Supplementary Data for

Genome-wide profiling of forum domains in *Drosophila melanogaster*

Nickolai A. Tchurikov,[1]* Olga V. Kretova,[1] Dmitri V. Sosin,[1] Ivan A. Zykov,[2] Igor F. Zhimulev,[2] Yuri V. Kravatsky[1]

[1] Engelhardt Institute of Molecular Biology Russian Academy of Sciences, Moscow, 119991, Russia, [2] Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, 630090, Russia

*****To whom correspondence should be addressed. Tel: 7-499-1359753; Fax: 7-499-1351405; Email: tchurikov@eimb.ru

Supplementary Figure 1

Pulse-field gel separation of forum DNA sample isolated from Schneider 2 cells as described in Materials and Methods. The 1% agarose gel was run at switching time of 25 sec. Lambda ladder was used as a marker. The lanes corresponding to the lambda ladder (Lambda) and forum DNA sample from Schneider 2 cells (S2) were scanned using "Quantity One" software (Bio-Rad). The lengths of bands in the lambda ladder are indicated in kb.

Supplementary Table 1

Nucleotide sequences of 45 random RAFT clones. Sequences corresponding to mobile elements are shown in red. Sau3A site is shown in bold. The nucleotides at the fragmentation sites are shown in blue. Some sequences were cloned from 2 to 5 times (indicated as the numbers of reads in the Annotation).

Supplementary Table 2

Localization of the RAFT probe hybridization sites on 3L and 3R chromosome arms. The coincidence with SuUR protein localizations in the wild-type line (2 copies of *SuUR*) and transgenic line (4 copies of *SuUR*) are shown as well as the correspondence of the RAFT hybridization sites with the known regions of the late replication (1).

Supplementary Figure 2

Overviews of 2L and 2R arms, X chromosome (from section 11) and chromosome 4. Positions of FT along the chromosomes are shown in red (Forum). The binding profiles of Pc and the distribution of H3K27me3 marks are shown (2,3). TF – transcription factors binding sites are shown in blue (4).

Supplementary Figure 3

Organization of 715 kb forum domain possessing 5S rRNA gene cluster in 2R chromosome. Positions of FT along the chromosomes are shown in red (Forum). The bracket indicates position of the 5S rRNA gene cluster. The binding profiles of Pc and the distribution of H3K27me3 marks are shown (2,3). TF – transcription factor binding site is shown in blue (4). Positions of mobile elements are shown in dark blue.

Supplementary Figure 4

Forum domains can possess clusters of coordinately expressed genes. Regions from the 2L, 2R, and X chromosomes and all of chromosome 4 are shown. The red brackets indicate the forum domains possessing clusters of actively transcribed genes in S2 cells, as visualized by IGB using S2 poly(A) RNA unique mapper data (http://modencode.oicr.on.ca/fgb2/gbrowse/fly/). The blue brackets indicate the forum domains containing the silent or weakly expressed clusters of genes. The red squares along the FT line show the positions of the mapped forum termini that delimit the forum domains.

The genes located in the active large forum domain shown at the very top of the figure have different molecular functions and are involved in different biological processes. For example, genes specifying ribosomal proteins RpL7 and RpS27A, proteasome protein (Pros35), translation elongation factor (eEFldelta), ribonucleoside diphosphate reductase large subunit (RnrL), mitochondrial ribosomal protein S7 (mRpS7), and protein retaining receptor (KdelR) have different functions but are located in the same forum domain and are actively co-transcribed in S2 cells. The data on early origins of replication correspond to the track with references to MacAlpine_K_WIG_934_939_941:70001.

Supplementary Figure 5

Relationships between FT, regions containing sequences of small non-coding 19-24 nt RNAs (snRNAs), Ago1 and Ago2 binding sites, nucleosomes, and histone modifications in the region of alfagamma-element in S2 cells. The Modencode genome browser was used (http://modencode.oicr.on.ca/fgb2/gbrowse/fly/) with the loaded tracks for snRNAs ["S2-DRSC (Rubin)", smallRNA RNA-Seq (Lai project, Lai subgroup), GEO accession number GSM361908]; Ago1 and Ago2 [S2 cells Ago1-HA Immunoprecipitation (Lai project, Hannon subgroup), GEO accession numbers GSM280088 and GSM280087, respectively]; and FT (RAFT652). The data for nucleosomes and histone modifications correspond to the tracks referred to as "HenikoffNUCL:70001" and "Karpen_HISMODS_S2:70001", respectively.

Supplementary Figure 6

Chromatin landscape in the region of 2L chromosome. The Modencode genome browser was used (http://modencode.oicr.on.ca/fgb2/gbrowse/fly/) with the loaded tracks for snRNAs [S2-DRSC (Rubin) S2 smallRNA RNA-Seq (Lai project, Lai subgroup), GEO accession number GSM361908]; Ago1, Ago2 [S2 cells Ago1-HA Immunoprecipitation (Lai project, Hannon subgroup), GEO accession numbers GSM280088 and GSM280087, respectively]; and FT (RAFT652). The data for nucleosomes and histone modifications correspond to the tracks referred to as "HenikoffNUCL:70001" and "Karpen_HISMODS_S2:70001", respectively. The data for 19 chromatin proteins correspond to the track with the following reference: Karpen_CHROMPROTS_S2:70001 "Chromatin Proteins ChIP-chip (Karpen project, Pirrotta subgroup)".

Supplementary Figure 7

The correlation between forum domains and syntenty conservation across *Drosophilids*. The windows from the http://flybase.org/cgi-bin/gbrowse/dmel/ browser are shown. The genes located in *Drosophila menalogaster* forum domains have the same orthologous genes in the *Drosophila* species *D. ananassae, D. erecta, D. grimshawi, D. mojavensis, D. persimilis, D. pseudoobscura, D. sechellia, D. simulans, D. virilis, D. willistoni,* and *D. yakuba* and also in the non-*Drosophilidae* orthologs *A. pisum, A. aegypti, A. gambiae, B. mori, B. floridae,* and *P. humanus*. The regions correspond to the 3R regions shown in Figure 7 (A and B).

Supplementary Tables 3 and 4 correspond to the RAFT652 data that can be uploaded into IGB (.gff file) or into the Modencode and FlyBase Browsers (.txt file).

Supplementary Figure 8

This phylogram tree was built using http://www.ebi.ac.uk/Tools/clustalw2/help.html#tree software. The 95 bp FT sequence was collected from the *Drosophila melanogaster* RAFT deep sequencing data (detailed analysis of the data will be published separately). The sequence corresponding to the right FT of the forum domain containing the histone gene cluster (around 2L:21,750,000 coordinate, Figure 5) was used for the BLAST search. The BLAST hits from different genomes were selected and analyzed.

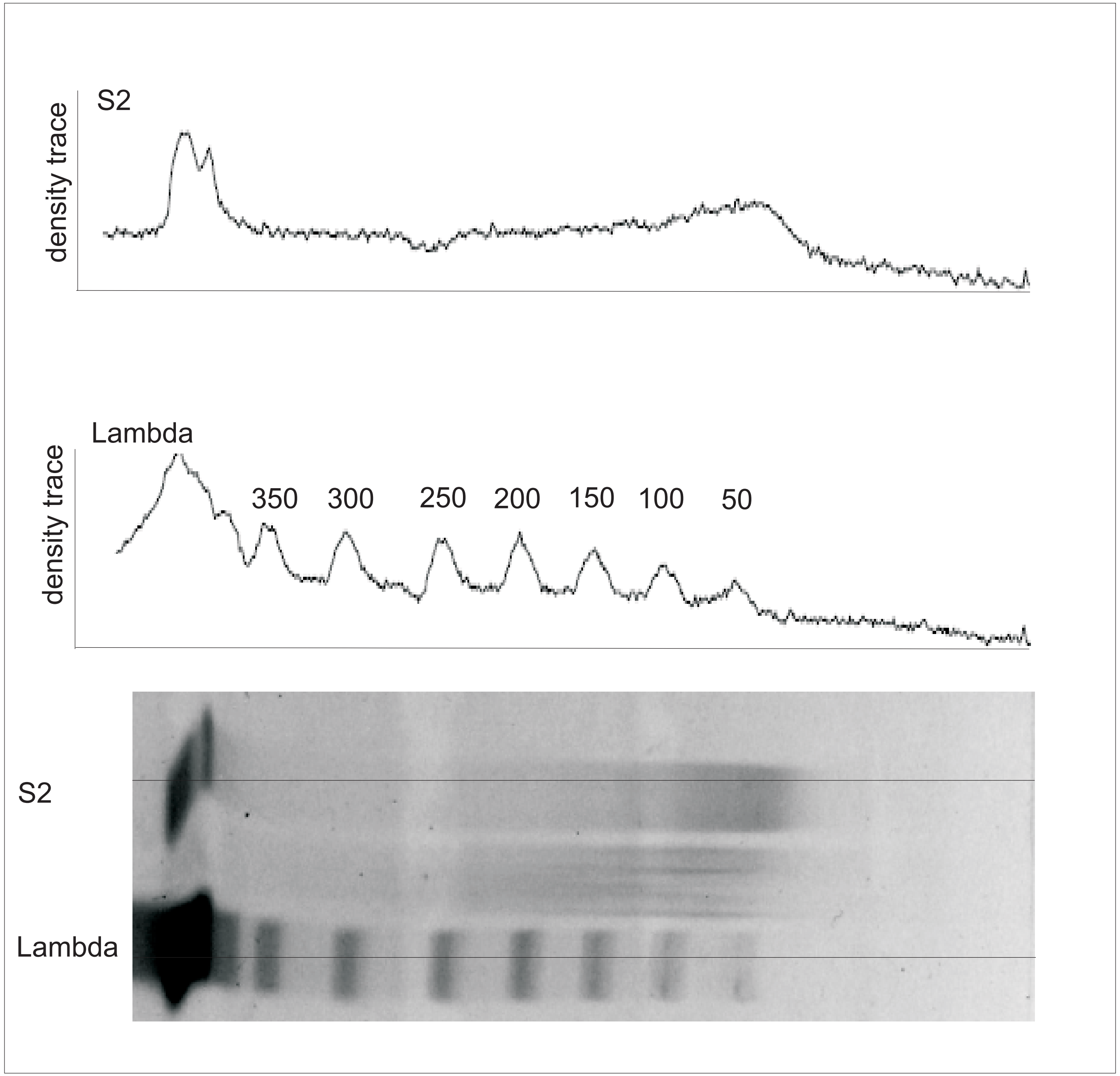
Supplementary Figure 9

The relationship between forum domains and color domains (5). (A) The 1 Mb region from the 2L chromosome (the region is shown in Figure 1 in Filion et al., 2010). (B) The relationship between forum domains and BLACK, GREEN, and YELLOW domains in the 9.68 Mb region of the 2L chromosome.

References:

1. Zhimulev,I.F., Belyaeva, E.S., Makunin, I.V., Pirrotta,V., Volkova,E.I., Alekseyenko, A.A., Andreyeva,E.N., Makarevich,G.F., Boldyreva,L.V., Nanayev,R.A., Demakova, O.V. (2003) Influence of the SuUR gene on intercalary heterochromatin in *Drosophila melanogaster* polytene chromosomes. *Chromosoma*, **111**, 377-398.
2. Schwartz,Y.B., Kahn,T.G., Nix,D.A., Li,X-Y., Bourgon,R., Biggin,M., Pirrotta,V. (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genet*., **38**, 599-727.
3. Kwong,C., Adryan,B., Bell,I., Meadows,L., Russel,S., Manak,R., White,R. (2008) Stability and dynamics of Polycomb target sites in *Drosophila* development. *PLoS Genet.* 4: 9,e1000178.
4. Bergman,C.M., Carlson,J.W., Celniker,S.E. (2005) *Drosophila* DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*. *Bioinformatics*, **21**, 1747-1749.
5. Filion,G.J., van Bemmel,J.G., Braunschweig,U., Talhout,W., Kind,J., Ward,L.D., Brugman,W., de Castro,I.J., Kerkhoven,R.M., Bussemaker,H.J., van Steensel,B.

(2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*, **143**, 212-224.



Supplementary Figure 1, Tchurikov et al.

Supplementary Table 1. Nucleotide sequences of random RAFT clones.

#	Clone	Sequence, 5'-3'	Length, bp	Chromosome, coordinate	Band	Annotation
1	22	GATTTTATTCGGGAATAAAAGGCGCCACACGGTTAACTTCGTCTGCAGGA**GATC**	54			gypsy, LTR-element
2	24	**GATC**AAAAGATTTTTATTAACCAATAGTCTAAAACAAGAAATAAGTTTGTATAAACTTTCAAGCATA	67	3R, 25589627	99B	
3	27	**GATC**GCACCACCTAAGGAAGTTGCACGCTGCACTGTCGCGACGCACGGAACAAAACCACGAAGGATATATGCTCACGTTGAAGGACAAGGAAAAACACAAAAAGCTAAAATTAACGGGACGTGAACGAAACAG	133	3R, 27808031	100D	3 reads
4	28	CAAATCAGTTAGGAGTGGATACTAAAGCTTAATATACGGTTTAAGGAAAATGCACATAGGGCAAAATAATTCCAATATTTAATAAAGTTCACATTTTAAGCACCCTTCTAAG**GATC**	116	2L, 14562985	35B	
5	30	GGAATAGGTCCGGAGGTTTACACATTTTTAGTCATTCTGAA**GATC**	45			invader, LTR-element
6	31	GCATTTGTTACAAACTCACATCTTTGCTTTCAGTTTCGCAACCAGTACGATAGCGATGTTACTGTGTGGAGTCCGCAGGGACGTCTCCACCAGGTGGAGTACGCAATGGAGGCCGTGAAACTTGGCACCGCAACCGTGGGCCTGAAAAATAAAGATTACGCTGTACTTGTCGCCCTGTGTAAGCCGACATCGGAGCTATCGGATACCCAGCGCAA**GATC**	219	2L, 10221481	31B	2 reads
7	33	GTACAATGATACGAAAGCGAAAGACAATTGAAGAAC**GATC**	40	3L, 23563585	80F	*Ago3*, intron
8	38	AGTAGAATATCATCAAGGGGAAAGATAATCATGTCGCAGACGCACTTGATCCCGANTAACAATCGGAGAACTTACAGCNTTAAATAGACAGATACTAAAGGTAACAACAA**GATC**	114			mgd1, LTR-element, 2 reads
9	44	ACAAATTGGAAACCATTCGAGTTGACAA**GATC**	32			3 reads, U
10	51	**GATC**GTTAGCACTGGTAATTAGCTGCTTAAAACAGTTATTGTTACATCTATGTGACAATTTTTAGCCAAGTTATAACGAAAATTTGGTTGTAAATATCAACATTTTGGCAGAATCTGTTTTTCCAAATTTCGGTCATCAAATAATCATTTATTTTGCCACAACATAAAAAAATAATTGTCTGAACATGGAATGTCATA	197	X, 9319496, 9366693, 9390221	8D	3 reads
11	56	**GATC**AGTGCGACTTTATGTAATCGGCACAATTCTGATTTCCCACACAGGATGCTGGTCAAGAGCTACGGACTTTTCTTCTTGGGAGTGCGTTTGG	95	2R, 1645462	41F	
12	57	AGAATATCTCCGTTGGAGGGCATGCT**GATC**	30	3R, 4848010	85C2	
13	58	CCATCCTACCCTCAGCTTGACACGCTT**GATC**	31			Rt1c, LINE
14	59	**GATC**AGGAACTTCATGTTGAATTTGGGCT	29	3L, 6131192	65A	
15	60	TATTTTTATTTATTTGCAATGTCAAA**GATC**	30	2L, 5747928	25F	
16	62	GAGAATCCGAGCACTTTGTTACAAAAATGTACAGAGAAAAAGCTCTGGATACTTTTATCAGAGGCCTTCGAGGAGATTTACCTCGCCTTTTGGCAATAAAAGAGCCAGCT**GATC**	114			HMS-Beagle, LTR element, 5 reads

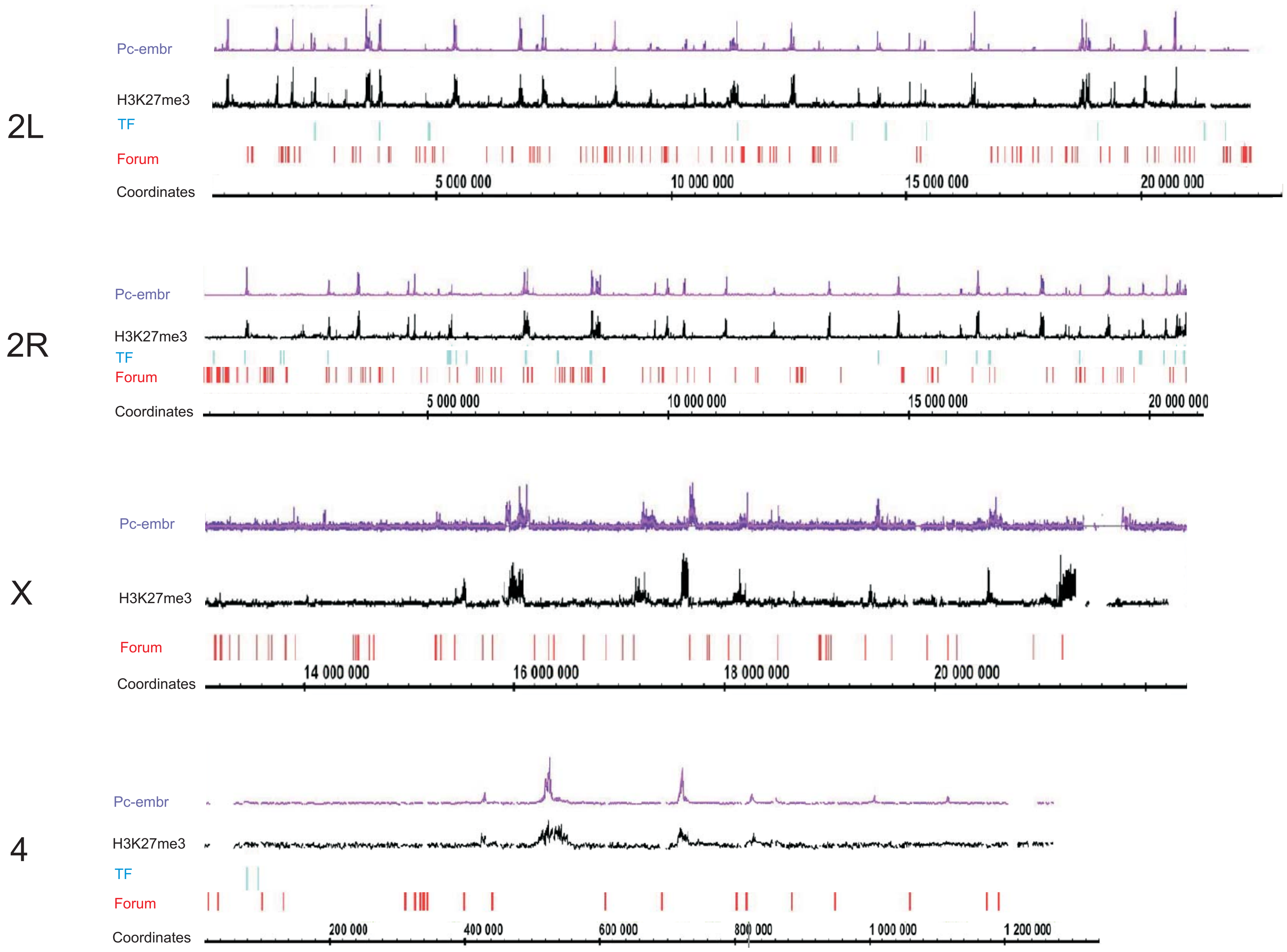
17	63	GGTATGCAATTATGGTTTCGGGTTCTATATACGTTAGTTTAGCG**GATC**	48	2L, 5315639	25D	2 reads
18	64	GNCCCCCATTATCTGGAATGTCCGTTGGAGACGGCCGTCTAGGCTGAAATCCGGACTAACACACAAGACCCG**GATC**	76			opus, LTR element
19	71	**GATC**GGCTTATGTATAAGCGTTAGATCGTATTTCATTGGCTGGAGTGTTTATGTAAGTGCATGTGGTCTTTGATGTGAATTATGTTTATTTGTGTTTGCAACAAAGTGTCACAATGTATGTACGTTACTTATATATATGTAGAGTAGAGTTATAGTGAGGGTAGATACACTACACCTCCCTTTTTAGAGAAATGACGTAATATA	204			Stalker , LTR element, 3 reads
20	76	AGCGCATAACTTCTCCGCTGATGCTGCGTTCAATATGCCCAAATAT**GATC**	50			17.6, LTR element
21	79	TGAGTTGAGGGAATCCTGGTCAAGGAATGTGATTGAACATACTACTAG**GATC**	52	3L, 4211456	64A	
22	82	GTATAACAAAACGAGAAATCCTATCGACCATAGCACGTTTATTT**GATC**	48			diver, LTR element
23	85	GGGGTTTGGTGGTGGTGTTGTTATGTACGCCGAGTAGAGCGCGTTCGTGTGTCAGAAATAGGGTGACTCAAGATGTCAGCAAGCTTTTCCTTCTAAAGAAATCGCCATGGCTCCATAAG**GATC**	123	3L, 22269778	79D	
24	87	TGCTCGAGTTGGGGATTTACGACCGGGACGGGGTCTAGGGCTCGCACTGCGACTGTCG**GATC**	62	3L, 21547755	78D	
25	88	ATGCTCGGCAATTAACCGAGAGTGCAGCTAACGGATAATTACACGGCGGAACTACCTAAGCCACACTAAGTCTGATAAGAGATTTTAGCCCTGATTGA**GATC**	102	3R, 13206611	90A3	
26	94	**GATC**AGTTGACATGTTCCTAGCCACAATGCATAATGGTTGCCCACACTCATAACCGGGGGTAATCTGCATAAACAGAGGCTATATCAAGTGCCATGCGCCGGCGGGTTCGTCAATCTCCTCCCGGGCCGGTTCCTTCACTCAAAATAGTTCATAATTG	158	3L, 16908874	73D	
27	95	CCAAGCGTTCATAGCGACGTCGCTTTTT**GATC**	32			18S rRNA gene, 2 reads
28	97	ATGCATAACCGTCGGGGTCTTCA**GATC**	27			1731, LTR element
29	99	**GATC**AAAACGGGGTAAAAAAGAAACGACTAGTTATCGATTTTAGGAAACTAAACCAGAGAACAGTCGACGATA	73	3R, 1830300	83D	

Supplementary Table 2. Comparison RAFT probe hybridization sites in polytene chromosomes with the localization of the SuUR antibodies binding sites and with the regions of the late replication on 3L and 3R chromosomes (Zhimulev et al., 2003).

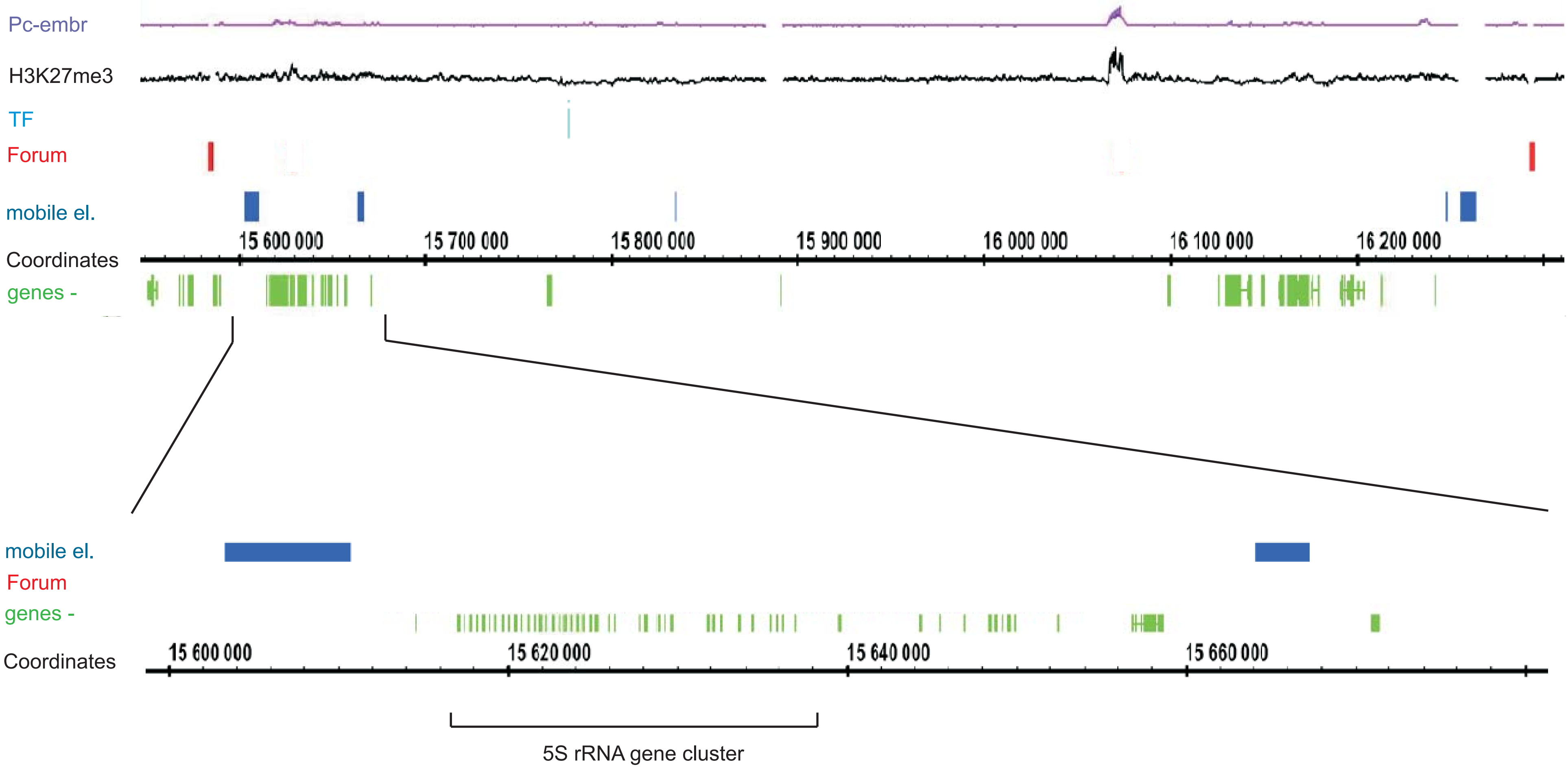
№	Regions of hybridization of RAFT probe	Coincidence (+) with SuUR localization in normal line (2 copies of *SuUR*)	Coincidence (+) with SuUR localization in line possessing 4 copies of *SuUR*	Coincidence (+) with the regions of the late replication
1.	61C	–	+	+
2.	62B	–	–	+
3.	62C	–	+	+
4.	62D	–	+	+
5.	64B	–	–	+
6.	64C	+	+	+
7.	64D	–	+	+
8.	64E	–	+	+
9.	65A	+	+	+
10.	65D	+	+	+
11.	65F	–	+	+
12.	66A	–	+	+
13.	66E	–	–	–
14.	67A	+	+	+
15.	67C	–	+	–
16.	67D	+	+	+
17.	68A	–	+	+
18.	68C	–	–	–
19.	68E	+	+	+
20.	69E	–	+	–
21.	70A	–	+	+
22.	70C	+	+	+
23.	70D	–	+	+
24.	71C	+	+	+
25.	72A	–	+	+
26.	73A	+	+	+
27.	73C	–	–	–
28.	74A	+	+	+
29.	75A	+	+	+
30.	75C	+	+	+
31.	75E	–	–	–
32.	76A	–	+	+
33.	76B	–	+	+
34.	76C	–	–	–
35.	77A	–	–	–
36.	77E	+	+	+
37.	78E	–	–	–
38.	79A	–	–	–

39.	79C	–	–	+
40.	79D	–	+	+
41.	79E	+	+	+
42.	80A	+	+	+
43.	80B	+	+	+
44.	80C	+	+	+
45.	81F	+	+	+
46.	82C	–	+	–
47.	83A	–	–	–
48.	83D	+	+	+
49.	83E	–	+	+
50.	84AC	+	+	+
51.	84D	+	+	+
52.	85AC	–	+	+
53.	85D	–	+	–
54.	85F	–	+	–
55.	86C	+	+	+
56.	86D	+	+	+
57.	86E	–	+	+
58.	86F	–	–	–
59.	87D	+	+	+
60.	87E	+	+	+
61.	87F	–	+	+
62.	88B	–	+	–
63.	88C	–	–	–
64.	88D	–	–	–
65.	88E	+	+	+
66.	89A	–	+	+
67.	90A	–	+	+
68.	90B	–	+	–
69.	90D	–	+	+
70.	91A	–	+	+
71.	91D	–	+	+
72.	92A	–	+	+
73.	92B	–	+	+
74.	92D	+	+	+
75.	93A	–	–	–
76.	93E	–	+	+
77.	94A	+	+	+
78.	94B	–	–	–
79.	94C	–	–	–
80.	94D	+	+	+
81.	95A	+	+	+
82.	96C	–	+	+
83.	97A	+	+	+
84.	97D	–	+	–
85.	98A	–	+	+
86.	98B	–	+	+

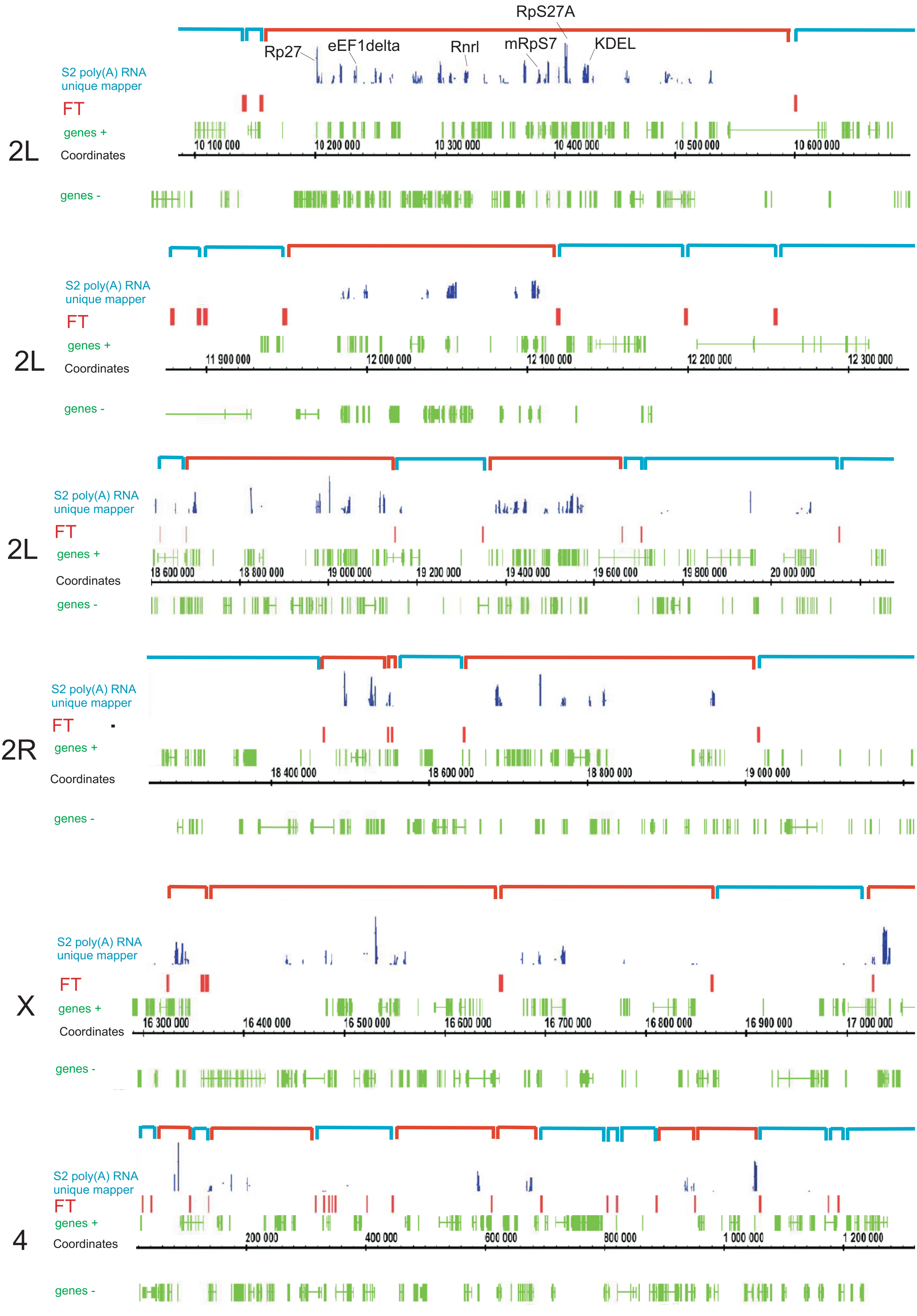
87.	98C	+	+	+
88.	98D	–	–	+
89.	98E	–	+	+
90.	98F	–	+	+
91.	100A	+	+	+
92.	100B	+	+	+
93.	100F	+	+	–
Coincidence (%)		37.6	79.6	74.2



Supplementary Figure 2, Tchurikov et al.



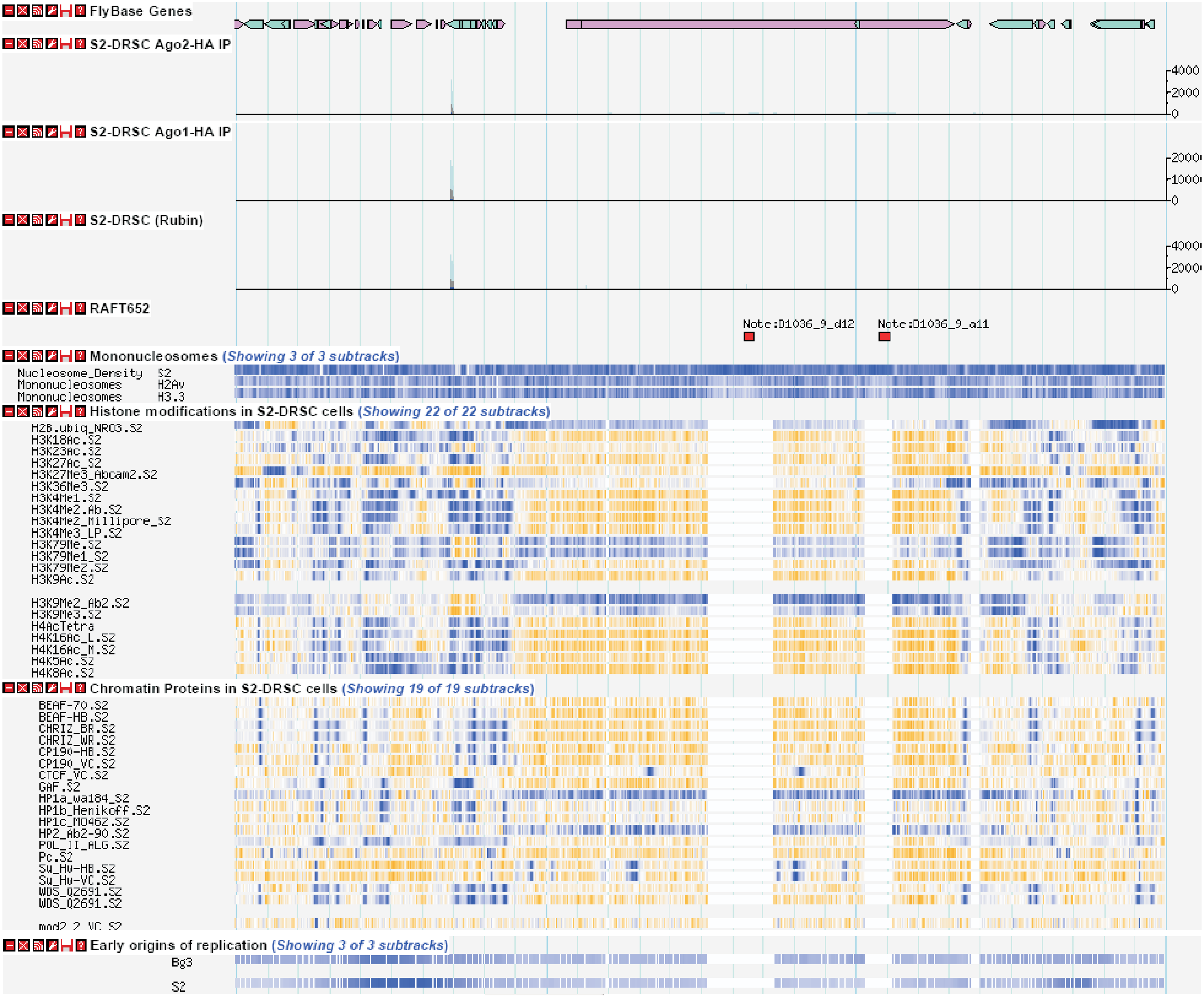
Supplementary Figure 3, Tchurikov et al.



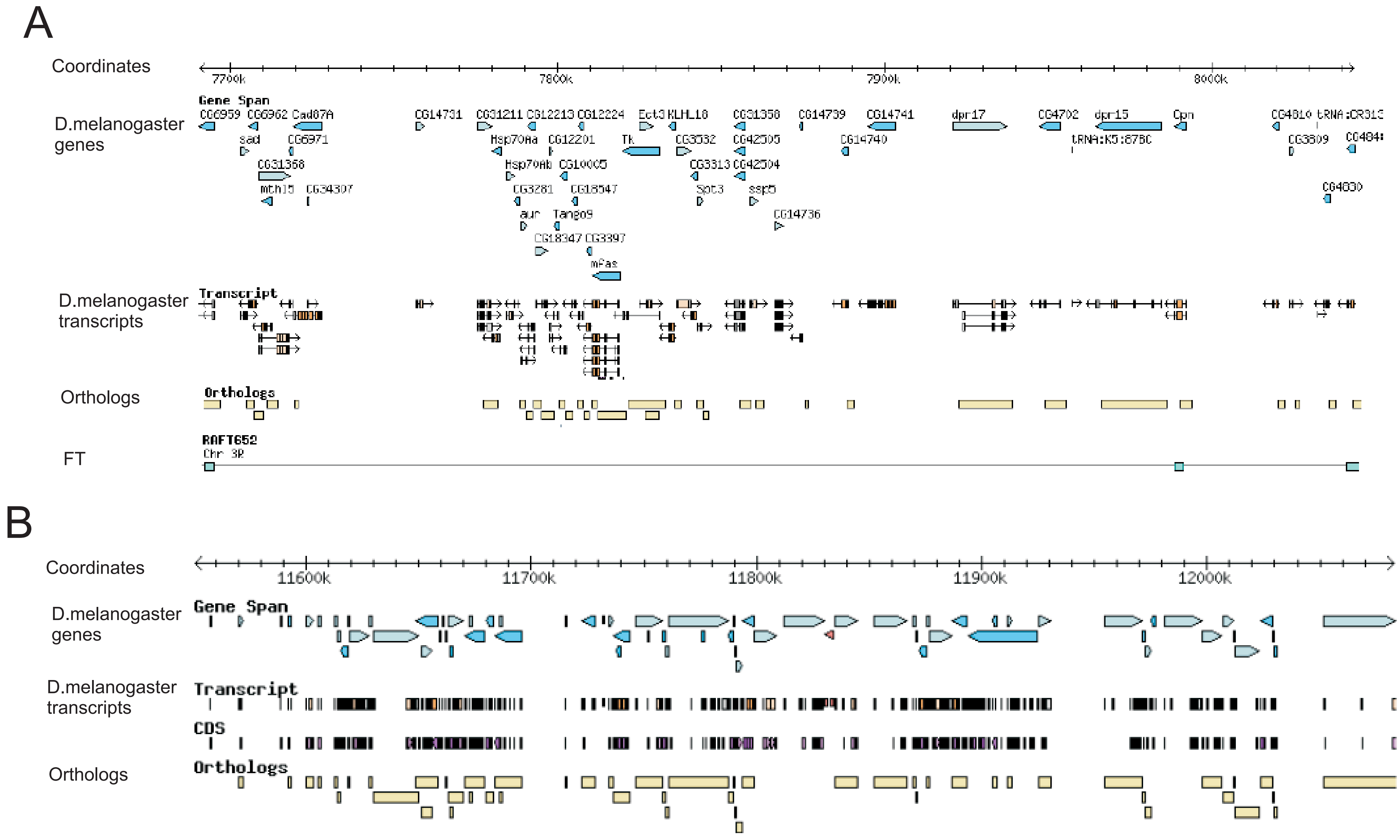
Supplementary Figure 4, Tchurikov et al.

Supplementary Figure 5, Tchurikov et al.

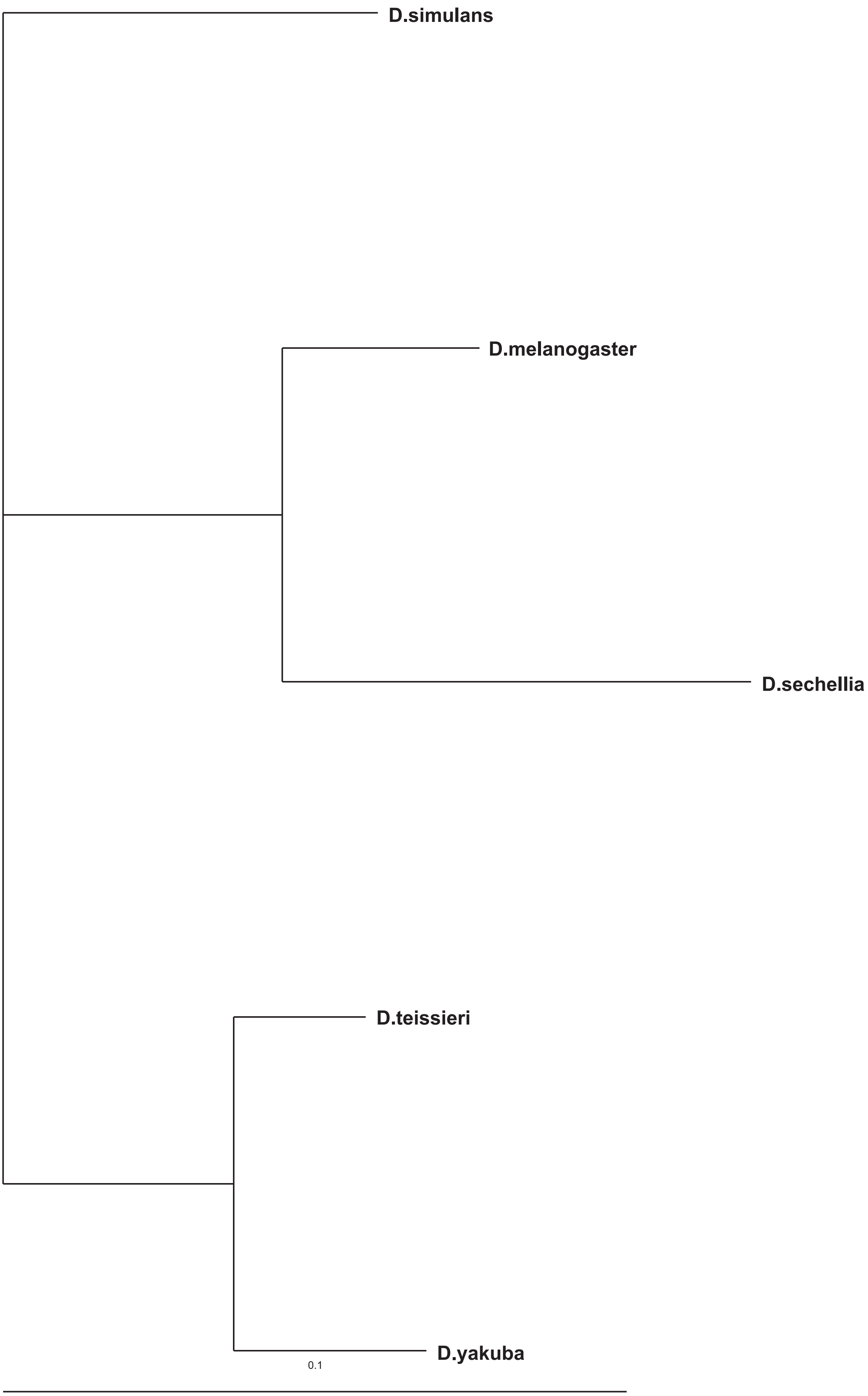
2L:19,500,000..19,800,000



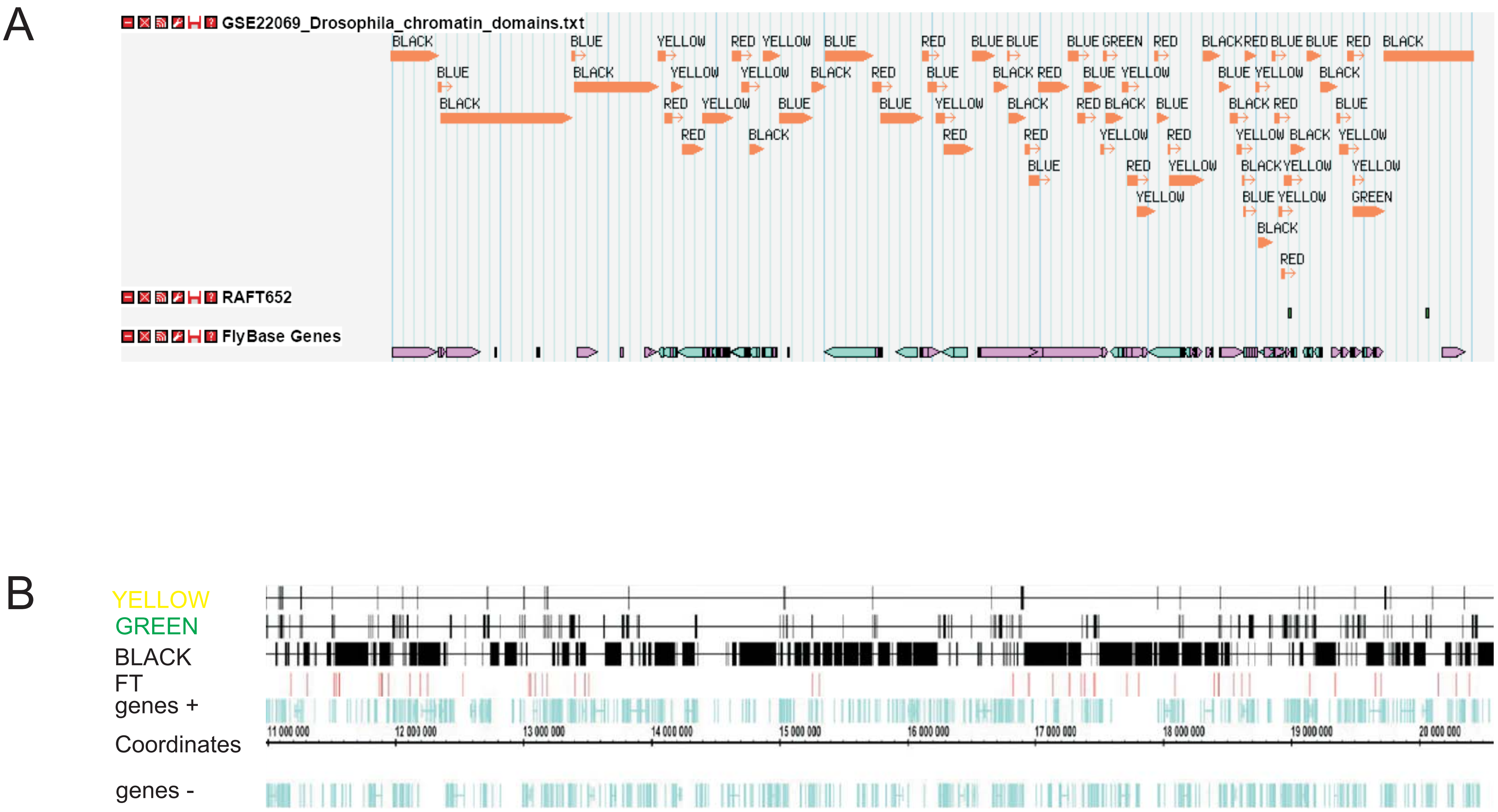
Supplementary Figure 6, Tchurikov et al.



Supplementary Figure 7, Tchurikov et al.



Tchurikov et al., Supplementary Figure 8



Tchurikov et al., Supplementary Figure 9