 12
Bm 20
22f
54f
13f
32f
71f
62f
0
4,000,000
8,000,000
12,000,000

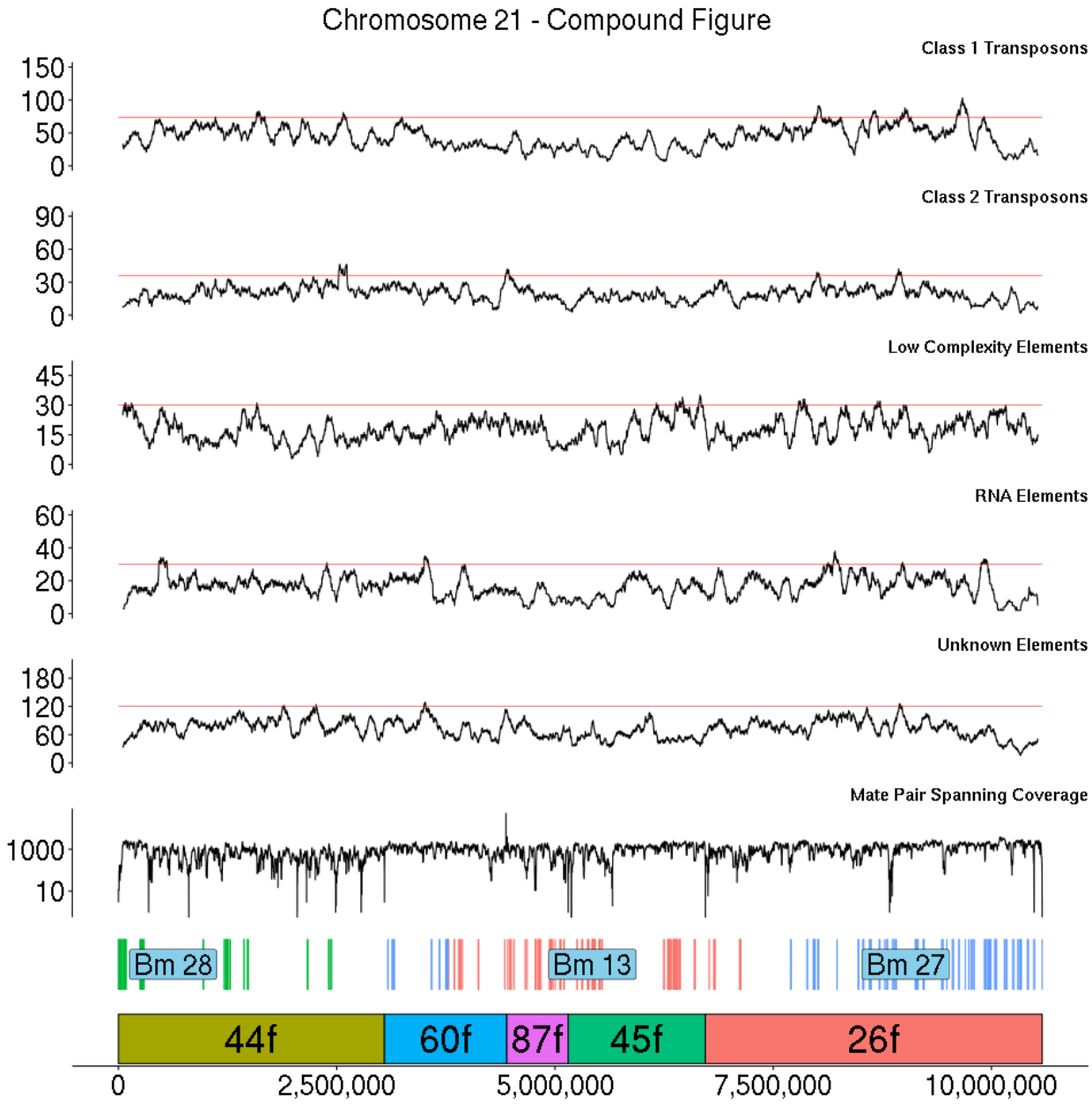

Chromosome 16 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 22
Bm 6
Bm 15
Bm 18
36f
4f
9f
85f
29f
0
3,000,000
6,000,000
9,000,000
12,000,0

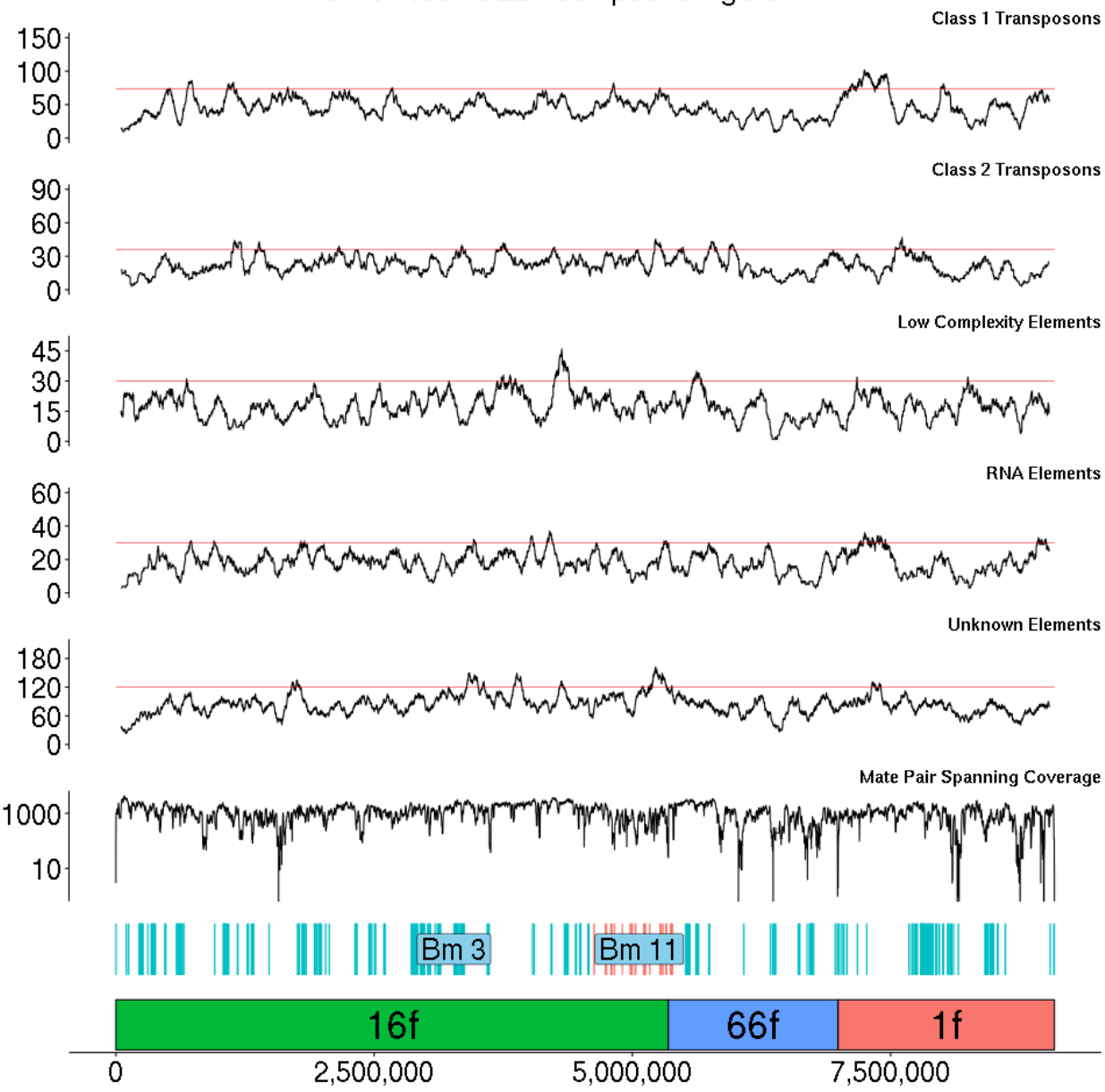

Chromosome 17 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 13
Bm 6
Bm 17
52f
21f
11f
0
3,000,000
6,000,000
9,000,000
12,000,0

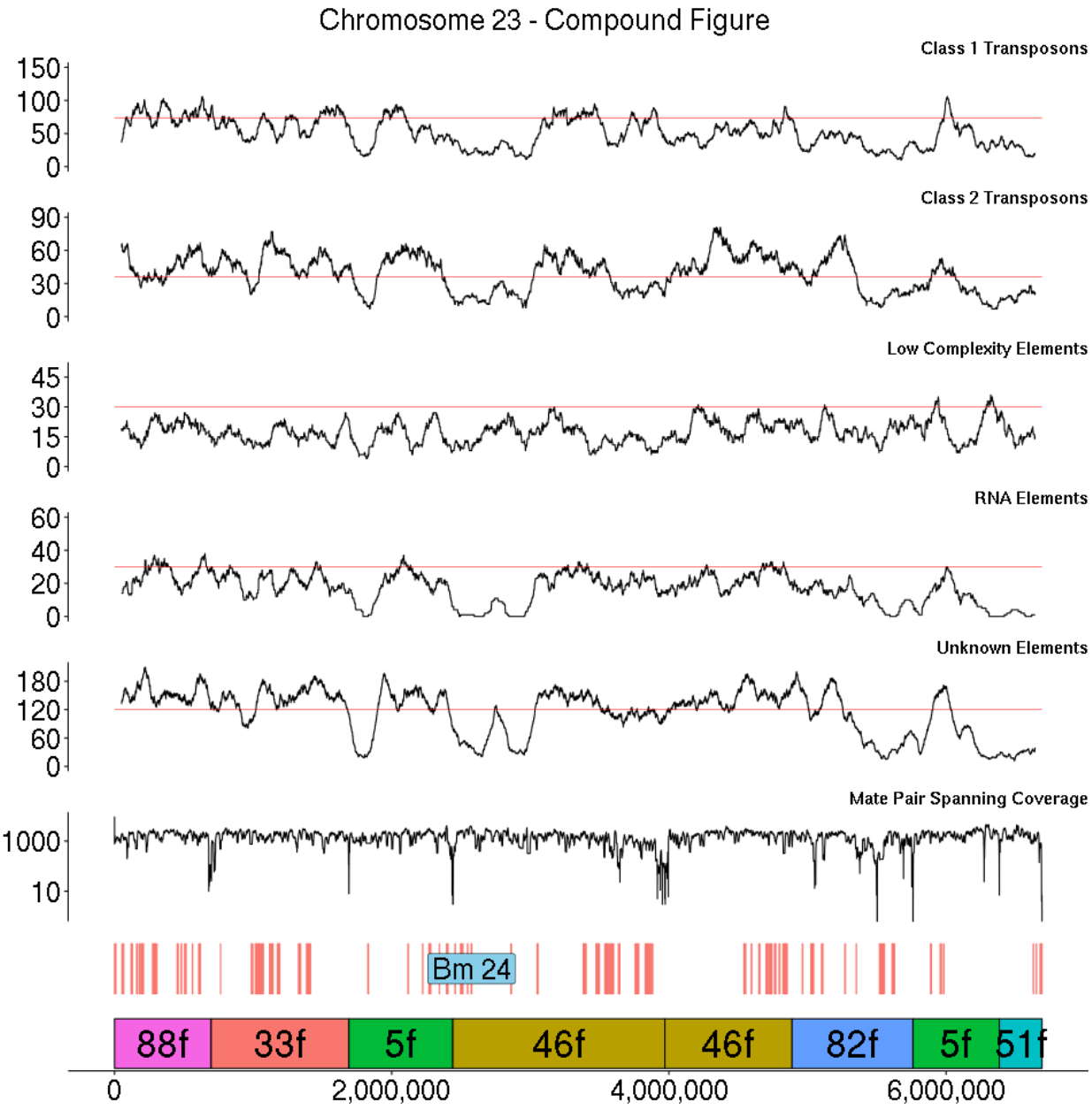

Chromosome 18 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 11
Bm 8
2f
74f
68f
58f
22f
0
3,000,000
6,000,000
9,000,000
12,000,0

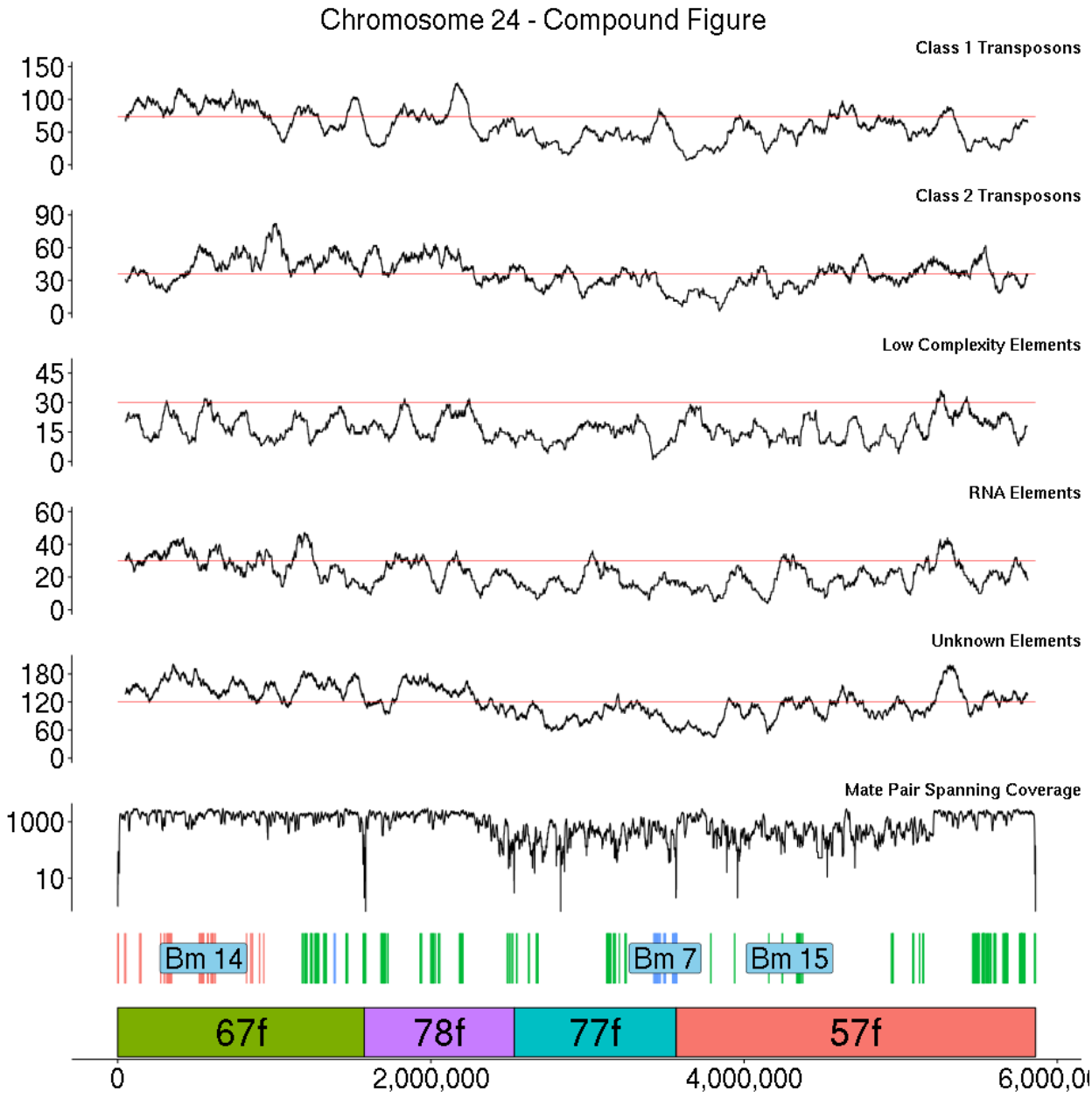

Chromosome 19 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 23
Bm 24
Bm 12
Bm 17
Bm 2
50f
31f
30f
1f
18f
18f
0
3,000,000
6,000,000
9,000,000

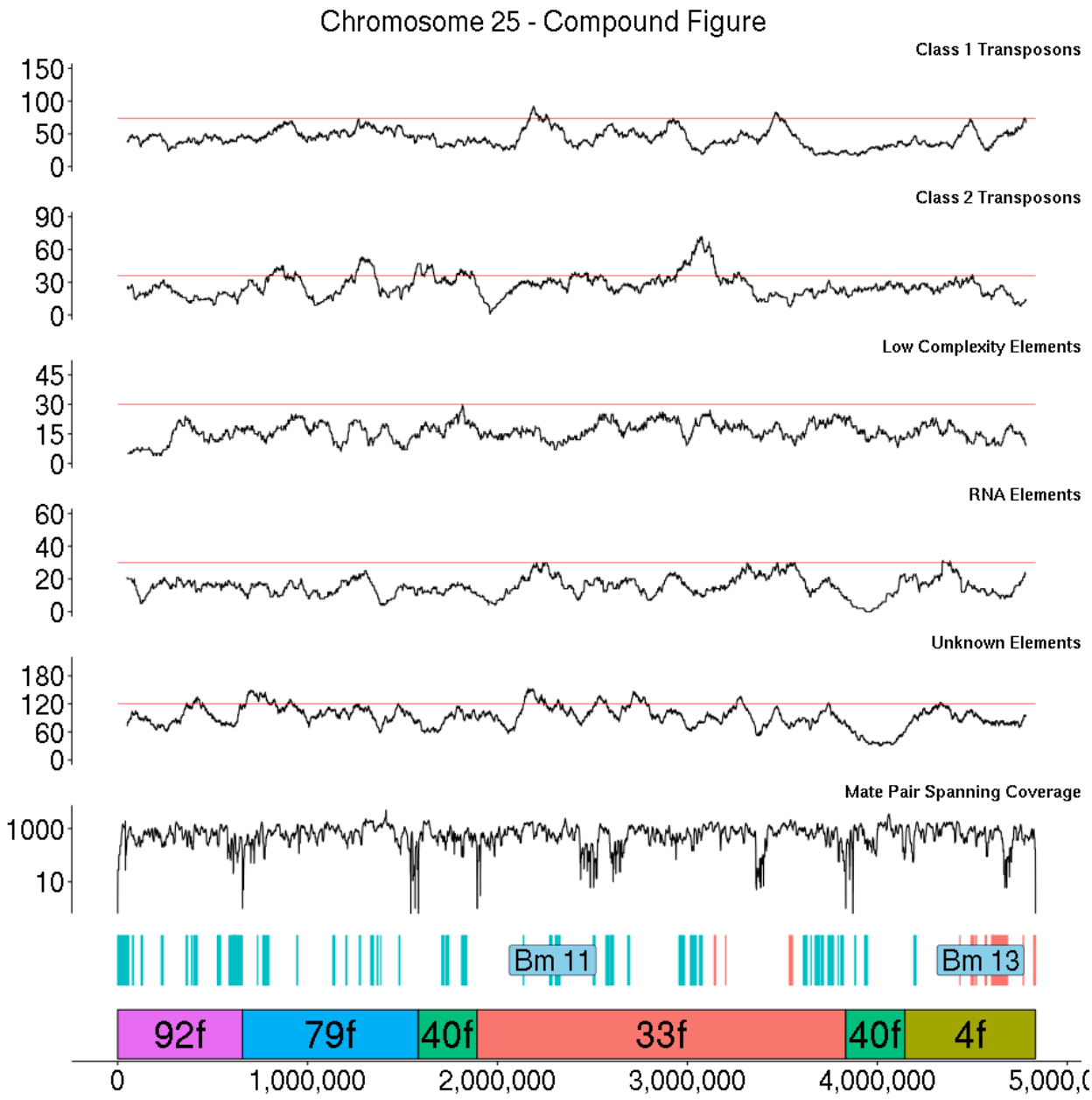

Chromosome 20 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1e+05
1e+03
1e+01
Bm 14
Bm 4
Bm 11
Bm 22
70f
73f
47f
2f
84f
0
3,000,000
6,000,000
9,000,000


Chromosome 21 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 28
Bm 13
Bm 27
44f
60f
87f
45f
26f
0
2,500,000
5,000,000
7,500,000
10,000,000


Chromosome 22 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 3
Bm 11
16f
66f
1f
0
2,500,000
5,000,000
7,500,000

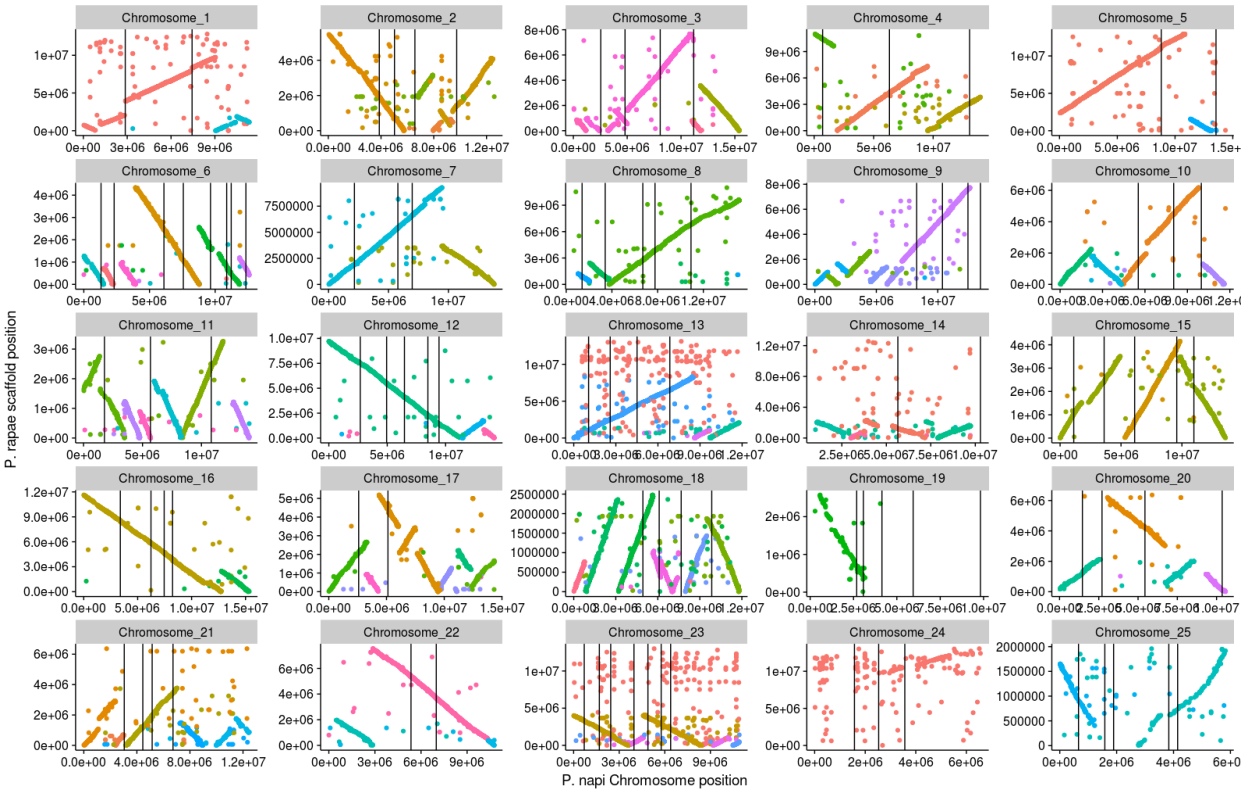

Chromosome 23 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 24
88f
33f
5f
46f
46f
82f
5f
51f
0
2,000,000
4,000,000
6,000,000


Chromosome 24 - Compound Figure
Class 1 Transposons
150
100
50
0
Class 2 Transposons
90
60
30
0
Low Complexity Elements
45
30
15
0
RNA Elements
60
40
20
0
Unknown Elements
180
120
60
0
Mate Pair Spanning Coverage
1000
10
Bm 14
Bm 7
Bm 15
67f
78f
77f
57f
0
2,000,000
4,000,000
6,000,0



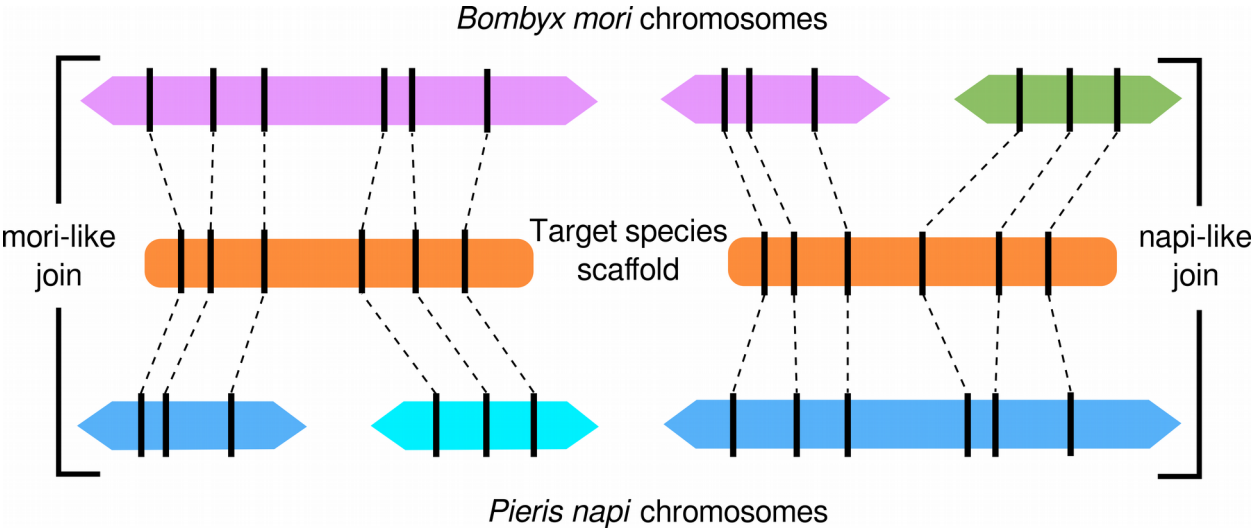
Supplementary fig 4: Repetitive elements were classified using Repeat masker v.4.0.5. The number of repeats in 10,000 base pair windows are plotted along each chromosome for Class 1 and Class 2 transposons, low complexity elements, RNA elements, and unknown elements. Mate-pair spanning depth, collinear block identity, and component scaffold identity are also shown.

Supplementary Figure 5: Alignment of *P. rapae* scaffolds to *P. napi* chromosomes



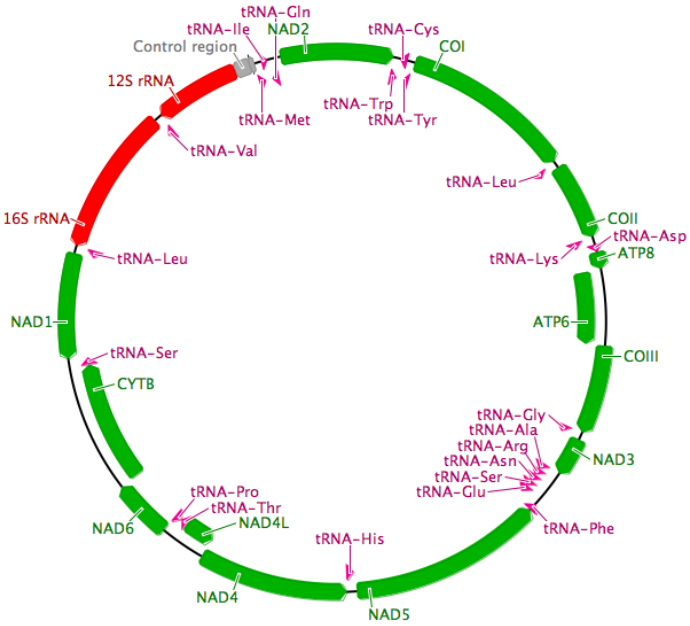
Scaffold joins in the chromosomal assembly of *P. napi* were validated by alignment with *P. rapae* scaffolds. LAST aligner v. 714[28] with default settings aligned the assemblies and found agreement between exons of the two species (mean size 151 bp). Verticle lines represent boundaries between *P. napi* scaffolds, showing where they were joined by the linkage map for chromosomal level assembly (e.g. compare with Supplemental Figure S2). While complete synteny between these two species is very unlikely we assumed that if a HiRise scaffold of *P. rapae* spans two scaffolds of *P. napi* that were joined by the linkage map that the relationship between the *P. napi* scaffolds has greater support. Noise was reduced by filtering out alignments if they had a score < 300 or if they belonged to a *P. rapae* scaffold that had less than 150 other alignments to a giver *P. napi* chromosome.

Supplementary Figure 6: Example of criteria used to determine napi-like and mori-like joins

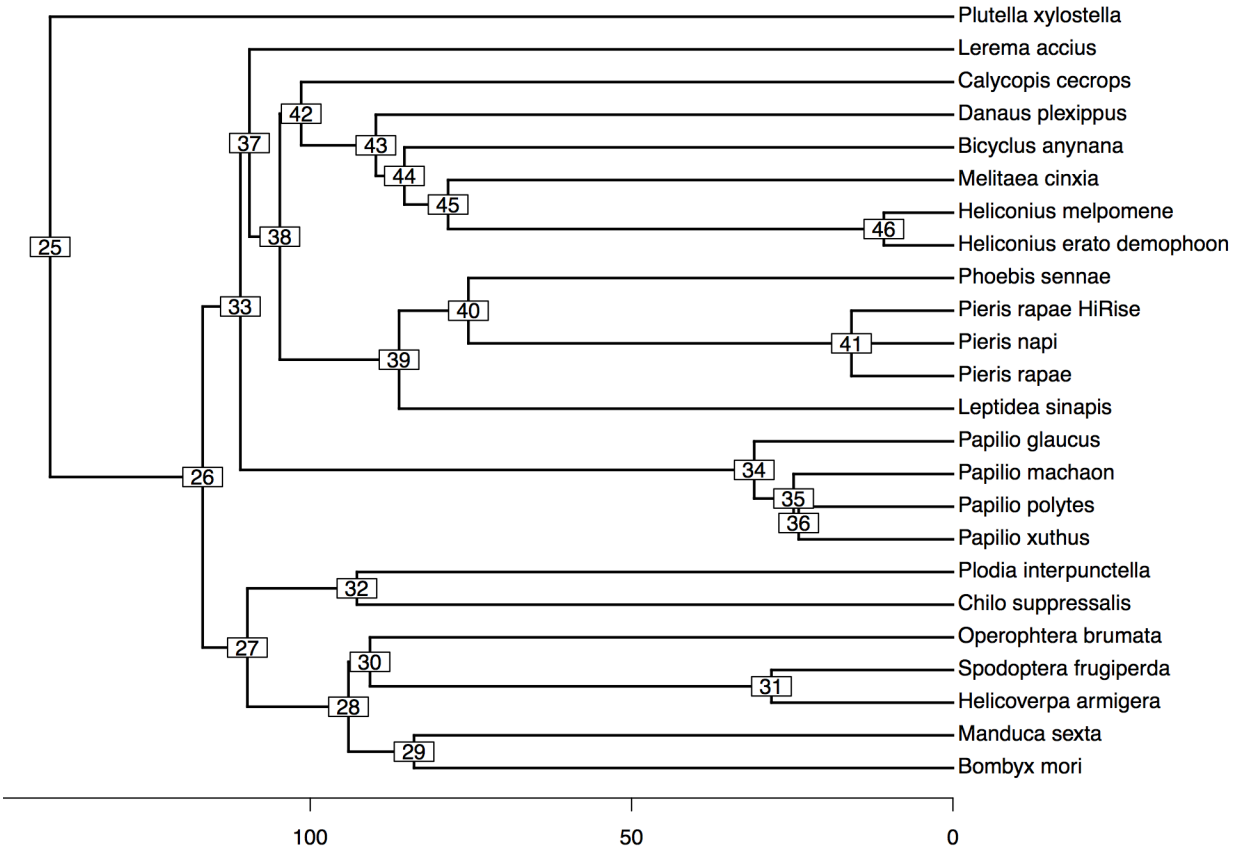


Two alternative examples of how 6 orthologous genes (represented by black rectangles) shared between *P. napi*, *B. mori*, and a target species could allow for the inference of the a napi-like or mori-like chromosome level synteny. In the left example 6 genes called by blastx reside on a single target species scaffold. If their orthologs on *B. mori* reside on a single chromosome and their orthologs on *P. napi* reside on two chromosomes the scaffold “joins” two *P. napi* chromosomes in the same manner as *B. mori* and is counted as supporting a mori-like chromosome. Conversely in the right example the 6 *P. napi* orthologs reside on a single chromosome and the *B. mori* orthologs are split, indicating the scaffold supports a napi-like scaffold.

Supplementary figure 7. The mtDNA of *P. napi*.

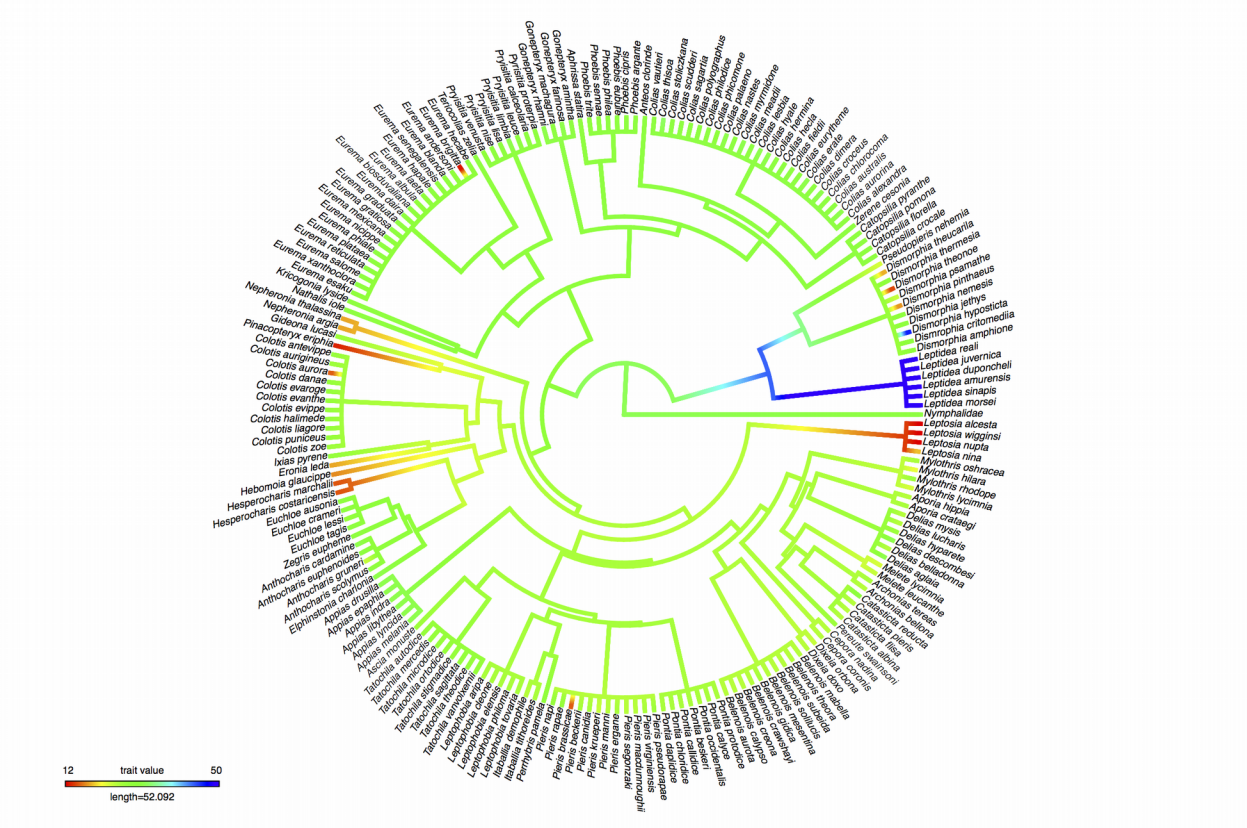


Supplementary figure 8. Chronogram of lepidopteran genomes with node labels.



A chronogram of currently available Lepidopteran genomes (n=24) with estimates of their chromosomal similarity relative to *B. mori* vs. *P. napi,* with time in million years ago (MYA).

Supplementary figure 9. Chronogram of haploid chromosome counts for Pieridae, showing species tip labels.



Ancestral state reconstruction of the chromosomal fusion and fission events across a chronogram of Pieridae (n=201 species). As only a time calibrated genus-level phylogeny exists for Pieridae, all genera with > 1 species were set to an arbitrary polytomy at 5 MYA, while deeper branches reflect calibrated nodes. The haploid chromosomal count of tips (histogram) and interior branches (color coding) are indicated, with the outgroup set to n=31 reflecting the butterfly chromosomal mode.

Supplementary references

1. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage arXiv : 1502 . 05331v1 [q-bio . GN] 18 Feb 2015. 1–25 (2016). doi:10.1101/gr.193474.115.Freely

2. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27,** 578–579 (2011).

3. Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci.* **108,** 1513–1518 (2011).

4. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **27,** 29–34 (1999).

5. Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Research* **42,** (2014).

6. Mitchell, A. *et al.* The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* **43,** D213–D221 (2015).

7. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43,** D1049-56 (2015).

8. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **44,** D471–D480 (2016).

9. Morgat, A. *et al.* UniPathway: A resource for the exploration and annotation of metabolic pathways. *Nucleic Acids Res.* **40,** (2012).

10. Fabregat, A. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **44,** D481–D487 (2016).

11. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204-12 (2015).

12. Jones, P. *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30,** 1236–1240 (2014).

13. Tate, R., Hall, B., DeRego, T. & Geib, S. Annie: the ANNotation Information Extractor (Version 1.0) [Software]. (2014). at <http://genomeannotation.github.io/annie>

14. Shimomura, M. *et al.* KAIKObase: an integrated silkworm genome database and data mining tool. *BMC Genomics* **10,** 486 (2009).

15. Zdobnov, E. M. *et al.* OrthoDB v9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45,** D744–D749 (2017).

16. Heikkila, M., Kaila, L., Mutanen, M., Pena, C. & Wahlberg, N. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. *Proc. R. Soc. B Biol. Sci.* **279,** 1093–1099 (2012).

17. Condamine, F. L., Sperling, F. A. H. & Kergoat, G. J. Global biogeographical pattern of swallowtail diversification demonstrates alternative colonization routes in the Northern and Southern hemispheres. *J. Biogeogr.* **40,** 9–23 (2013).

18. Edger, P. P. *et al.* The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci.* **112,** 8362–8366 (2015).

19. Wahlberg, N., Wheat, C. W. & Peña, C. Timing and Patterns in the Taxonomic Diversification of Lepidoptera (Butterflies and Moths). *PLoS One* **8,** e80875 (2013).

20. Kozak, K. M. *et al.* Multilocus species trees show the recent adaptive radiation of the mimetic heliconius butterflies. *Syst. Biol.* **64,** 505–524 (2015).

21. Mitchell, A., Mitter, C. & Regier, J. C. Systematics and evolution of the cutworm moths (Lepidoptera: Noctuidae): Evidence from two protein-coding nuclear genes. *Syst. Entomol.* **31,** 21–46 (2006).

22. Bushnell, B. BBTools. (2017). at <https://jgi.doe.gov/data-and-tools/bbtools/>

23. Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30,** 566–8 (2014).