

Supplementary information

1- Databases used for gene evidence in the Funannotate pipeline.

- MEROPS Database: version=12.0 date=2017-10-04 records=5,009
- UniProtKB Database: version=2021_03 date=2021-06-02 records=565,254
- dbCAN Database: version=9.0 date=2020-08-04 records=607
- Pfam Database: version=34.0 date=2021-03 records=19,179
- Repeat Database: version=1.0 date=2021-07-28 records=11,950
- GO ontology version=2021-07-02 date=2021-07-02 records=47,228
- InterProScan XML: version=86.0 date=2021-06-03 records=38,913
- BUSCO outgroups: version=1.0 date=2021-07-28 records=8
- Gene2Product: version=1.70 date=2021-06-15 records=34,039
- Downloading busco models: insecta

2- Ultrametric tree calculated for the studied species in Newick format.

```
((Eufriesea_mexicana:57.8594,Euglossa_dilemma:57.8594):42.8287,(((Bombus_terrestris:2.4391,(Bombus_impatiens:8.42918,(Bombus_vancouverensis:1.21064,Bombus_bifarius:1.21064):7.21854):14.0099):52.1207,(Frieseomelitta_varia:44.5763,(Melipona_quadrifasciata:31.4319,Melipona_bicolor:31.4319):13.1444):29.9835):15.9248,((Apis_mellifera:18.7074,Apis_cerana:18.7074):8.68206,Apis_dorsata:27.3895):63.0952):10.2034):4.31195,Habropoda_labriosa:105);
```

Supplementary Tables

Table SI. List of bee genome assemblies used for comparative analyses.

Species	NCBI accession	Scaffolds	Contigs	Length (Mb)	CDS	Scaffolds N50	Scaffolds L50	N in 100 kbp (%)	Reference
<i>Apis cerana</i>	GCF_001442555.1	2,430	10,706	228.3	20,933	1.4 Mb	42	2.31	Park et al. (2015) <i>BMC genomics</i> . DOI: 10.1186/1471-2164-16-1
<i>Apis dorsata</i>	GCF_000469605.1	4,040	45,204	230.3	20,012	732.1 kb	87	4.95	Fouks et al. (2021) <i>Genome Res.</i> DOI: 10.1101/gr.272310.120
<i>Apis mellifera</i>	GCF_003254395.2	176	227	225.2	23,471	13.6 Mb	7	0	Wallberg et al. (2019) <i>BMC Genomics</i> . DOI: 10.1186/s12864-019-5642-0
<i>Bombus bifarius</i>	GCF_011952205.1	1,249	1,251	266.8	23,896	2.2 Mb	30	0	Heraghty et al. (2020) <i>G3</i> . DOI: 10.1534/g3.120.401437
<i>Bombus impatiens</i>	GCF_000188095.3	5,460	16,060	246.9	24,471	1.4 Mb	54	0.79	Sad et al. (2015) <i>Genome Biol.</i> DOI: 10.1186/s13059-015-0623-3
<i>Bombus terrestris</i>	GCF_000214255.1	5,678	10,672	248.7	22,091	3.5 Mb	18	0.24	Sad et al. (2015) <i>Genome Biol.</i> DOI: 10.1186/s13059-015-0623-3
<i>Bombus vancouverensis</i>	GCF_011952275.1	1,162	1,164	282.1	24,285	3.1 Mb	24	0	Heraghty et al. (2020) <i>G3</i> . DOI: 10.1534/g3.120.401437
<i>Eufriesea mexicana</i>	GCF_001483705.1	187,611	212,650	596.3	15,659	351.9 kb	377	0	Kapheim et al. (2015) <i>Science</i> . DOI: 10.1126/science.aaa4788
<i>Euglossa dilemma</i>	GCA_002201625.1	22,698	110,426	588.2	15,904	143.6 kb	419	5.3	Brand et al. (2017) <i>G3</i> . DOI: 10.1534/g3.117.043687
<i>Frieseomelitta varia</i>	GCF_011392965.1	2,173	6,729	275.4	23,628	467 kb	176	1.12	Freitas et al. (2020) <i>BMC Genomics</i> . DOI: 10.1186/s12864-020-06784-8
<i>Habropoda laboriosa</i>	GCF_001263275.1	27,566	50,205	297	12,256	1.8 Mb	47	0	Kapheim et al. (2015) <i>Science</i> . DOI: 10.1126/science.aaa4788
<i>Melipona bicolor</i>	GCA_030673865.1	241	244	259.9	20,278	6.2 Mb	12	0	This study
<i>Melipona quadrifasciata</i>	GCA_001276565.1	2,866	38,604	256.3	14,257	1.9 Mb	42	0	Kapheim et al. (2015) <i>Science</i> . DOI: 10.1126/science.aaa4788
<i>Tetragonula carbonaria</i>	GCA_010645115.1	NA	90201	299.6	NA	13.6 kb	5949	0	NA
<i>Tetragonula clypearis</i>	GCA_010645135.1	NA	77144	294	NA	14.9 kb	5525	0	NA
<i>Tetragonula davenporti</i>	GCA_010645165.1	NA	47286	284.6	NA	18.5 kb	4237	0	NA
<i>Tetragonula hockingsi</i>	GCA_010645185.1	NA	112753	307.1	NA	10.5 kb	7605	0	NA
<i>Tetragonula mellipes</i>	GCA_011634685.1	NA	133257	337.6	NA	16.4 kb	5285	0	NA

Table SII. Orthogroups exclusive to eusocial corbiculate bees and containing *M. bicolor*.

HOG	<i>A. cerana</i>	<i>A. dorsata</i>	<i>A. mellifera</i>	<i>B. bifarius</i>	<i>B. impatiens</i>	<i>B. terrestris</i>	<i>B. vancouverensis</i>	<i>F. varia</i>	<i>M. bicolor</i>	<i>M. quadrifasciata</i>	Annotation in <i>M. bicolor</i>	GO in <i>M. bicolor</i>
N0.HOG 0002289	XP_0169 14686.1, XP_0169 14687.1, XP_0169 13788.1	XP_031366 627.1	XP_0011202 93.2, NP_0010116 14.1	XP_0333191 82.1, XP_0333191 84.1	XP_012248547. 1, XP_012248546. 1, XP_033180239. 1	XP_012170987. 1, XP_012170988. 1	XP_033185943.1	KAF3421744 .1, KAF3421743 .1	K0M31_01 2811-T1	KOX69497.1	Phosphatidylcholine 2- acylhydrolase (EC 3.1.1.4)	GO:0006644, GO:0050482, GO:0016042, GO:0005576, GO:0102567, GO:0102568, GO:0004623, GO:0046872
N0.HOG 0003494	XP_0285 22782.1, XP_0285 20410.1, XP_0285 20409.1, XP_0169 05282.1	XP_031370 965.1	XP_0263006 28.1	XP_0333192 59.1	XP_024221752. 1	XP_012171089. 2	XP_033199797.1, XP_033182849.1		K0M31_00 9555-T1, K0M31_00 9554-T1	KOX67239.1, KOX67232.1	Uncharacterized protein	
N0.HOG 0005010	XP_0169 19808.1	XP_031365 150.1	XP_0065681 69.1, XP_624119.2	XP_0333155 48.1	XP_024220196. 1	XP_012168725. 1, XP_003398715. 1	XP_033198891.1	KAF3430281 .1, KAF3426545 .1, KAF3429429 .1	K0M31_00 2407-T1	KOX77185.1	Putative RNA-binding protein EEED8.10 (Fragment)	GO:0003723
N0.HOG 0008548	XP_0169 04248.1	XP_031368 500.1	XP_0167661 72.1	XP_0332976 22.1	XP_003487765. 1	XP_012165659. 1	XP_033183336.1	KAF3422505 .1, KAF3422504 .1	K0M31_00 6915-T1, K0M31_00 6914-T1	KOX73507.1	Uncharacterized protein	GO:0016021
N0.HOG 0008949	XP_0169 16307.1	XP_006611 907.1	NP_0010116 13.1	XP_0333106 62.1	XP_003491768. 1, XP_033178254. 1	XP_012175465. 1	XP_033196951.1, XP_033196952.1	KAF3424888 .1	K0M31_02 0235-T1	KOX78578.1	Apidaecins type 22	GO:0005576, GO:0042742
N0.HOG 0009274	XP_0285 22658.1	XP_031367 232.1, XP_031367 231.1	XP_0262953 88.1, XP_0262954 34.1	XP_0333069 01.1	XP_033175218. 1	XP_020724381. 1, XP_020724377. 1	XP_033204903.1		K0M31_01 1481-T1	KOX71343.1	MYND-type domain-containing protein	
N0.HOG 0009323	XP_0285 25174.1	XP_031368 067.1	XP_0263009 59.1, XP_0263009 58.1, XP_0263009 57.1, XP_0065602 49.3	XP_0333206 93.1	XP_033177771. 1	XP_020722146. 1	XP_033197784.1		K0M31_01 9774-T1	KOX71531.1	uncharacterized protein LOC110119859	
N0.HOG 0009429	XP_0169 07290.2	XP_006609 518.1	XP_0262982 10.1	XP_0333037 72.1	XP_012237456. 1	XP_012169489. 1	XP_033190083.1	KAF3420998 .1	K0M31_00 2469-T1, K0M31_00 2470-T1	KOX77138.1	LIM/homeobox protein Awh	GO:0005634, GO:0046872, GO:0003677

N0.HOG 0009691	XP_0169 04427.1	XP_031365 192.1	XP_0263018 22.1	XP_0333025 16.1	XP_033180106. 1	XP_020719386. 1	XP_033183844.1		K0M31_00 5632-T1, K0M31_00 5631-T1	KOX77512.1	Uncharacterized protein,Cytoglobin-2	GO:0005344, GO:0019825, GO:0020037, GO:0005833, GO:0005576, GO:0046872
N0.HOG 0009694	XP_0169 05072.1	XP_006610 175.1	XP_0065619 09.1	XP_0332983 08.1	XP_012237621. 1	XP_003397817. 1	XP_033202021.1	KAF3423249 .1	K0M31_01 3427-T1	KOX75389.1	Uncharacterized protein	
N0.HOG 0009730	XP_0169 11234.1	XP_006609 717.1	XP_0262954 25.1	XP_0333011 61.1	XP_024223348. 1	XP_012176274. 1	XP_033184964.1	KAF3425246 .1	K0M31_01 1809-T1	KOX68466.1	Cytochrome c oxidase subunit 6A1, mitochondrial	GO:0004129, GO:0005751
N0.HOG 0009760	XP_0169 22059.1, XP_0285 19992.1	XP_006610 171.1	XP_0262985 43.1	XP_0332981 95.1	XP_012239961. 1	XP_012167339. 1	XP_033201885.1		K0M31_00 8604-T1	KOX74284.1	keratin-associated protein 10-8- like	
N0.HOG 0009765	XP_0285 20355.1	XP_031369 824.1	XP_0065611 49.2, XP_0167664 11.2	XP_0333104 93.1	XP_012248793. 1	XP_012169860. 2	XP_033198641.1		K0M31_00 0254-T1, K0M31_00 0255-T1		PDZ domain-containing protein 2	
N0.HOG 0009897	XP_0169 13059.1, XP_0169 13058.1	XP_006613 452.1	XP_0263002 37.1	XP_0333089 53.1	XP_024227403. 1	XP_020721432. 1	XP_033190734.1		K0M31_01 3097-T1		uncharacterized protein LOC100652324	
N0.HOG 0009898	XP_0169 13182.1	XP_031368 418.1	XP_0065609 11.2	XP_0333095 16.1	XP_012243091. 1	XP_003397027. 1	XP_033195976.1		K0M31_01 8409-T1	KOX70140.1	Uncharacterized protein	
N0.HOG 0009911	XP_0169 17430.1	XP_006614 906.1	XP_0262951 47.1	XP_0333006 64.1	XP_024223318. 1	XP_020724099. 1	XP_033184326.1		K0M31_01 3335-T1	KOX79962.1	uncharacterized protein LOC112212842	GO:0006334, GO:0003677, GO:0000786
N0.HOG 0010049	XP_0169 09912.1	XP_031371 339.1	XP_0032509 47.2	XP_0333161 57.1	XP_024225515. 1	XP_020722860. 1	XP_033198982.1		K0M31_00 8297-T1		Uncharacterized protein	
N0.HOG 0010070	XP_0169 19457.1	XP_031366 306.1	XP_0167708 09.1	XP_0333150 19.1	XP_012241182. 1	XP_003398842. 1	XP_033206859.1		K0M31_00 0208-T1		P17/29C-like protein DDB_G0287399	
N0.HOG 0010098	XP_0285 25478.1	XP_031364 702.1	XP_0167693 45.2	XP_0332983 52.1	XP_024221480. 1	XP_020720473. 1	XP_033202325.1		K0M31_00 8781-T1		uncharacterized protein LOC110119589	

Table SIII. Gene families under significant changes (expanding or contracting) in the nodes of all eusocial corbiculate bees and the node of *M. bicolor*.

Family changes in the node of all eusocial corbiculate			
Gene family ID	Genes annotation in <i>M. bicolor</i>	Gos in <i>M. bicolor</i>	Number of changes
N0.HOG0000029	Fatty acyl-CoA reductase (EC 1.2.1.84)	GO:0102965,GO:0016021,GO:0080019,GO:0006629	+1
N0.HOG0000059	null		-1
N0.HOG0000082	Putative odorant receptor 13a,Odorant receptor,Putative odorant receptor 92a	GO:0005549,GO:0016021,GO:0004984,GO:0007165,GO:0005886	+1
N0.HOG0000120	Lysophospholipase (EC 3.1.1.3) (EC 3.1.1.4) (EC 3.1.1.5) (Phospholipase A2) (Phospholipase B/lipase) (Phospholipase B1, membrane-associated, membrane-associated) (Triacylglycerol lipase),Phospholipase B1	GO:0004806,GO:0006629,GO:0016324,GO:0016021,GO:0102568,GO:0102567,GO:0004622,GO:0004623,GO:0004620,GO:0102545	+1
N0.HOG0000170	Maltase 2,Alpha-glucosidase	GO:0003824,GO:0016021,GO:0005975	+1
N0.HOG0000235	Elongation of very long chain fatty acids protein (EC 2.3.1.199) (Very-long-chain 3-oxoacyl-CoA synthase)	GO:0006633,GO:0016021,GO:0102337,GO:0102336,GO:0102338,GO:0102756	+1
Family changes in the node of <i>M. bicolor</i>			
N0.HOG0000030	Zinc finger imprinted 3,Uncharacterized protein,Zinc finger protein,Zinc finger protein CKR1,longitudinals lacking protein isoform X37, isoforms A/B/D/L,Longitudinals lacking protein		+4
N0.HOG0000056	Uncharacterized protein	GO:0005549,GO:0016021,GO:0005886,GO:0004984	-2
N0.HOG0000108	Uncharacterized protein		+1
N0.HOG0000135	Uncharacterized protein	GO:0016021	+1
N0.HOG0000222	Acyltransferase (EC 2.3.1.-)	GO:0016021,GO:0016747,GO:0005789	+2
N0.HOG0000225	Putative odorant receptor 13a,Odorant receptor	GO:0005549,GO:0016021,GO:0004984,GO:0007165,GO:0005886	+4
N0.HOG0000235	Elongation of very long chain fatty acids protein (EC 2.3.1.199) (Very-long-chain 3-oxoacyl-CoA synthase)	GO:0006633,GO:0016021,GO:0102337,GO:0102336,GO:0102338,GO:0102756	-2
N0.HOG0000342	organic cation transporter protein-like	GO:0016021,GO:0022857	+1
N0.HOG0000404	LOW QUALITY PROTEIN: uncharacterized protein LOC100746456,LOW QUALITY PROTEIN: uncharacterized protein LOC117206071,Fibrillin-2	GO:0005634,GO:0005576,GO:0005509,GO:0005520,GO:0016021,GO:0008270	+2
N0.HOG0000561	Ig-like domain-containing protein (Fragment)		+1
N0.HOG0000685	null		+1
N0.HOG0000868	Orexin receptor type 2,allatotropin receptor isoform X1 (allatotropin receptor precursor)	GO:0007631,GO:0016021,GO:0005886,GO:0004983,GO:0016499	+1

Table SIV. Results of the selection tests of the genes associated with the bee venom using the Meliponini ancestral node as the foreground node. The *branch* model tests compare the entire gene evolution in the foreground (f) node to the remaining background (b) nodes, while the branch-site models search for codon selection in the foreground node. Significant values are highlighted in bold.

Orthogroup	Gene	<i>branch</i>					<i>branch-site</i>						
		mean ω (<i>M0</i>)	ω (f)	ω (b)	p-value <i>M0</i>	p-value <i>b-neut</i>	mean ω (<i>M1</i>)	codon positions under selection	number of codons under selection	p-value <i>M1</i>	p-value <i>bsA1</i>	p-value <i>bsA-M1</i>	p-value <i>bsA1-bsA</i>
N0.HOG0001460	<i>Venom carboxylesterase-6-like</i>	0,199	0,179	0,201	0,742	0,000	0,253	NA	0	1,000	1,000	0,974	0,822
N0.HOG0004097	<i>Venom dipeptidyl peptidase 4</i>	0,181	0,302	0,172	0,016	0,000	0,283	62, 85, 316	3	0,000	0,638	0,000	0,070
N0.HOG0004530	<i>Venom serine carboxypeptidase</i>	0,155	0,169	0,154	0,742	0,000	0,245	NA	0	1,000	1,000	0,980	1,000
N0.HOG0006230	<i>Venom peptide isomerase</i>	0,136	0,075	0,146	0,151	0,000	0,195	NA	0	1,000	1,000	1,000	1,000
N0.HOG0005126	<i>Clq-like venom protein</i>	0,221	0,090	0,258	0,151	0,000	0,294	NA	0	1,000	0,638	0,356	0,151
N0.HOG0000658	<i>Cysteine-rich venom protein</i>	0,156	0,018	0,179	0,000	0,000	0,205	NA	0	1,000	1,000	1,000	1,000
N0.HOG0004893	<i>Venom protease-like</i>	0,238	0,268	0,235	0,742	0,000	0,293	NA	0	1,000	0,647	0,471	0,245
N0.HOG0002840	<i>Venom allergen 5</i>	0,199	0,279	0,194	0,742	0,046	0,294	NA	0	1,000	1,000	1,000	1,000
N0.HOG0006648	<i>Toxin 3FTx-Lei1</i>	0,200	0,120	0,218	0,742	0,030	0,224	NA	0	1,000	1,000	1,000	1,000
N0.HOG0005748	<i>Venom serine protease</i>	0,296	0,333	0,292	0,742	0,000	0,403	NA	0	0,225	0,638	0,041	0,174
N0.HOG0004751	<i>Hyaluronidase isoform x1</i>	0,151	0,141	0,153	0,742	0,000	0,205	NA	0	1,000	0,647	0,576	0,294

Supplementary Figures

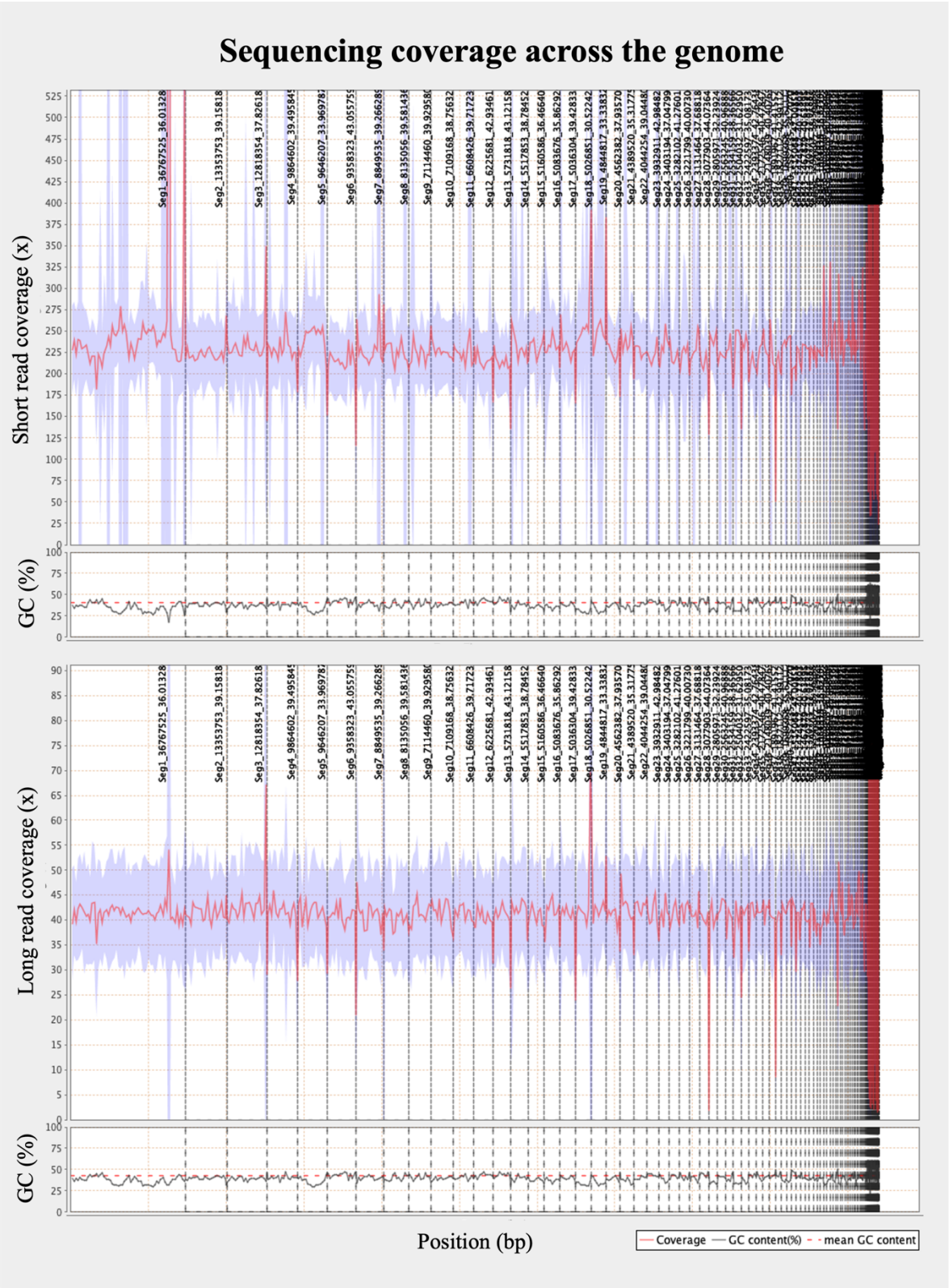


Figure S1 Sequencing coverage and mapping quality of the alignment of short reads (upper plot) and long reads (lower plot) against the final genome assembly of *M. bicolor*. Below each coverage plot the mean GC content is shown and 400 windows were used to represent the genome regions. Each scaffold is represented by a segment that are separated by dotted lines.

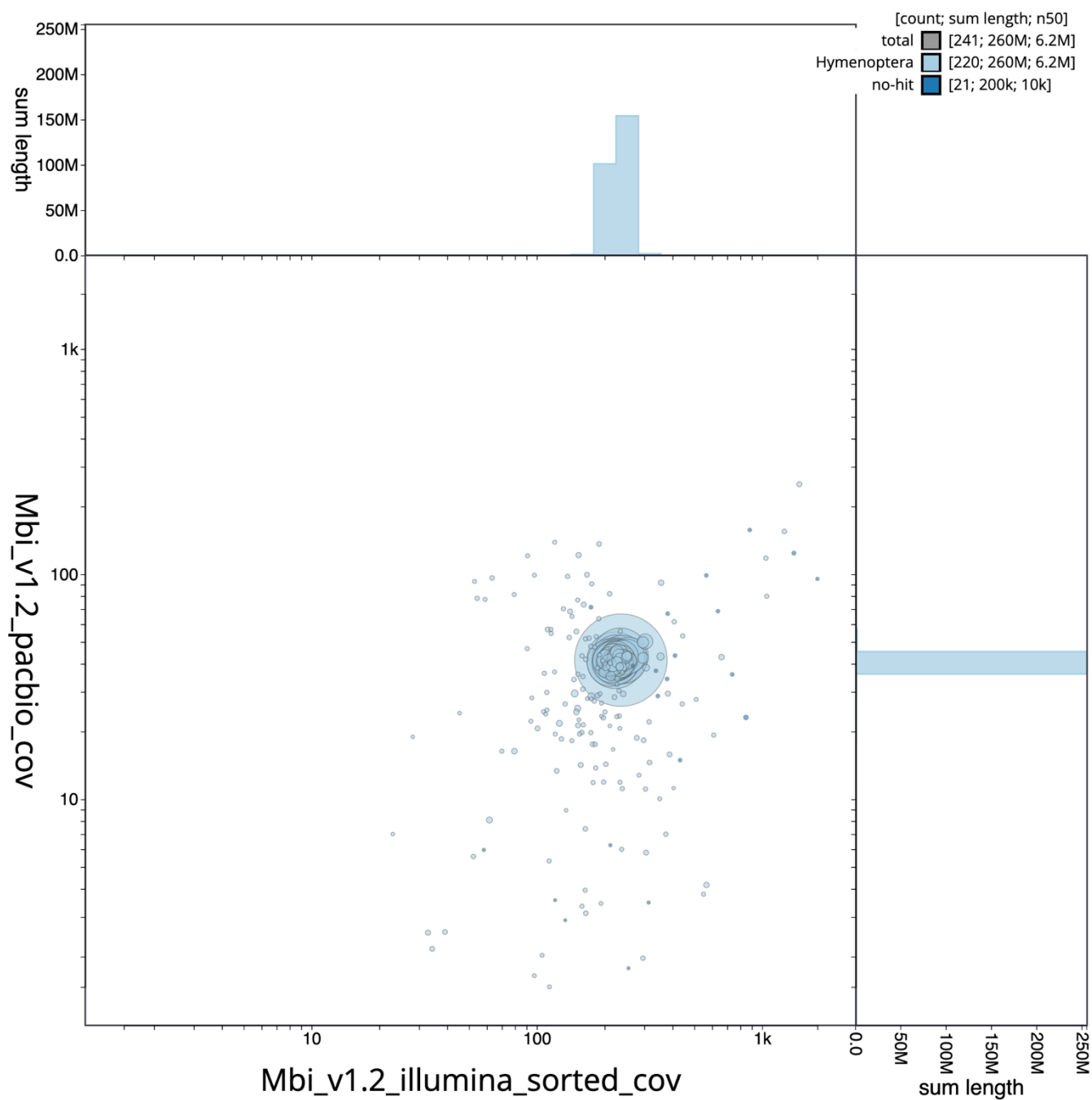


Figure S2 Sequencing coverage of short reads (x-axis) against long reads (y-axis) showing that coverage of the assembled scaffolds was mostly congruent between the two sequencing technologies as we removed outlier scaffolds (i.e., scaffolds with coverage <1x) during quality trimming steps. Circle size represents scaffold size.

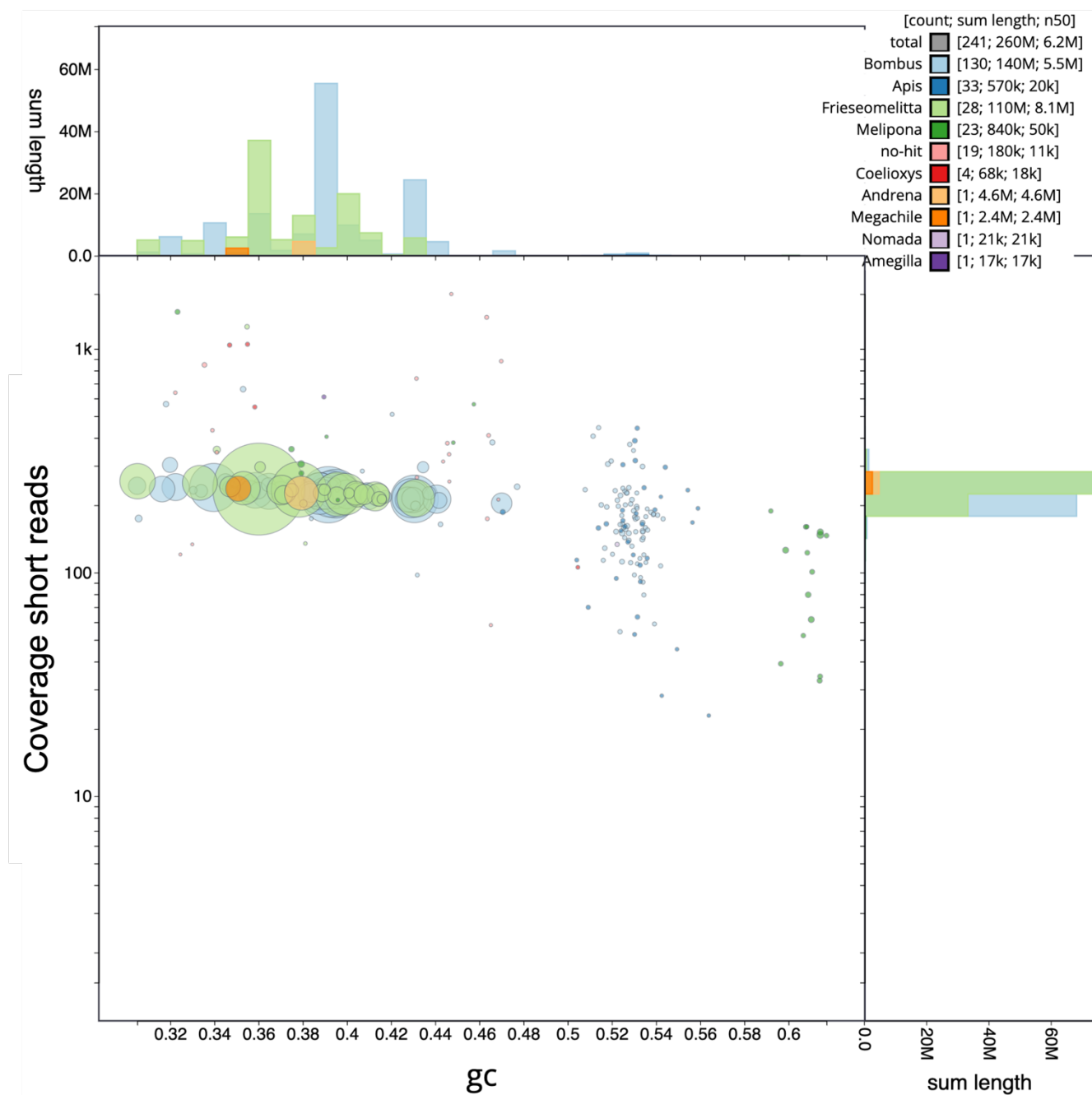


Figure S3 Blob plot of short reads base coverage against the GC proportion of scaffolds in the genome assembly of *Melipona bicolor*. Sequences are colored by matching genus sequences in databases. Circle size represents scaffold size. The histograms show the sequence length distribution in each axis.

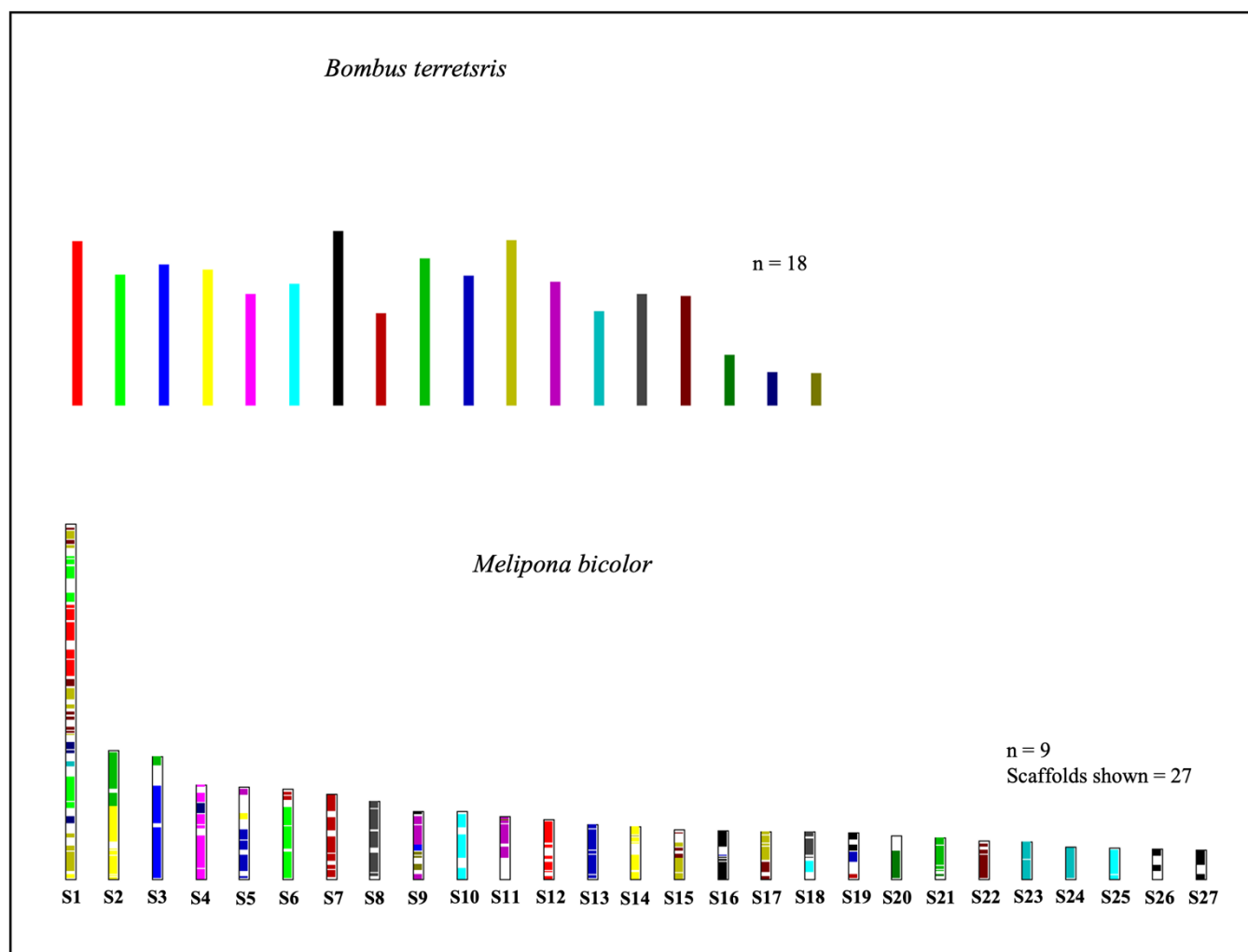


Figure S4 Genomic synteny inferred between the genome assembly of *B. terrestris* (upper bars), and *M. bicolor* (lower bars). The same colors represent syntenic chromosome regions in the *B. terrestris* genome. For *M. bicolor* the largest 27 scaffolds are presented instead of the chromosomes, these represent 80% of the total genome assembly.

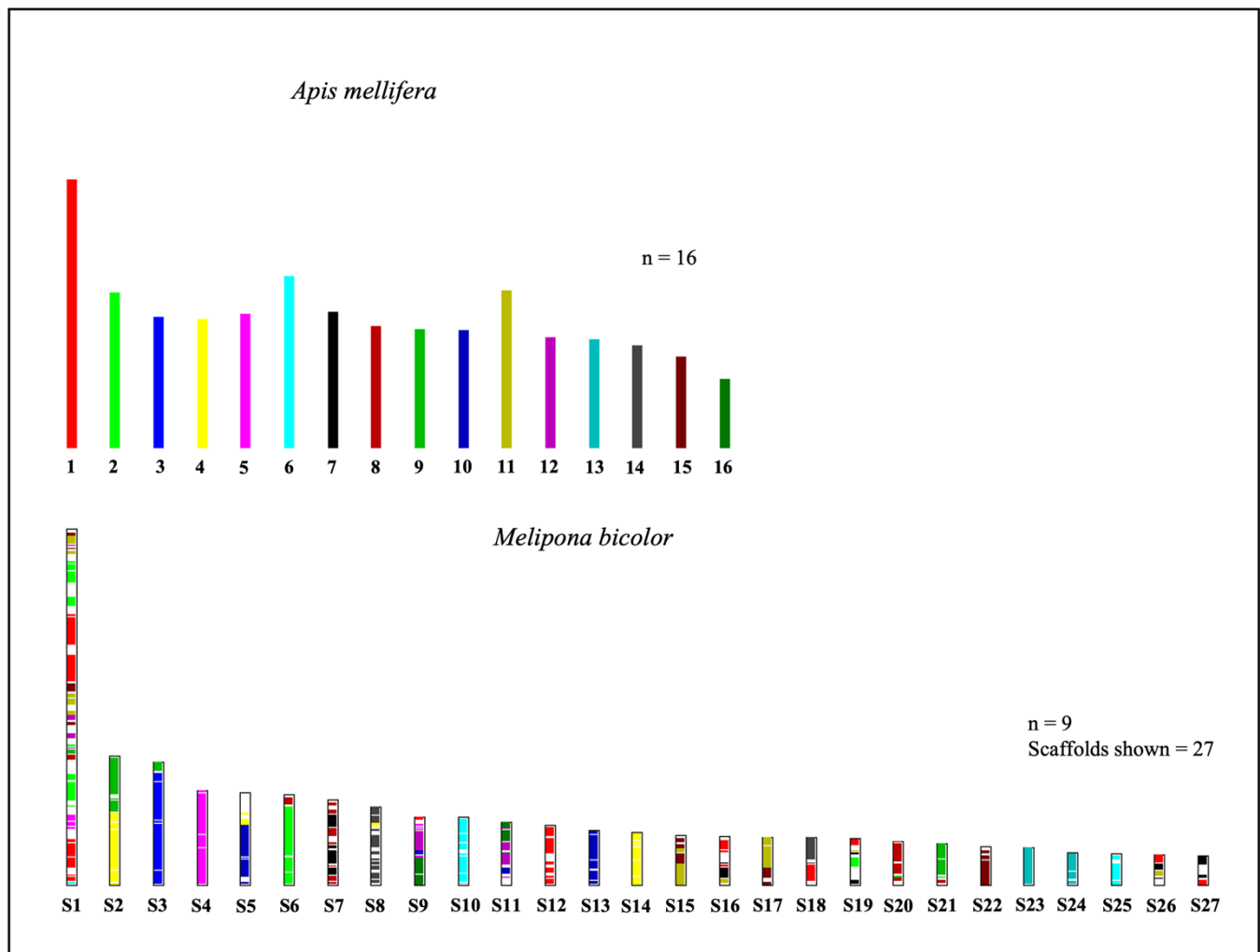


Figure S5 Genomic synteny inferred between the genome assembly of *A. mellifera* (upper bars), and *M. bicolor* (lower bars). The same colors represent syntenic chromosome regions in the *A. mellifera* genome. For *M. bicolor* the largest 27 scaffolds are presented instead of the chromosomes, these represent 80% of the total genome assembly.