SUPPLEMENTARY MATERIALS

"Cinteny: Efficient Multiple Organism-based Analysis of Genome Rearrangements"

Amit U. Sinha and Jaroslaw Meller

Contact: jmeller@chmcc.org; Server available at http://cinteny.cchmc.org

SP1. Interactive assessment of synteny and evolutionary distances with Cinteny

Cinteny is a flexible and efficient tool for analysis of synteny and evolutionary distances in terms of genome rearrangements (the reversal distance) for multiple genomes. In addition to annotated genomes, which are available for interactive browsing and assessment of synteny and evolutionary distances in terms of orthologous genes, Cinteny can be used with user provided discrete objects, such as sequence tags or other evolutionarily conserved markers. The graphical layer is primarily based on specifically designed library of graphical objects and views, optimized to enable intuitive interpretation of the results.

Since a typical query, involving computation of the reversal distance for a pair of mammalian genomes, takes only about a CPU second, one can easily assess the effects of various approximations and different levels of coarse-graining. In particular, Cinteny offers the option of using multiple genomes to define a set of common, evolutionarily conserved markers (e.g. orthologs identified in all genomes included in the analysis, as specified by the user). This approach yields a natural coarse-graining, in which only the most conserved markers are taken into account when identifying syntenic blocks and computing reversal distances.

Furthemore, parameters affecting identification of synteny blocks can be interactively adjusted, leading to different levels of aggregation by choosing the minimum length of synteny blocks, maximum gap between adjacent markers, the minimum number of markers in a block and other parameters. In addition, the effect of paralogs (multiple copies of the same marker in a genome) may be assessed by choosing a range of options, from removing all paralogs to using the ones that are contained within the largest conserved blocks. In what follows, we demonstrate examples of interrogating evolutionary relatedness of genomes with the Cinteny server, using different types of queries and different setup of parameters.

SP2. Examples of queries and inter-genome comparisons using the Cinteny server

Cinteny can be used to visualize synteny around a single gene (or marker), to identify synteny blocks and compute reversal distance between pairs of chromosomes as well as whole genomes. In figures included below, we show an example of each type of query, demonstrating the effects of various parameters on the results. In particular, the level of coarse-graining is varied from none (referred to as NOAGG), through intermediate aggregation with the minimum length of synteny blocks and the maximum gap between adjacent markers set to 100 kb (denoted as INTAGG), to default aggregation adjusted for mammalian genomes, with the minimum length of synteny blocks set to 300 kb and the maximum gap

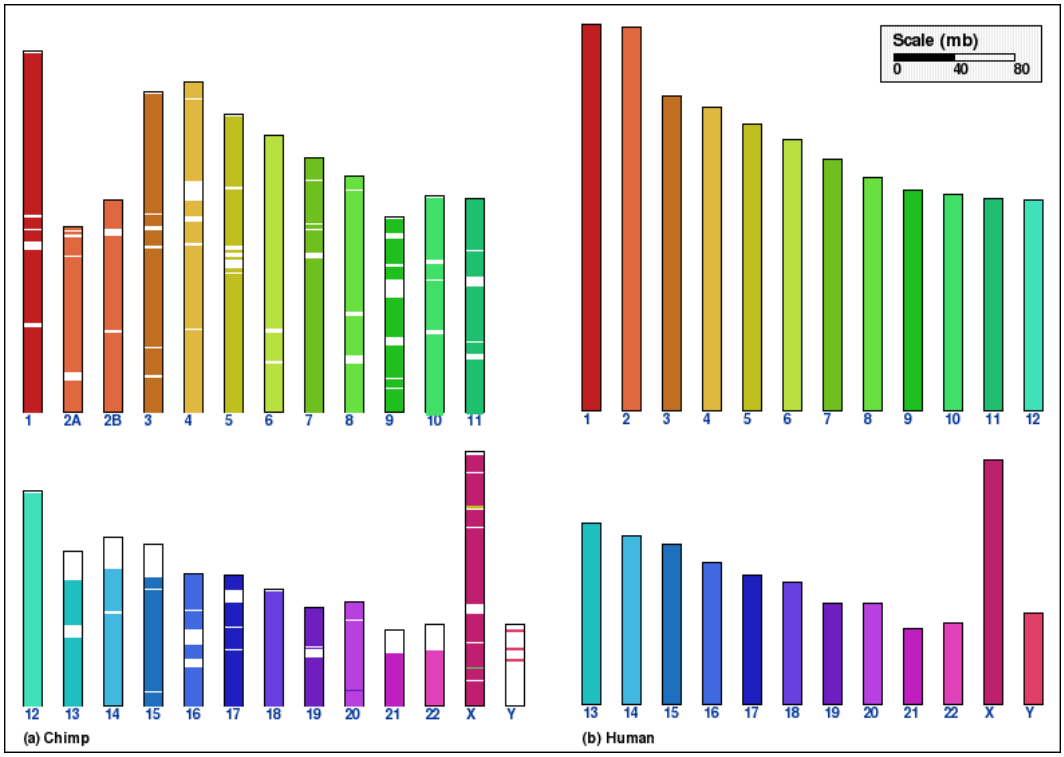
between adjacent markers set to 1 Mb (denoted as DEFAGG). The minimum number of markers is set to two and paralogs contained within the largest synteny blocks are used, unless otherwise stated. All the graphical representations shown below are cross-linked, as well as linked to external resources, such as NCBI (NCBI, 2005), enabling interactive browsing and exploring synteny.

As discussed in the manuscript, finding synteny blocks and the reversal distance typically involves identifying genes (or sequence tags) that are shared by the two species of interest. For example, the synteny between human and mouse may be analyzed using 15,645 orthologs, as identified by Homologene (NCBI, 2005). Alternatively, multiple organisms may be used, in order to define a reduced set of markers common to multiple species. For example, one may use a subset of 6,424 orthologs common to human, chimpanzee, dog, mouse and rat genomes, in order to find synteny between human and mouse or other pairs of genomes.

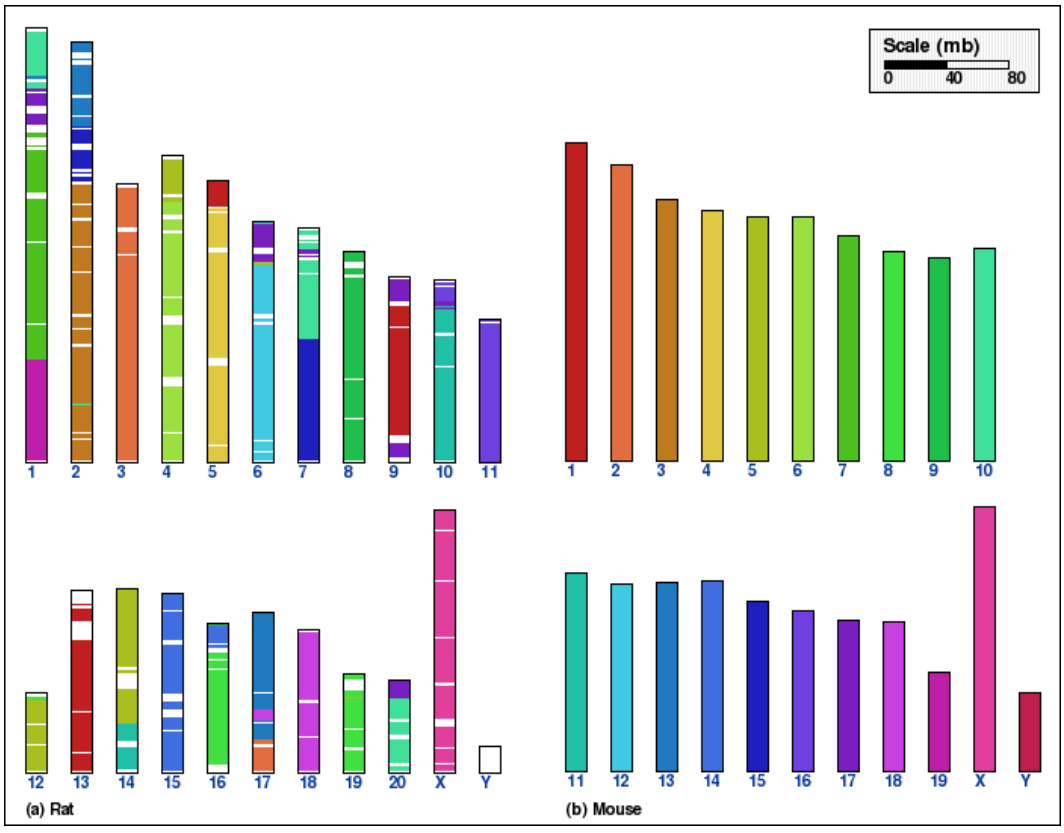
The effects of such defined coarse-graining are demonstrated for the whole genome comparisons (see Figure 1), and for X chromosomes (Figure 3), for chimpanzee vs. human, and rat vs. mouse. The reversal distances are computed in each case with different level of aggregation and coarse-graining. As can be seen from the figures, the reversal distance (referred to as RD) varies significantly in absolute terms, depending on the particular choice of parameters. On the other hand, the ratio of the distances between chimpanzee and human, and between rat and mouse genomes, for example, remains relatively constant for a reasonable range of coarse-graining parameters. Note also, that appropriate choice of coarse-graining parameters may be highly genome-specific, as illustrated in Figure 2. Other features and examples of queries are described at http://cinteny.cchmc.org.

Figure 1. Whole genome comparison for two pairs of mammalian genomes, using alternative aggregation approaches. In panels A and B, the intermediate aggregation (INTAGG) is used, leading to reversal distances of 21 and 128, respectively. In panels C and D, the same setup of parameters is used, except that only orthologs common to five mammalian genomes, as specified in the text, are used. In the latter case, reversal distances of 14 and 86, respectively, are obtained.

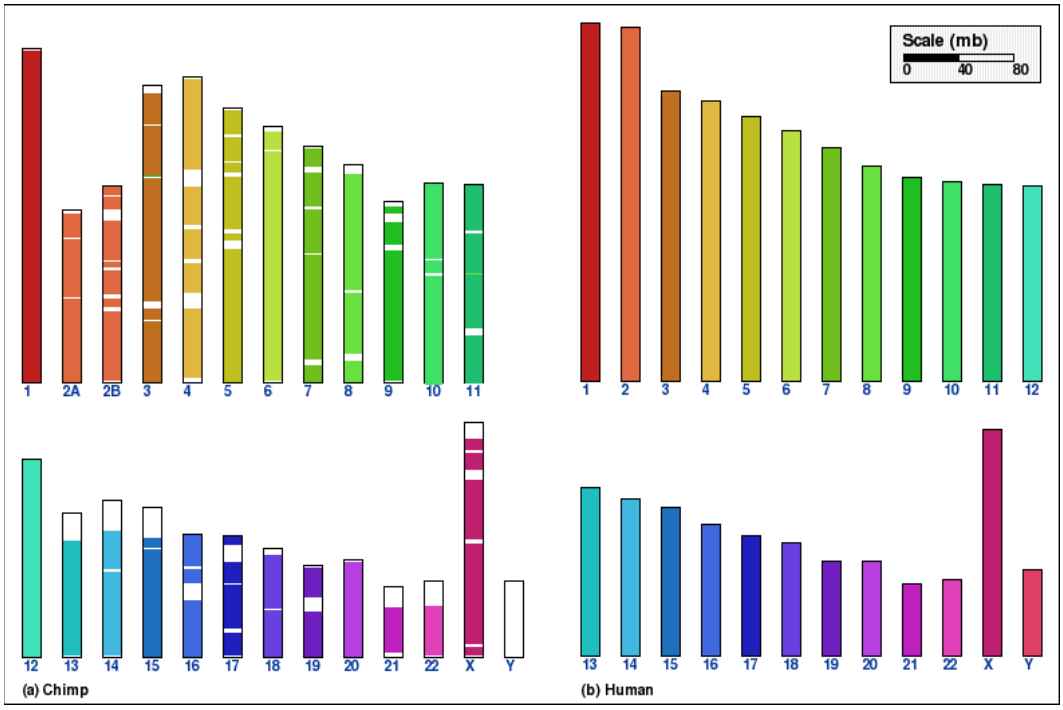
A.


Scale (mb)
0 40 80
1 2A 2B 3 4 5 6 7 8 9 10 11
1 2 3 4 5 6 7 8 9 10 11 12
12 13 14 15 16 17 18 19 20 21 22 X Y
13 14 15 16 17 18 19 20 21 22 X Y
(a) Chimp
(b) Human

B.


Scale (mb)
0 40 80
1 2 3 4 5 6 7 8 9 10 11
1 2 3 4 5 6 7 8 9 10
12 13 14 15 16 17 18 19 20 X Y
11 12 13 14 15 16 17 18 19 X Y
(a) Rat
(b) Mouse

C.


Scale (mb)
0 40 80
1 2A 2B 3 4 5 6 7 8 9 10 11
1 2 3 4 5 6 7 8 9 10 11 12
12 13 14 15 16 17 18 19 20 21 22 X Y
13 14 15 16 17 18 19 20 21 22 X Y
(a) Chimp
(b) Human

D.

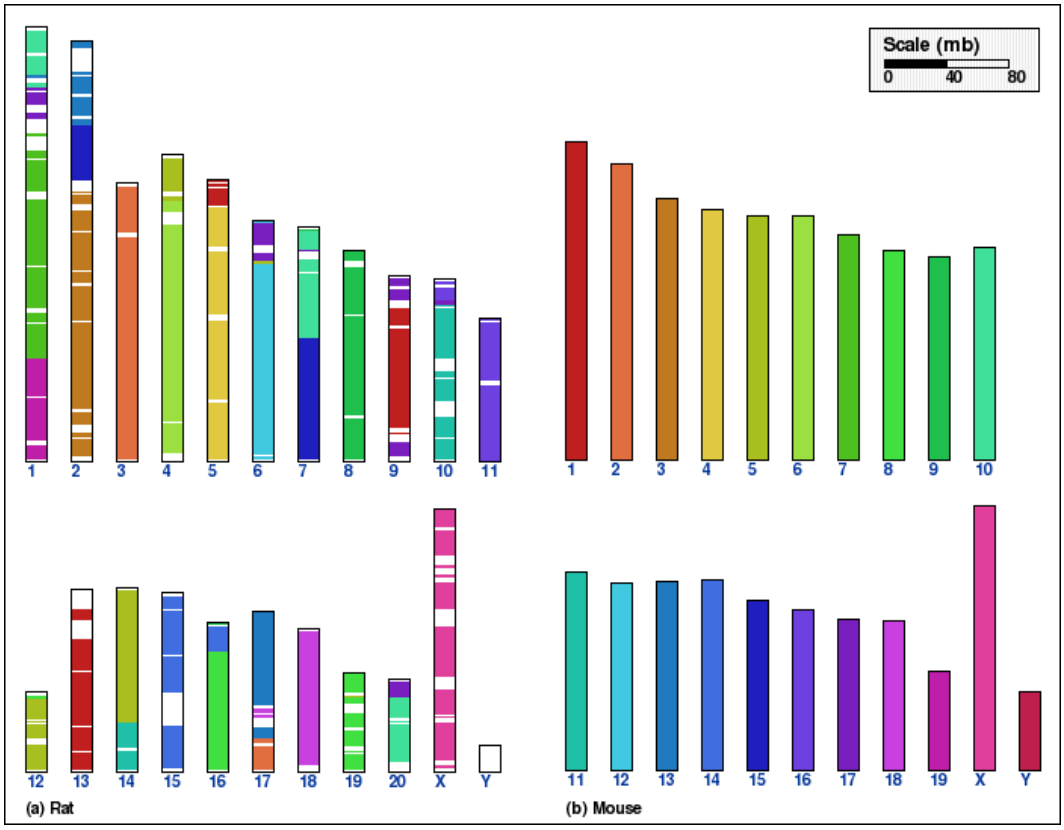
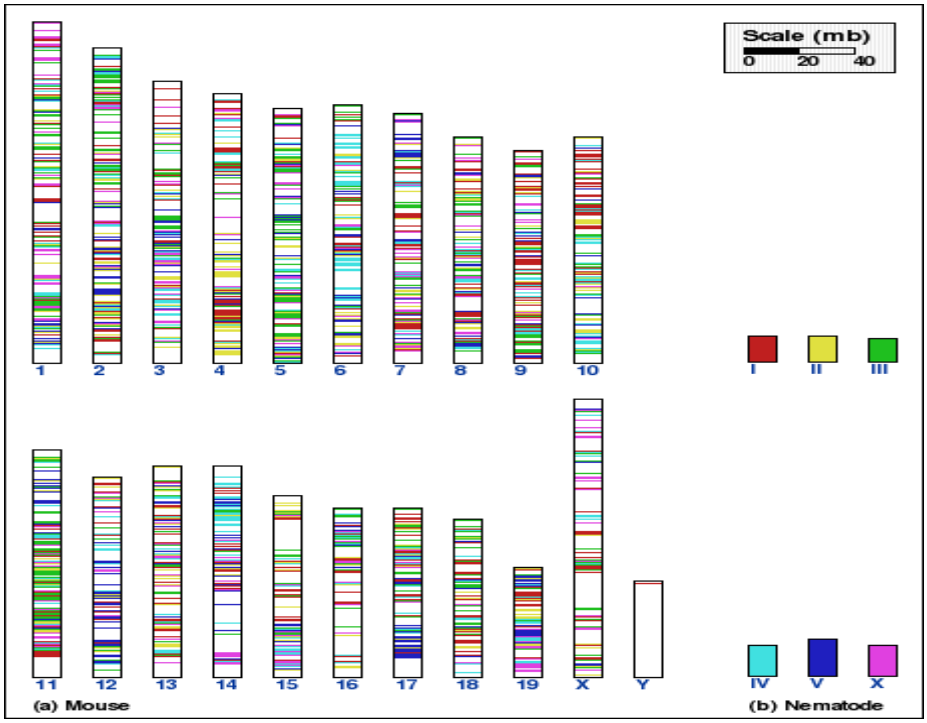
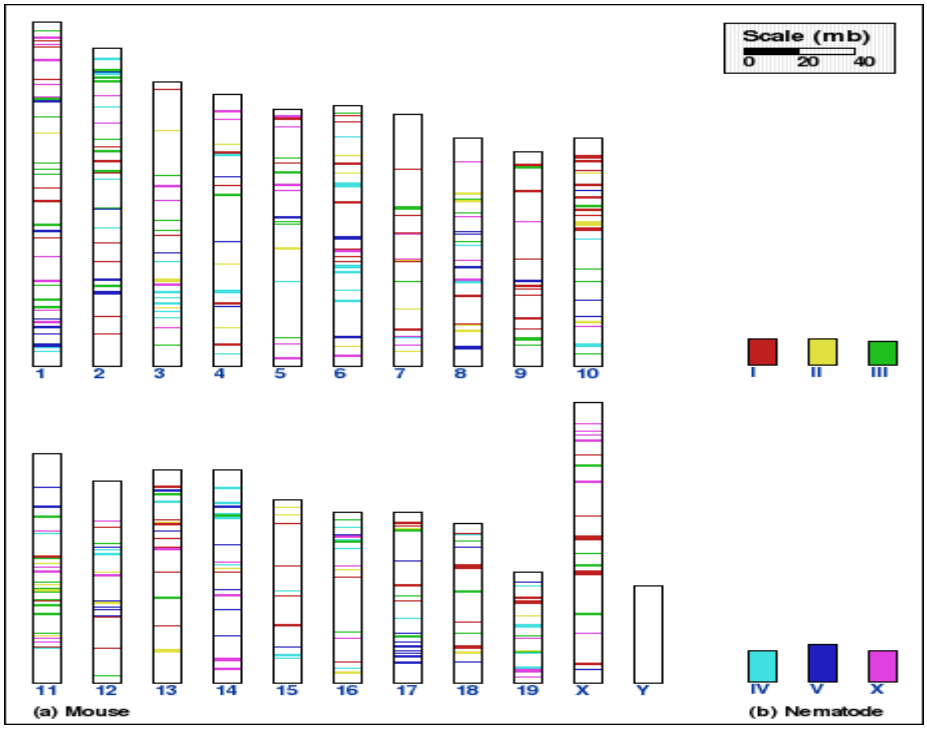

Scale (mb)
0 40 80
1 2 3 4 5 6 7 8 9 10 11
1 2 3 4 5 6 7 8 9 10
12 13 14 15 16 17 18 19 20 X Y
11 12 13 14 15 16 17 18 19 X Y
(a) Rat
(b) Mouse

Figure 2. Whole genome comparison for mouse vs. nematode (C. elegans), panels A and B; and for drosophila vs. nematode, panels C and D. Syntenic blocks obtained without any coarse-graining (NOAGG) are shown in panels A and C, respectively. On the other hand, an intermediate aggregation, with the minimal length of synteny blocks set to 100 kb, is used to show only sufficiently long blocks in panels B and D, respectively.

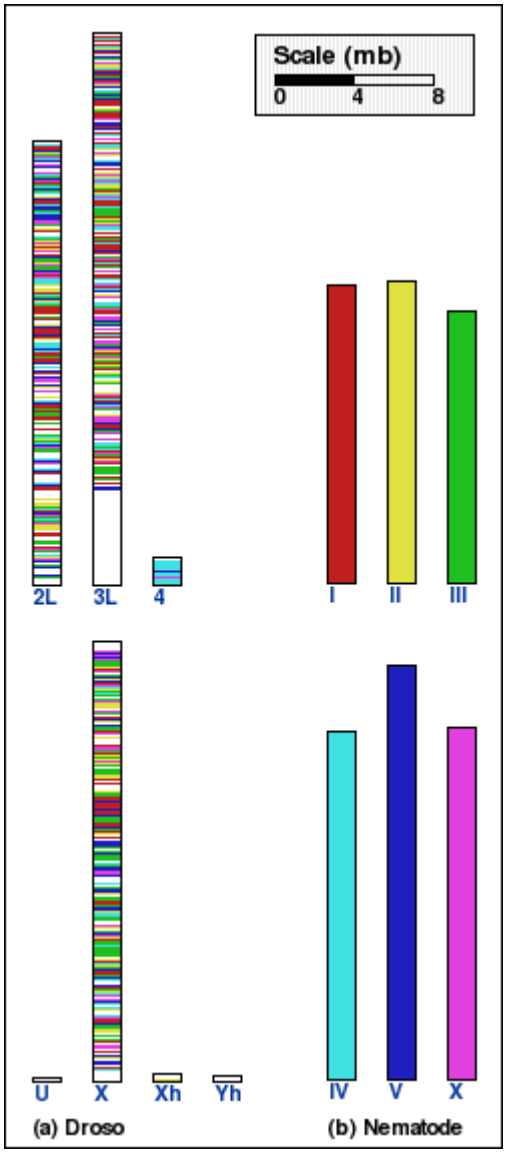
A.



B.



C.


Scale (mb)
0 4 8
2L
3L
4
I
II
III
U
X
Xh
Yh
IV
V
X
(a) Droso
(b) Nematode

D.

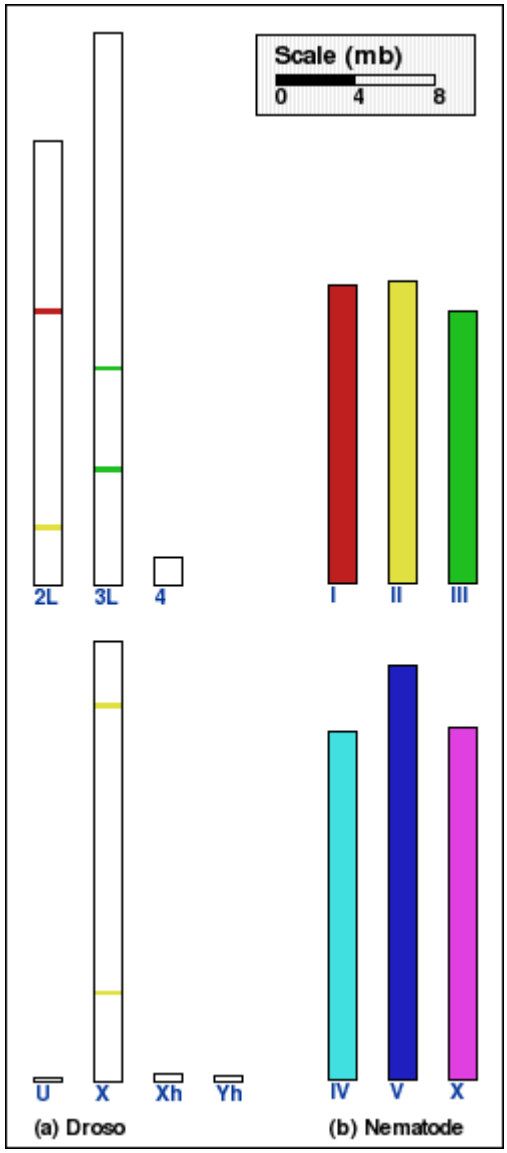
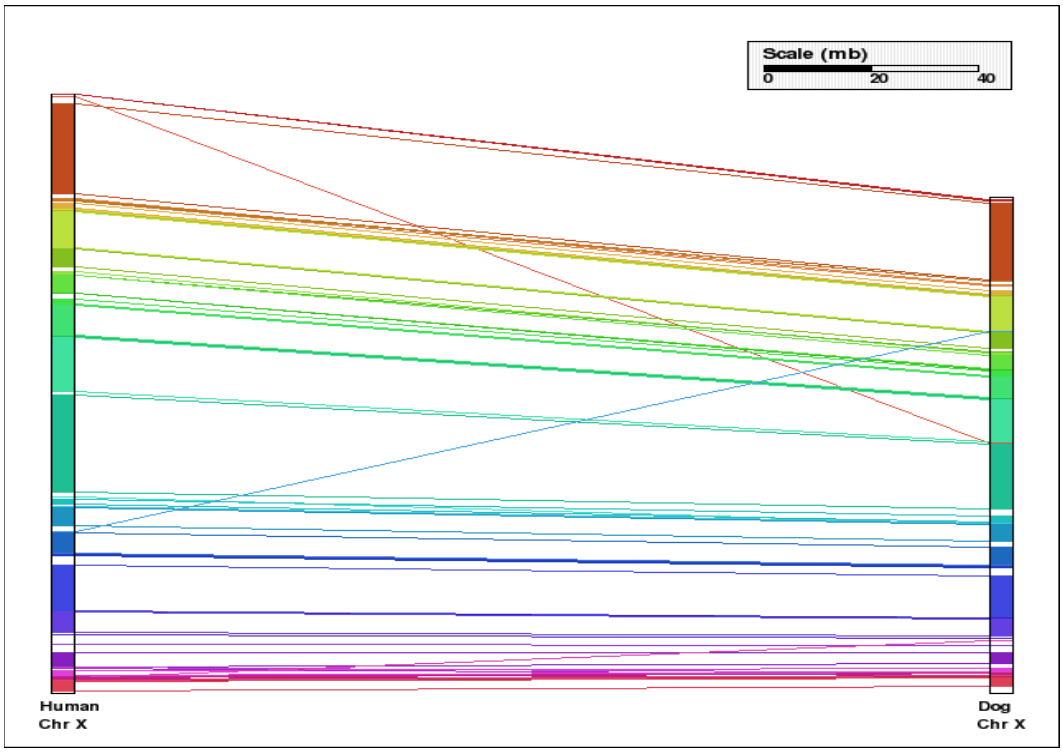
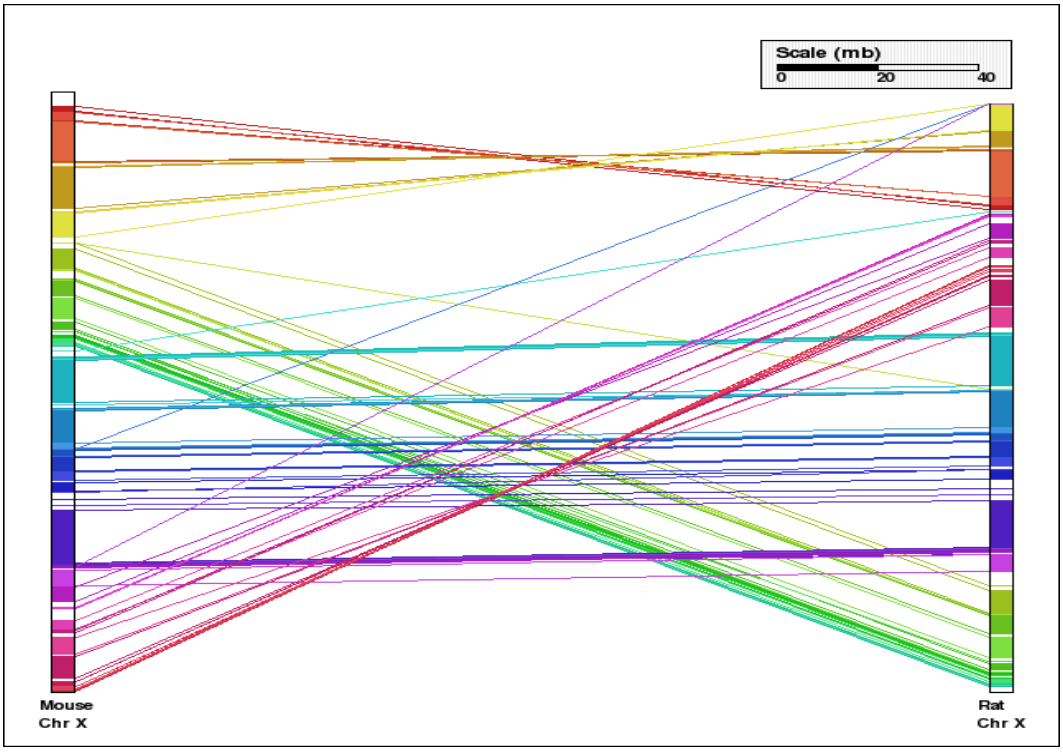

Scale (mb)
0 4 8
2L
3L
4
I
II
III
U
X
Xh
Yh
IV
V
X
(a) Droso
(b) Nematode

Figure 3. Analysis of synteny for X chromosomes in selected mammalian genomes. Human and dog, as well as mouse and rat X chromosomes are compared in panels A, C E and B, D, F, respectively. First, no coarse-graining and orthologs identified for each pair of genomes independently are used (panels A and B, with the resulting RD distances of 27 and 52, respectively), next no aggregation, however, with a set of orthologs common to all five mammalian genomes considered here, is used (panels C and D, with the resulting distances of 11 and 14), and finally, the default coarse-graining (DEFAGG), with the same common subset of orthologs, is used (panels E and F, with the resulting distances of 0 and 4, respectively).

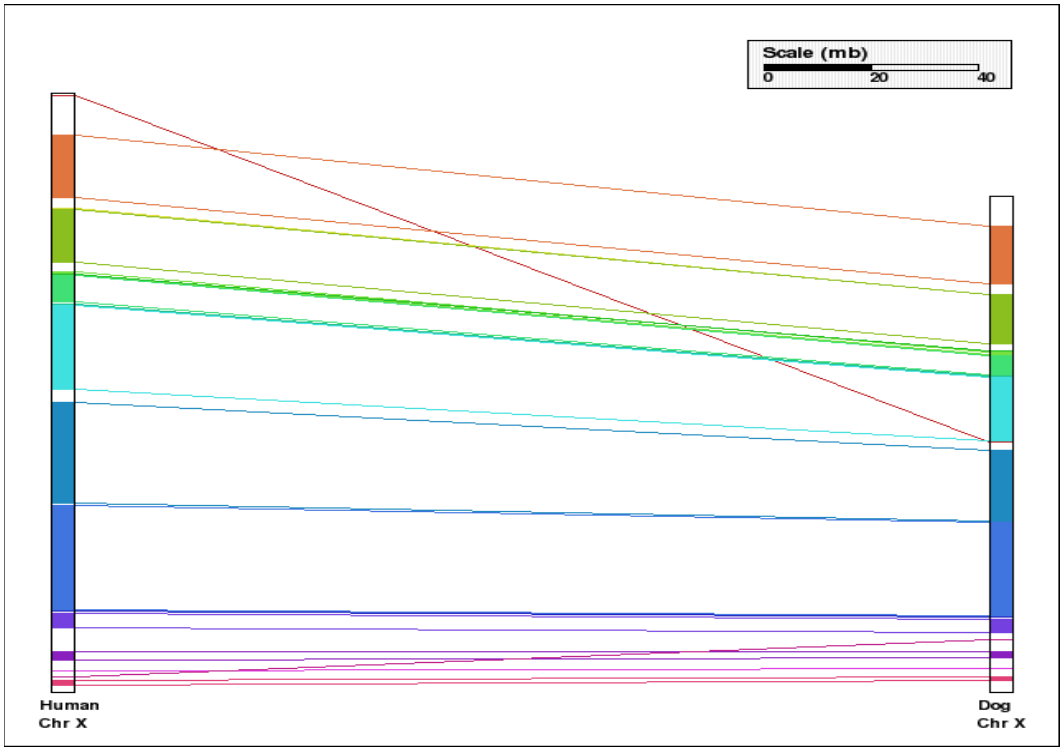
A.


Scale (mb)
0
20
40
Human
Chr X
Dog
Chr X

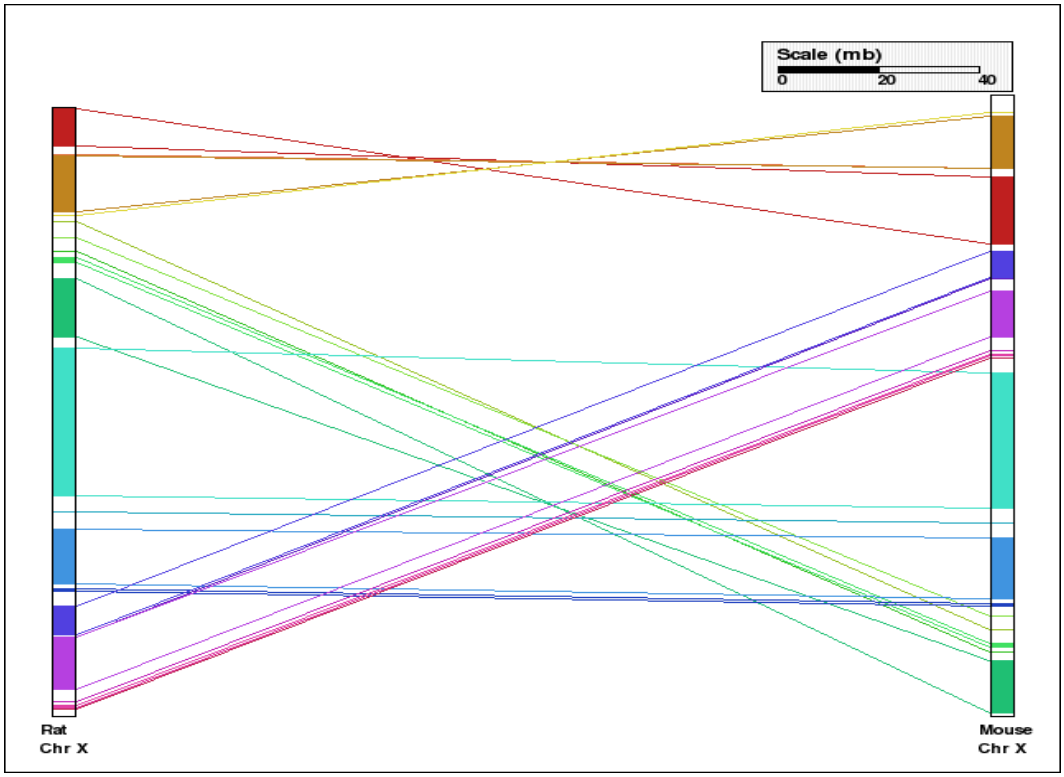
B.


Scale (mb)
0
20
40
Mouse
Chr X
Rat
Chr X

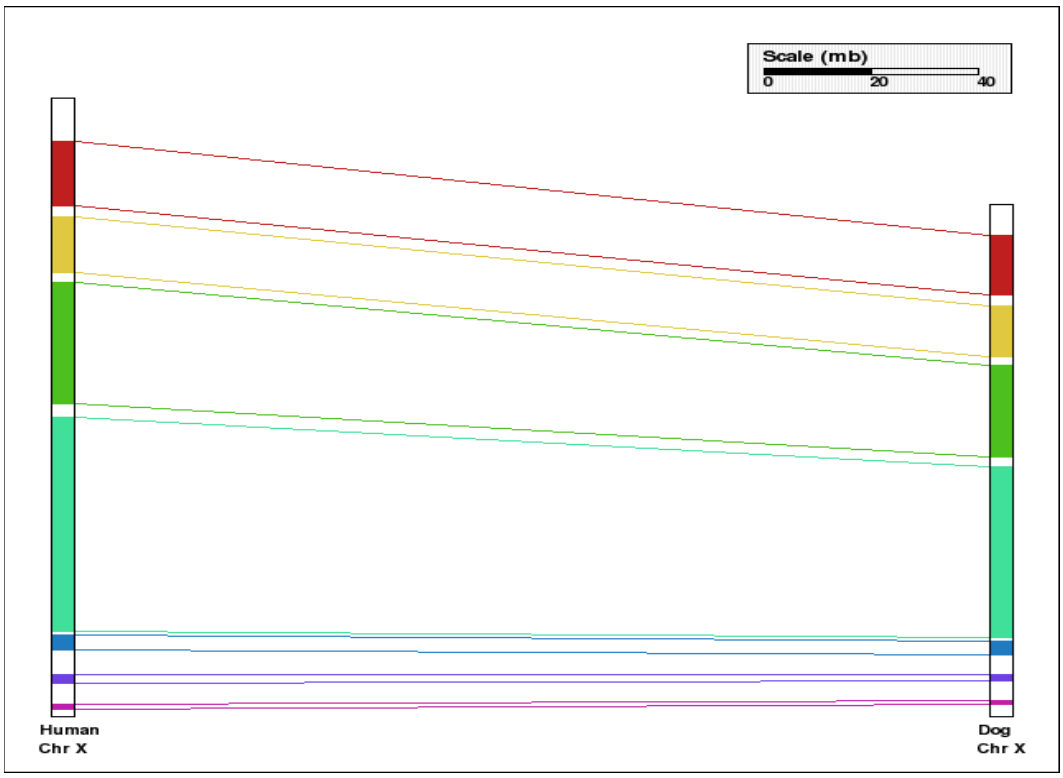
C.


Scale (mb)
0
20
40
Human
Chr X
Dog
Chr X

D.


Scale (mb)
0
20
40
Rat
Chr X
Mouse
Chr X

E.


Scale (mb)
0
20
40
Human
Chr X
Dog
Chr X

F.

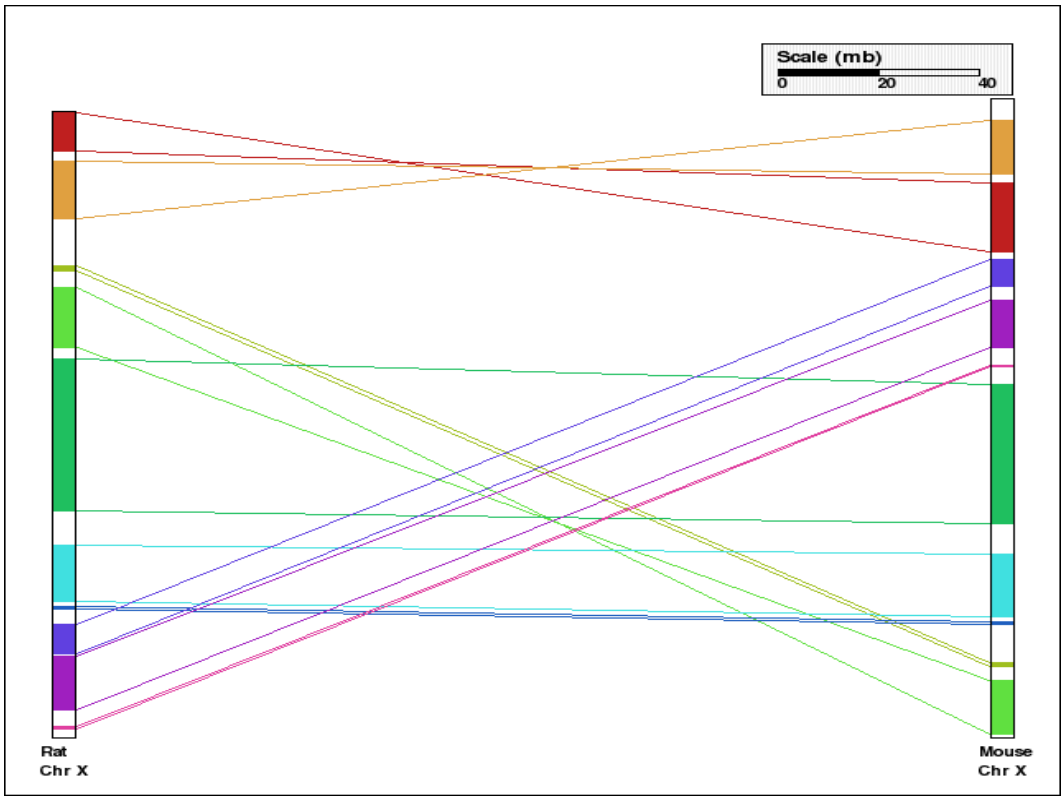
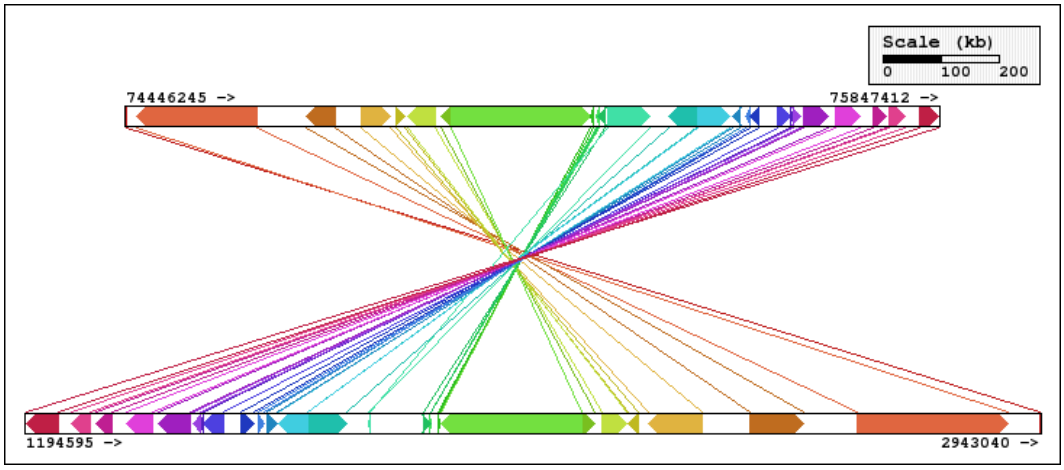

Scale (mb)
0
20
40
Rat
Chr X
Mouse
Chr X

Figure 4. Syntenic blocks around the replication protein A1 (RPA1) gene in the mouse and human (panel A), as well as mouse and rat genomes (panel B), respectively. A set of orthologs common to all three genomes and default coarse-graining are used.

A.



B.