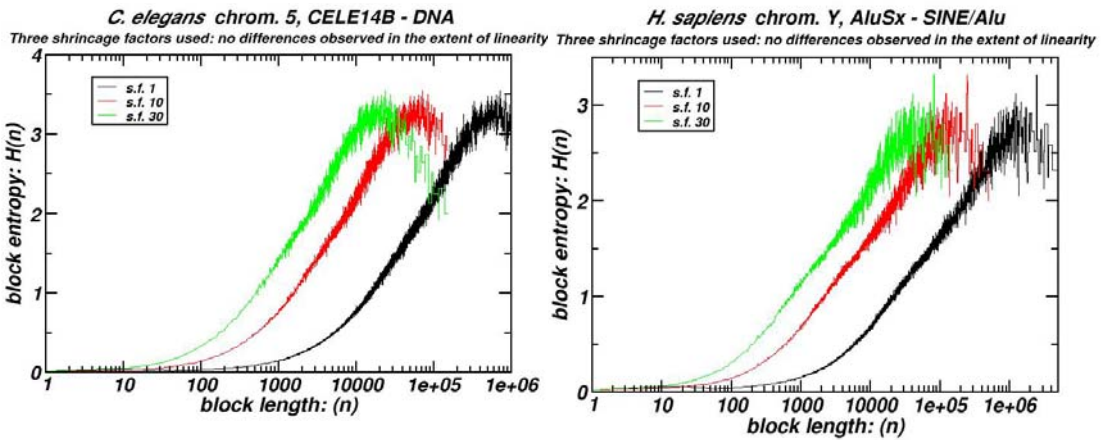
Supplementary Material

Supplementary material contents

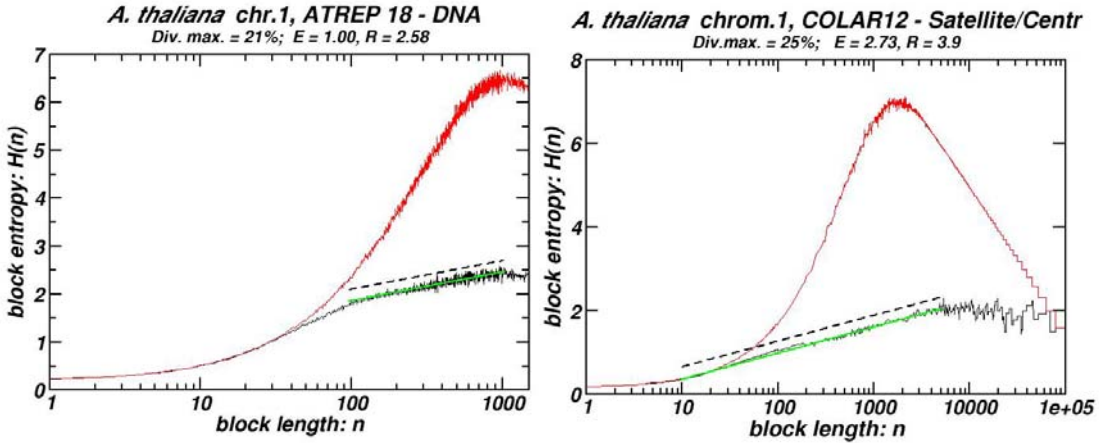
- Figures showing the independence of the derived fractality measures on the value of the *shrinkage factor* (*s.f.*) used.
- 33 entropic scaling plots: *supplement to Figure 1*.
- 33 box-counting plots: *supplement to Figure 2*.
- 3x10 plots for the manipulations described in the "Results" as alternatives of the box-counting study the first 10 cases: *supplement to Figure 3*.
- Three cases of chromosomal regions with high SIM vs. the hosting chromosomes, not exhibiting fractality: *supplement to Figure 4*.
- Details of the simulations.

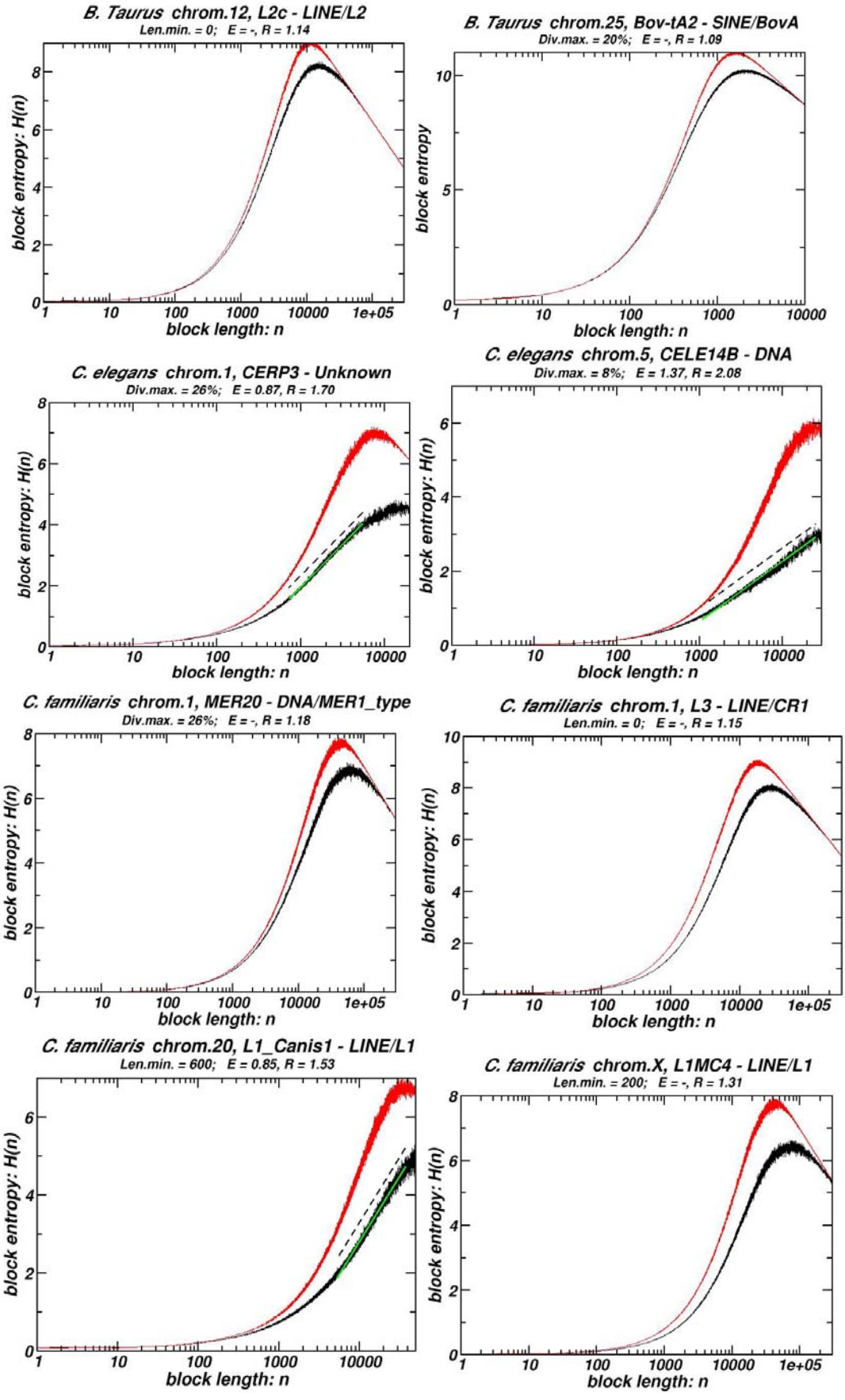
1. Figures Showing the Independence of the Derived Fractality Measures on the Used Shrinkage Factor (s.f.)

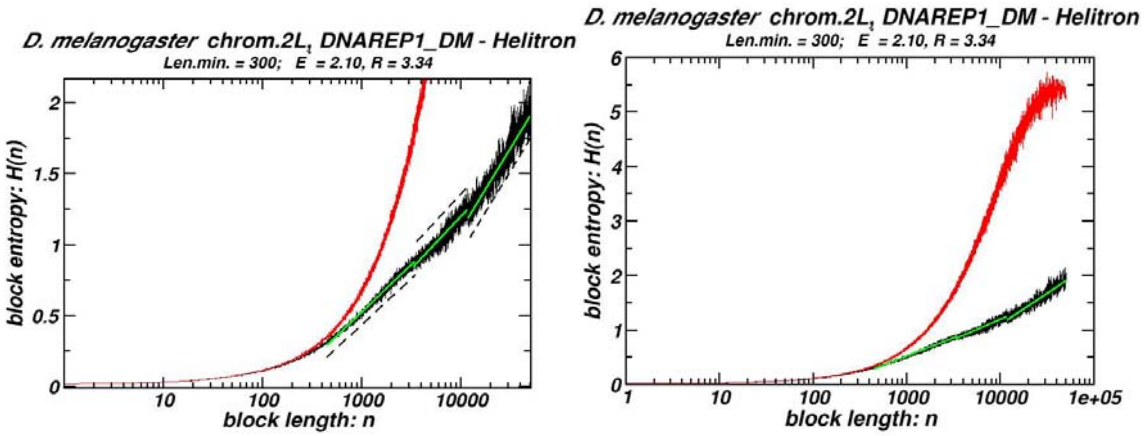


The above examples are typical for the independence of the values of E (extent of linearity) for different *shrinkage factor* choices. The case *s.f.* = 1 corresponds to no shrinkage.

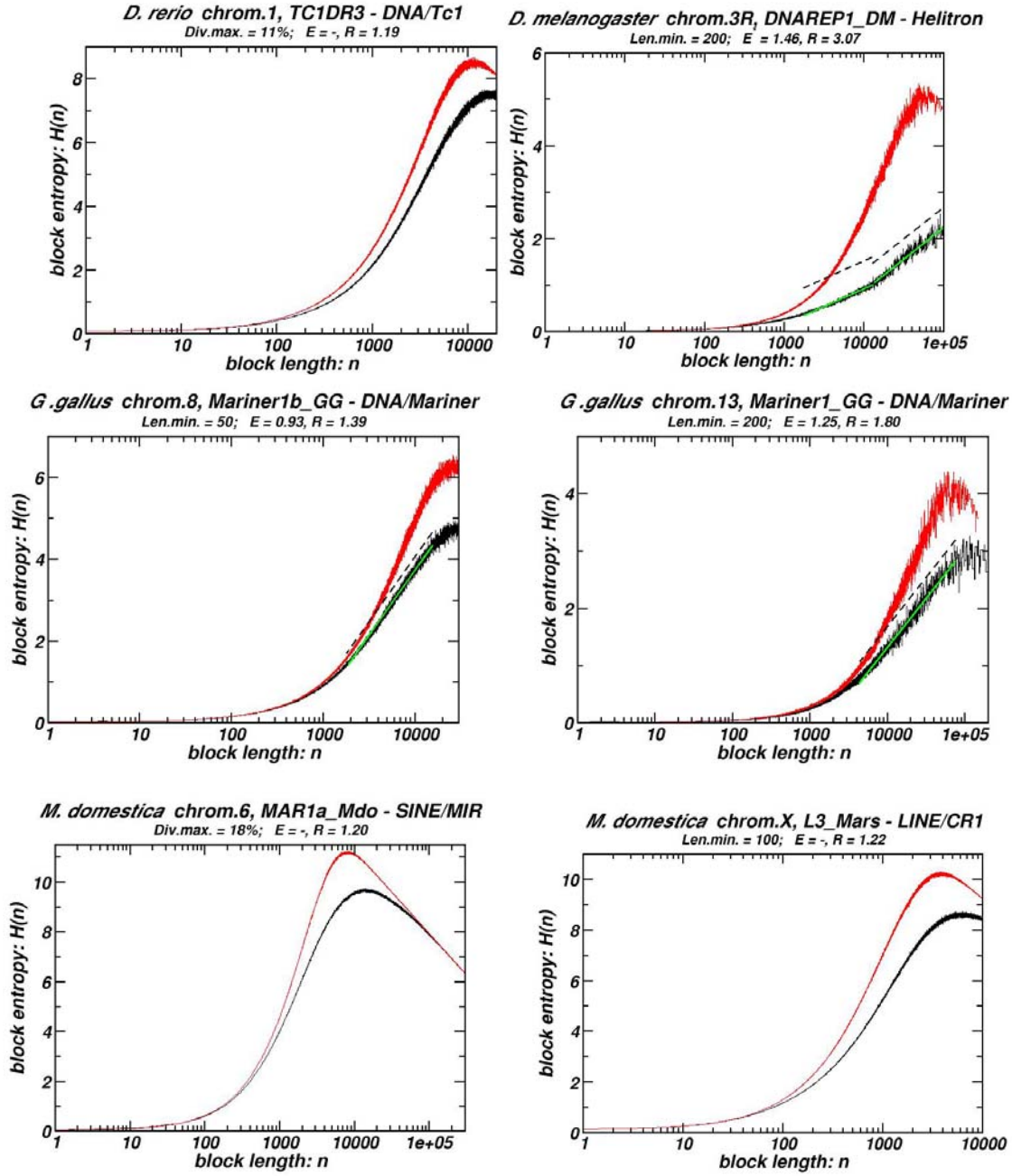
2. Entropic Scaling Plots

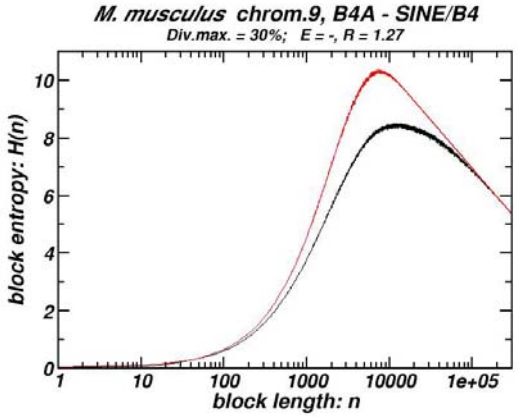


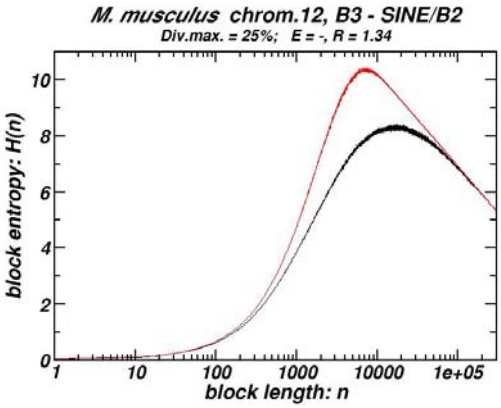

B. Taurus chrom.12, L2c - LINE/L2
Len.min. = 0; E = -, R = 1.14
B. Taurus chrom.25, Bov-tA2 - SINE/BovA
Div.max. = 20%; E = -, R = 1.09
C. elegans chrom.1, CERP3 - Unknown
Div.max. = 26%; E = 0.87, R = 1.70
C. elegans chrom.5, CELE14B - DNA
Div.max. = 8%; E = 1.37, R = 2.08
C. familiaris chrom.1, MER20 - DNA/MER1_type
Div.max. = 26%; E = -, R = 1.18
C. familiaris chrom.1, L3 - LINE/CR1
Len.min. = 0; E = -, R = 1.15
C. familiaris chrom.20, L1_Canis1 - LINE/L1
Len.min. = 600; E = 0.85, R = 1.53
C. familiaris chrom.X, L1MC4 - LINE/L1
Len.min. = 200; E = -, R = 1.31
block entropy: H(n)
block entropy
block length: n

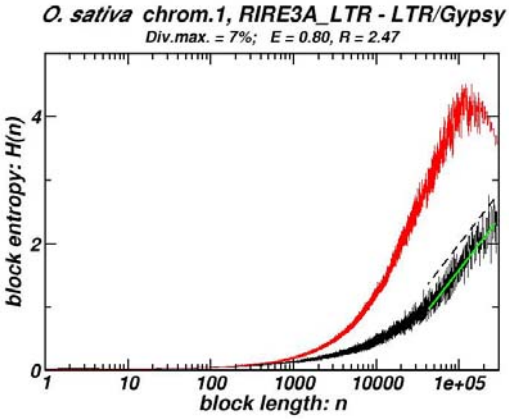


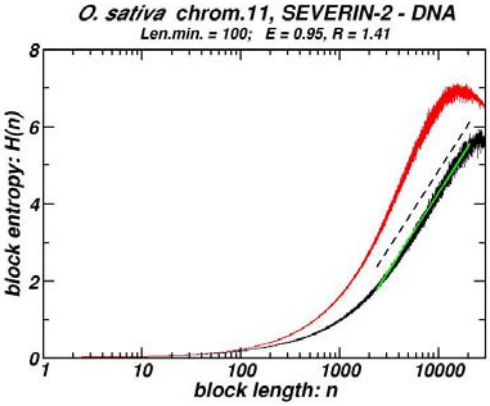
Left plot: detail showing the linear segments. Right plot: the full-scale plot.

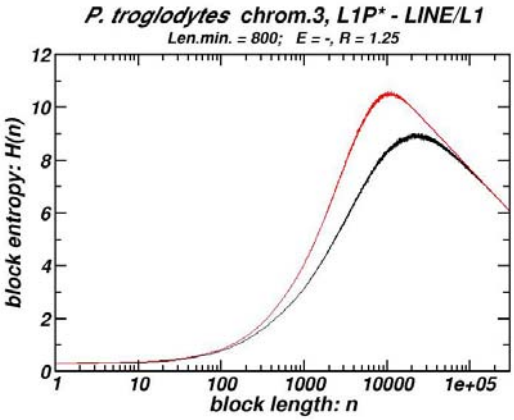


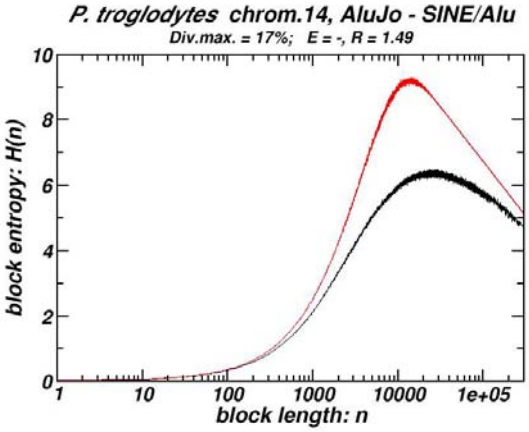

M. musculus chrom.9, B4A - SINE/B4
Div.max. = 30%; E = -, R = 1.27
block entropy: H(n)
block length: n

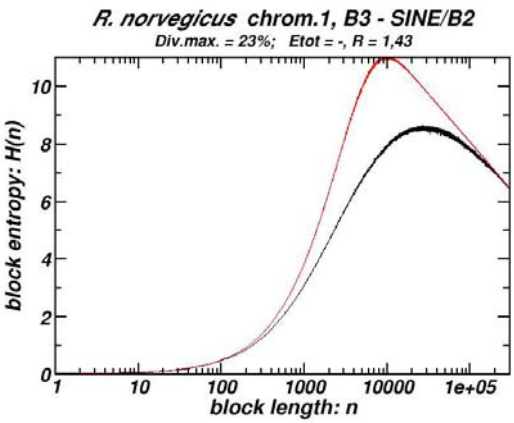

M. musculus chrom.12, B3 - SINE/B2
Div.max. = 25%; E = -, R = 1.34
block entropy: H(n)
block length: n

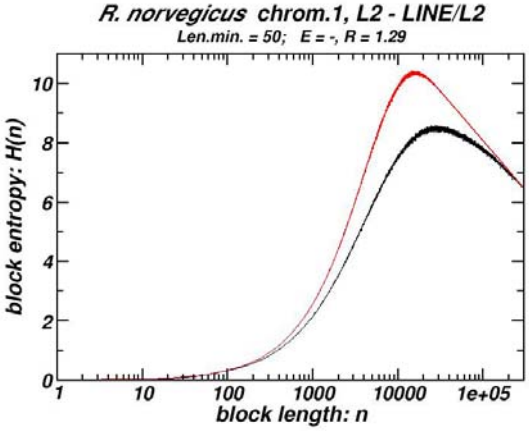

O. sativa chrom.1, RIRE3A_LTR - LTR/Gypsy
Div.max. = 7%; E = 0.80, R = 2.47
block entropy: H(n)
block length: n

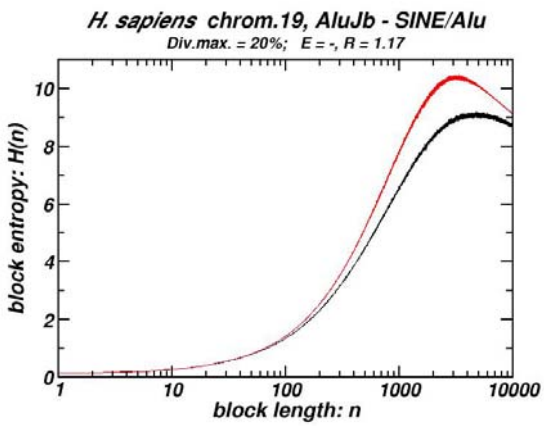

O. sativa chrom.11, SEVERIN-2 - DNA
Len.min. = 100; E = 0.95, R = 1.41
block entropy: H(n)
block length: n

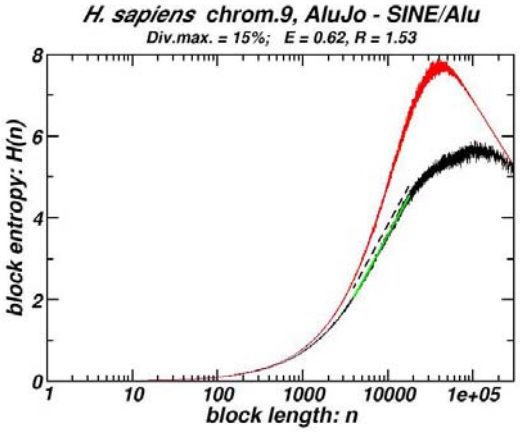

P. troglodytes chrom.3, L1P* - LINE/L1
Len.min. = 800; E = -, R = 1.25
block entropy: H(n)
block length: n

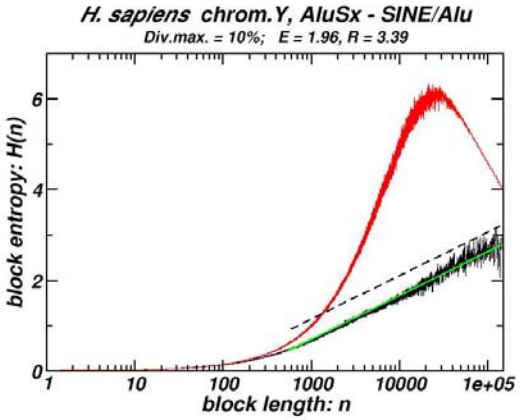

P. troglodytes chrom.14, AluJo - SINE/Alu
Div.max. = 17%; E = -, R = 1.49
block entropy: H(n)
block length: n

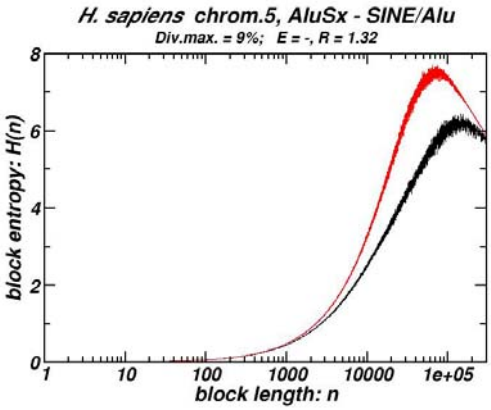

R. norvegicus chrom.1, B3 - SINE/B2
Div.max. = 23%; Etot = -, R = 1,43
block entropy: H(n)
block length: n

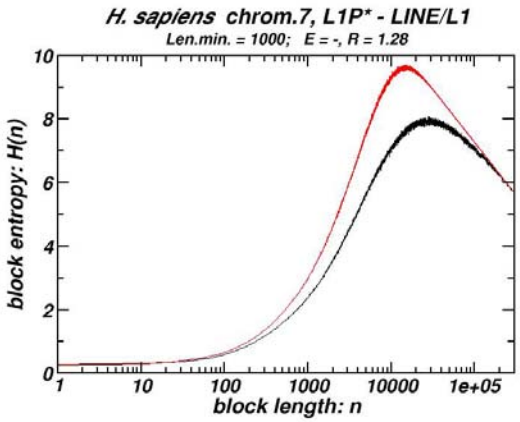

R. norvegicus chrom.1, L2 - LINE/L2
Len.min. = 50; E = -, R = 1.29
block entropy: H(n)
block length: n

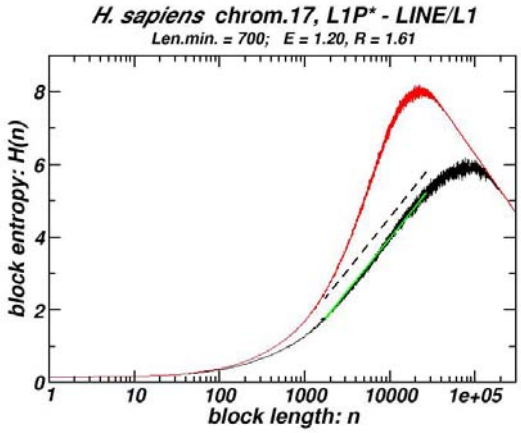

H. sapiens chrom.19, AluJb - SINE/Alu
Div.max. = 20%; E = -, R = 1.17
block entropy: H(n)
block length: n

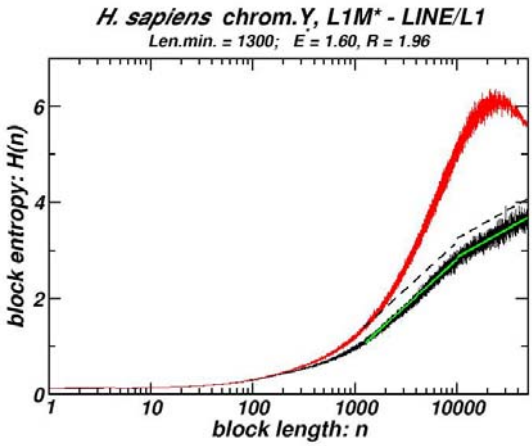

H. sapiens chrom.9, AluJo - SINE/Alu
Div.max. = 15%; E = 0.62, R = 1.53
block entropy: H(n)
block length: n

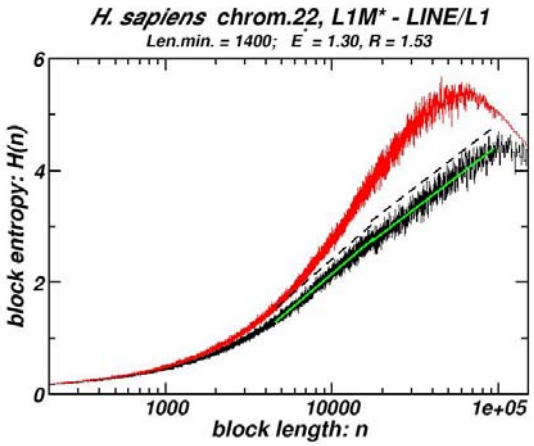

H. sapiens chrom.Y, AluSx - SINE/Alu
Div.max. = 10%; E = 1.96, R = 3.39
block entropy: H(n)
block length: n


H. sapiens chrom.5, AluSx - SINE/Alu
Div.max. = 9%; E = -, R = 1.32
block entropy: H(n)
block length: n

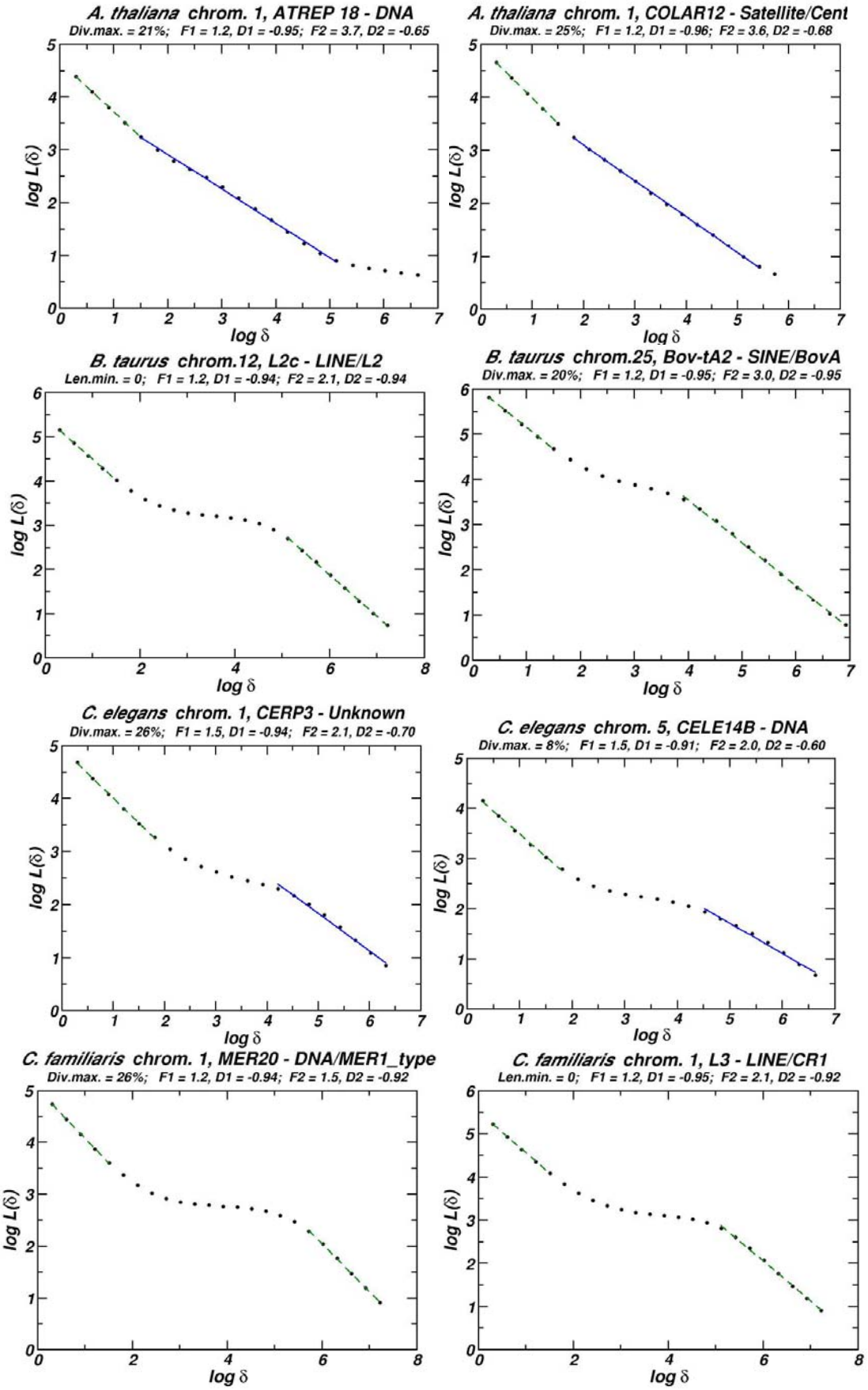

H. sapiens chrom.7, L1P* - LINE/L1
Len.min. = 1000; E = -, R = 1.28
block entropy: H(n)
block length: n

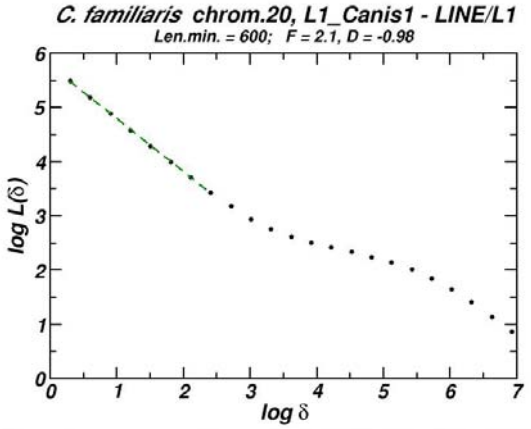

H. sapiens chrom.17, L1P* - LINE/L1
Len.min. = 700; E = 1.20, R = 1.61
block entropy: H(n)
block length: n

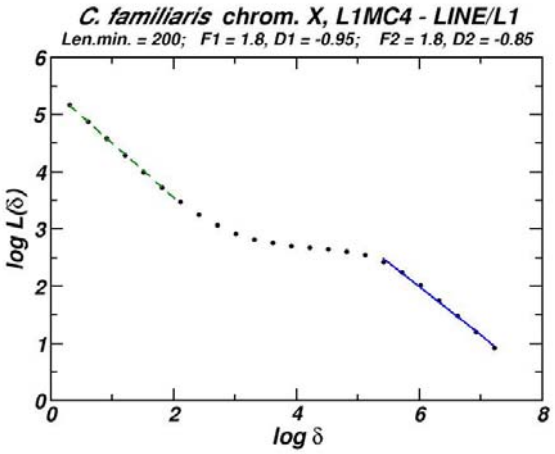

H. sapiens chrom.Y, L1M* - LINE/L1
Len.min. = 1300; E = 1.60, R = 1.96
block entropy: H(n)
block length: n

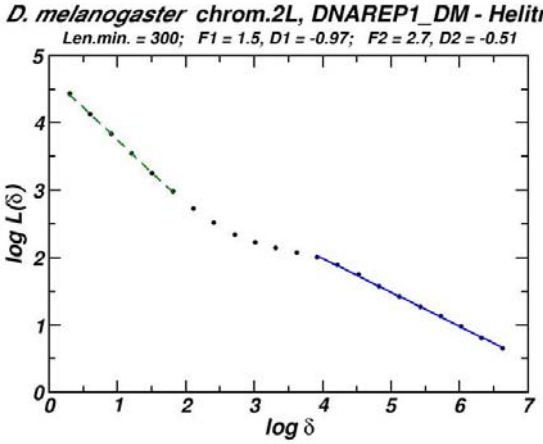

H. sapiens chrom.22, L1M* - LINE/L1
Len.min. = 1400; E = 1.30, R = 1.53
block entropy: H(n)
block length: n

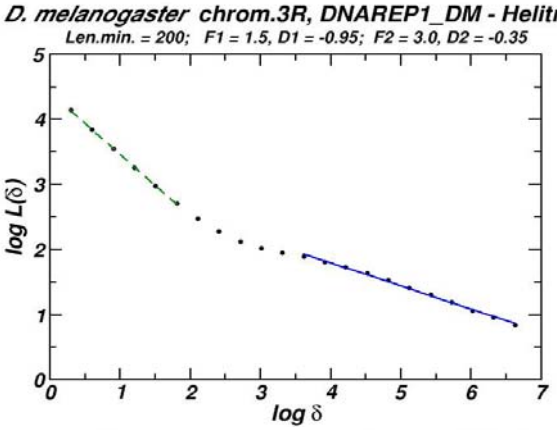
3. Box-Counting Plots

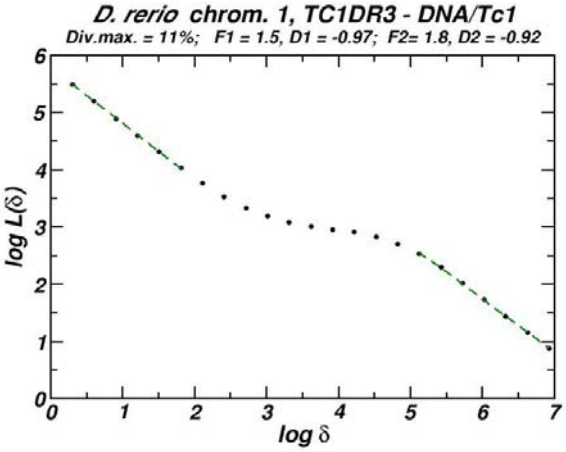


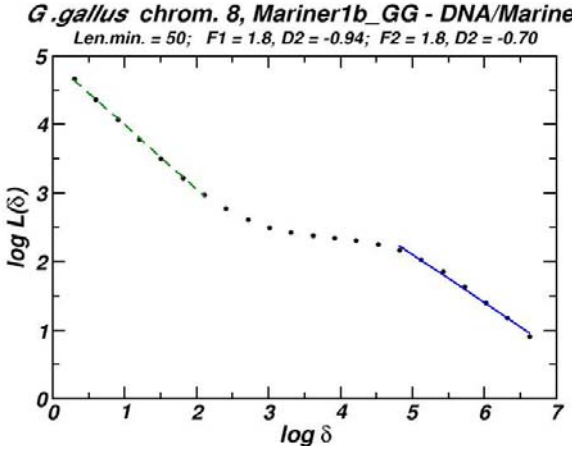

C. familiaris chrom.20, L1_Canis1 - LINE/L1
Len.min. = 600; F = 2.1, D = -0.98
log L(δ)
log δ

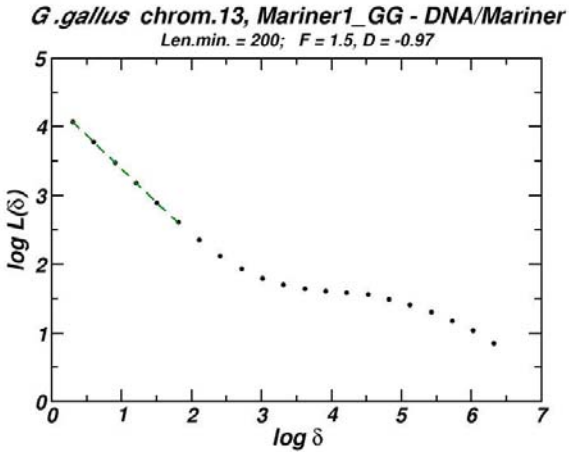

C. familiaris chrom. X, L1MC4 - LINE/L1
Len.min. = 200; F1 = 1.8, D1 = -0.95; F2 = 1.8, D2 = -0.85
log L(δ)
log δ

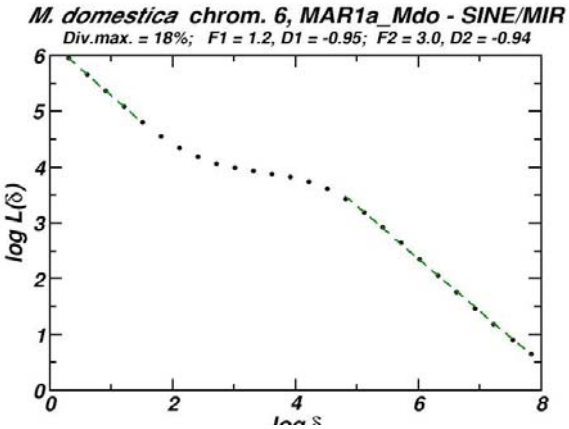

D. melanogaster chrom.2L, DNAREP1_DM - Helitron
Len.min. = 300; F1 = 1.5, D1 = -0.97; F2 = 2.7, D2 = -0.51
log L(δ)
log δ

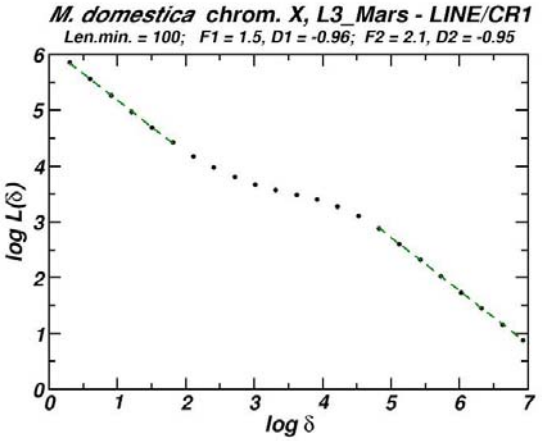

D. melanogaster chrom.3R, DNAREP1_DM - Helitron
Len.min. = 200; F1 = 1.5, D1 = -0.95; F2 = 3.0, D2 = -0.35
log L(δ)
log δ

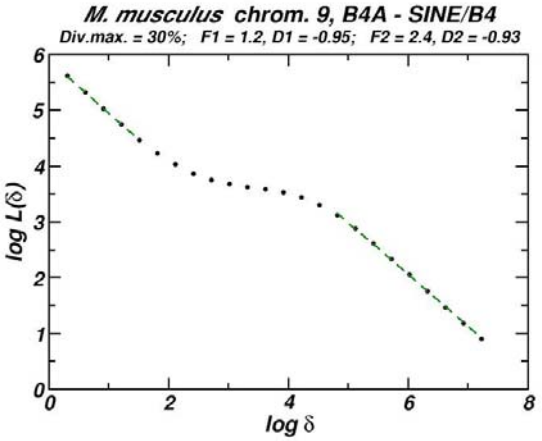

D. rerio chrom. 1, TC1DR3 - DNA/Tc1
Div.max. = 11%; F1 = 1.5, D1 = -0.97; F2= 1.8, D2 = -0.92
log L(δ)
log δ

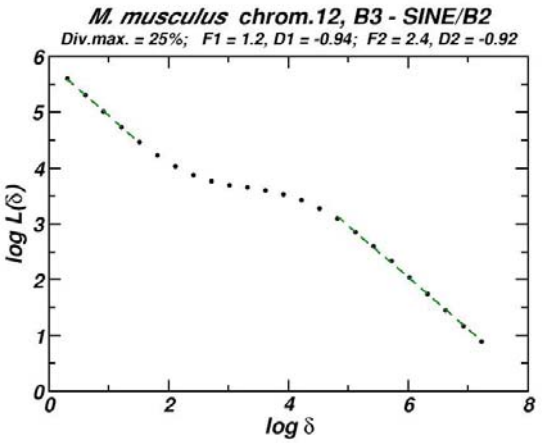

G .gallus chrom. 8, Mariner1b_GG - DNA/Mariner
Len.min. = 50; F1 = 1.8, D2 = -0.94; F2 = 1.8, D2 = -0.70
log L(δ)
log δ

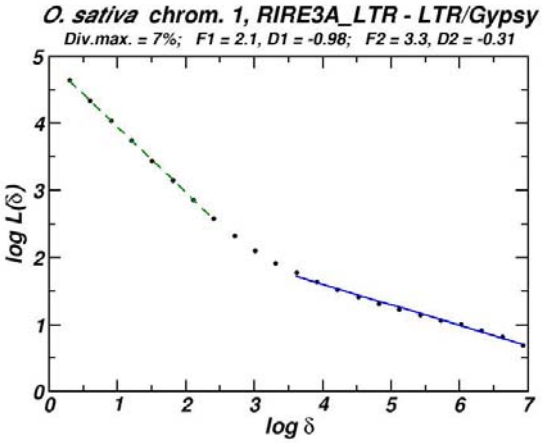

G .gallus chrom.13, Mariner1_GG - DNA/Mariner
Len.min. = 200; F = 1.5, D = -0.97
log L(δ)
log δ

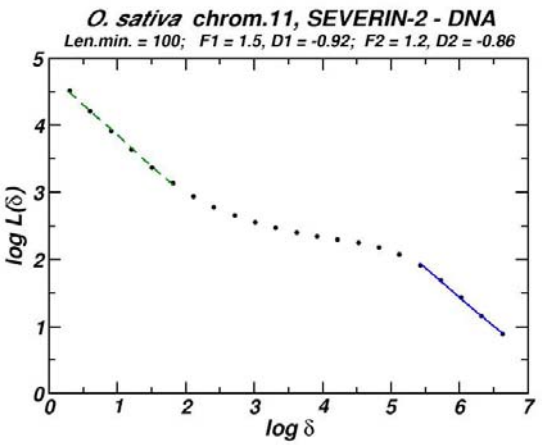

M. domestica chrom. 6, MAR1a_Mdo - SINE/MIR
Div.max. = 18%; F1 = 1.2, D1 = -0.95; F2 = 3.0, D2 = -0.94
log L(δ)
log δ

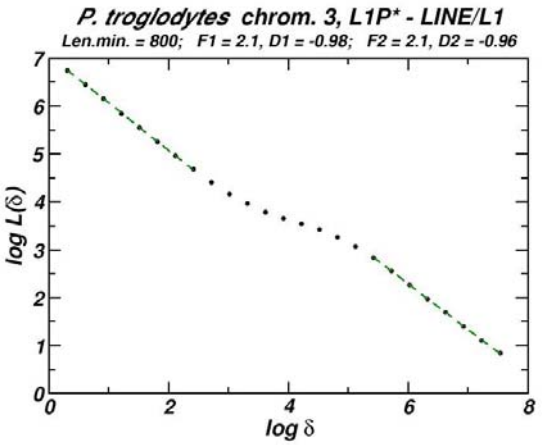

M. domestica chrom. X, L3_Mars - LINE/CR1
Len.min. = 100; F1 = 1.5, D1 = -0.96; F2 = 2.1, D2 = -0.95
log L(δ)
log δ

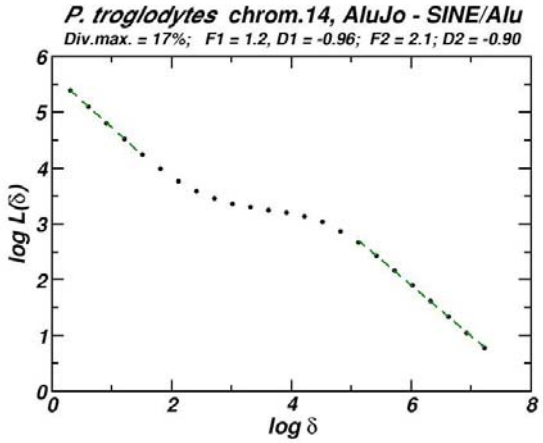

M. musculus chrom. 9, B4A - SINE/B4
Div.max. = 30%; F1 = 1.2, D1 = -0.95; F2 = 2.4, D2 = -0.93
log L(δ)
log δ

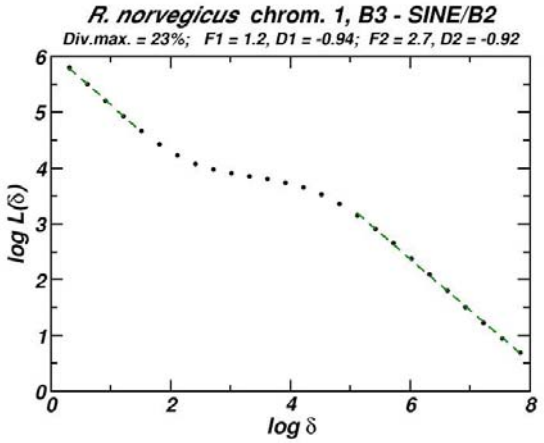

M. musculus chrom.12, B3 - SINE/B2
Div.max. = 25%; F1 = 1.2, D1 = -0.94; F2 = 2.4, D2 = -0.92
log L(δ)
log δ

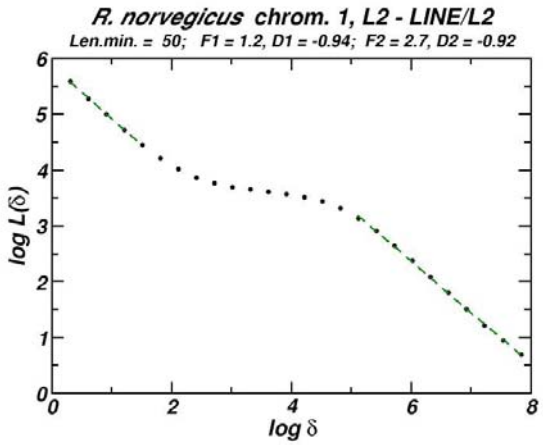

O. sativa chrom. 1, RIRE3A_LTR - LTR/Gypsy
Div.max. = 7%; F1 = 2.1, D1 = -0.98; F2 = 3.3, D2 = -0.31
log L(δ)
log δ

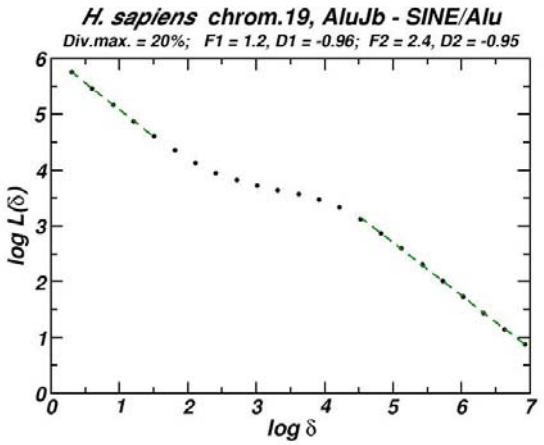

O. sativa chrom.11, SEVERIN-2 - DNA
Len.min. = 100; F1 = 1.5, D1 = -0.92; F2 = 1.2, D2 = -0.86
log L(δ)
log δ

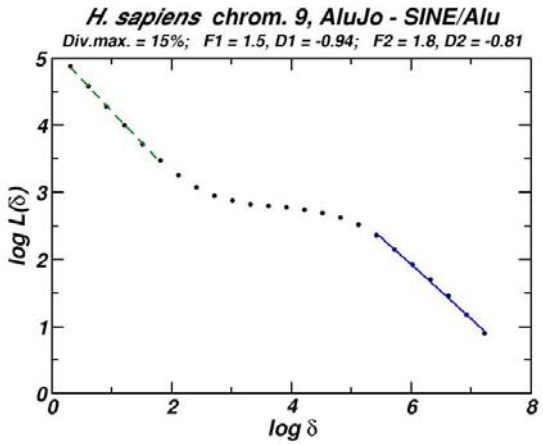

P. troglodytes chrom. 3, L1P* - LINE/L1
Len.min. = 800; F1 = 2.1, D1 = -0.98; F2 = 2.1, D2 = -0.96
log L(δ)
log δ

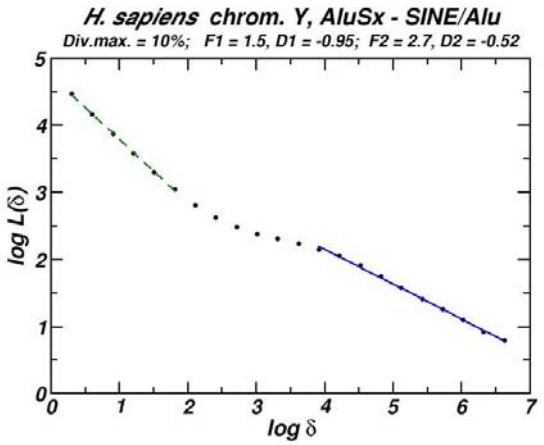

P. troglodytes chrom.14, AluJo - SINE/Alu
Div.max. = 17%; F1 = 1.2, D1 = -0.96; F2 = 2.1; D2 = -0.90
log L(δ)
log δ

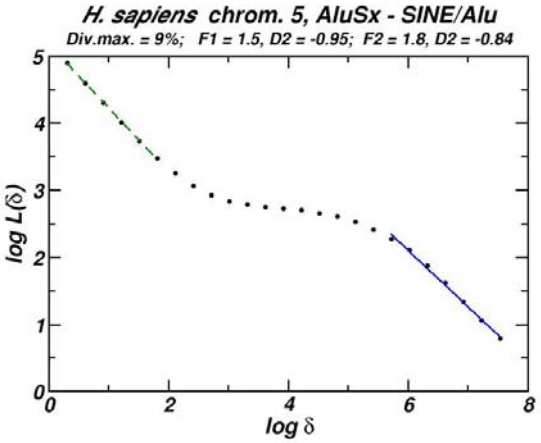

R. norvegicus chrom. 1, B3 - SINE/B2
Div.max. = 23%; F1 = 1.2, D1 = -0.94; F2 = 2.7, D2 = -0.92
log L(δ)
log δ

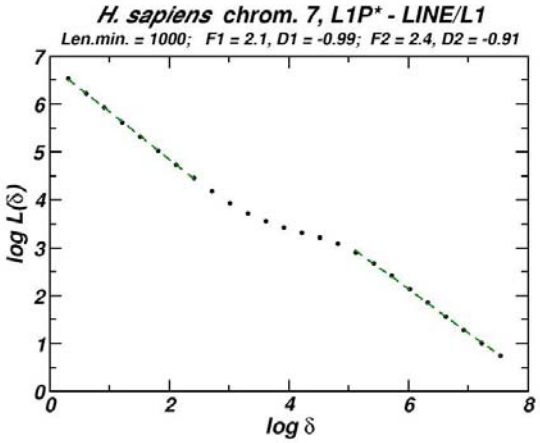

R. norvegicus chrom. 1, L2 - LINE/L2
Len.min. = 50; F1 = 1.2, D1 = -0.94; F2 = 2.7, D2 = -0.92
log L(δ)
log δ

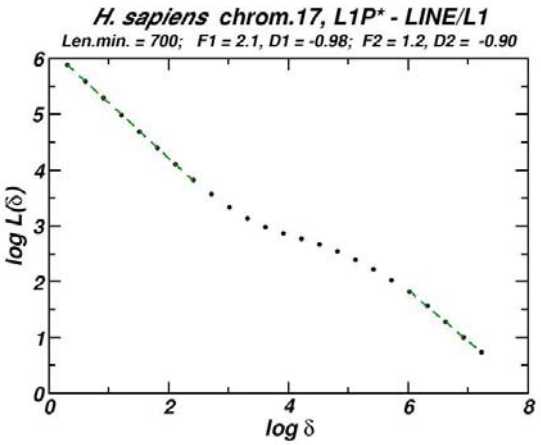

H. sapiens chrom.19, AluJb - SINE/Alu
Div.max. = 20%; F1 = 1.2, D1 = -0.96; F2 = 2.4, D2 = -0.95
log L(δ)
log δ

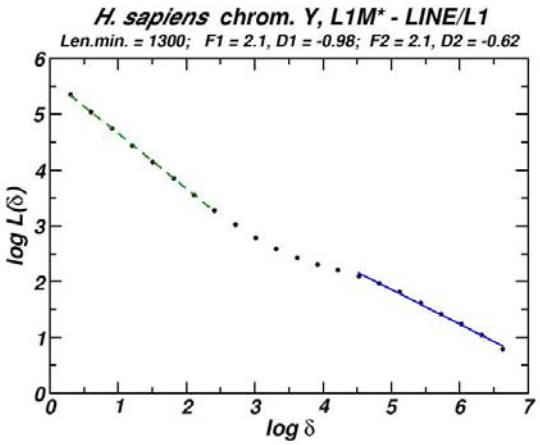

H. sapiens chrom. 9, AluJo - SINE/Alu
Div.max. = 15%; F1 = 1.5, D1 = -0.94; F2 = 1.8, D2 = -0.81
log L(δ)
log δ

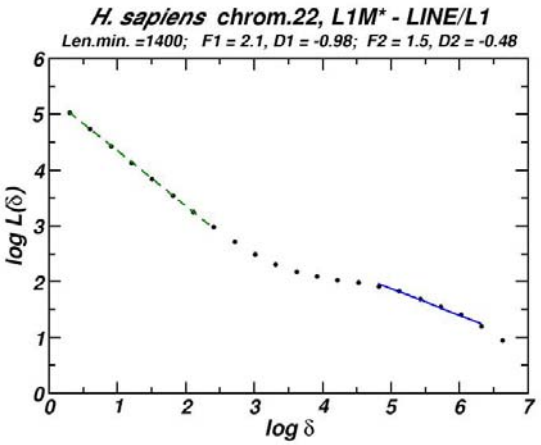

H. sapiens chrom. Y, AluSx - SINE/Alu
Div.max. = 10%; F1 = 1.5, D1 = -0.95; F2 = 2.7, D2 = -0.52
log L(δ)
log δ


H. sapiens chrom. 5, AluSx - SINE/Alu
Div.max. = 9%; F1 = 1.5, D2 = -0.95; F2 = 1.8, D2 = -0.84
log L(δ)
log δ

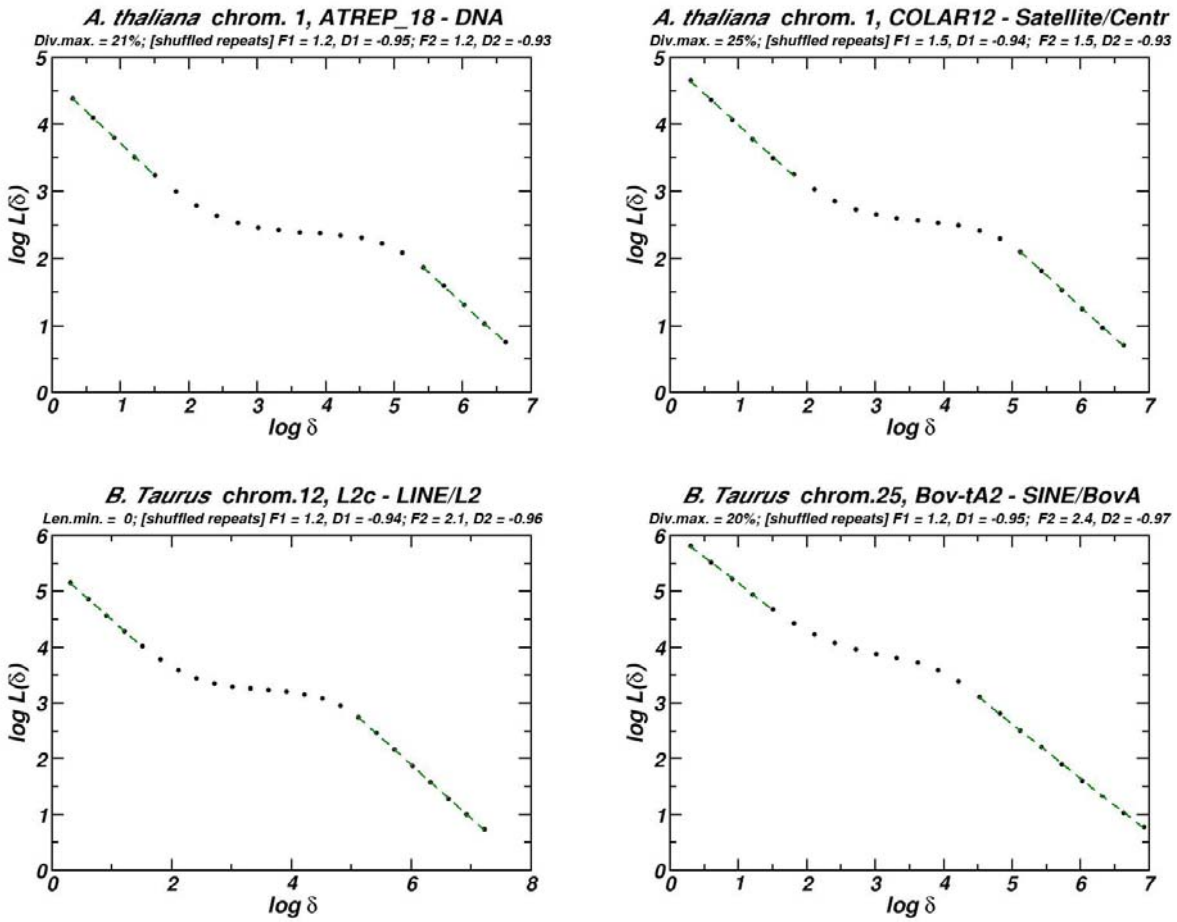

H. sapiens chrom. 7, L1P* - LINE/L1
Len.min. = 1000; F1 = 2.1, D1 = -0.99; F2 = 2.4, D2 = -0.91
log L(δ)
log δ

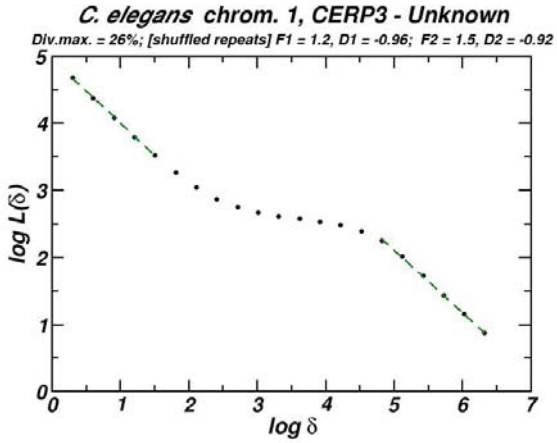

H. sapiens chrom.17, L1P* - LINE/L1
Len.min. = 700; F1 = 2.1, D1 = -0.98; F2 = 1.2, D2 = -0.90
log L(δ)
log δ

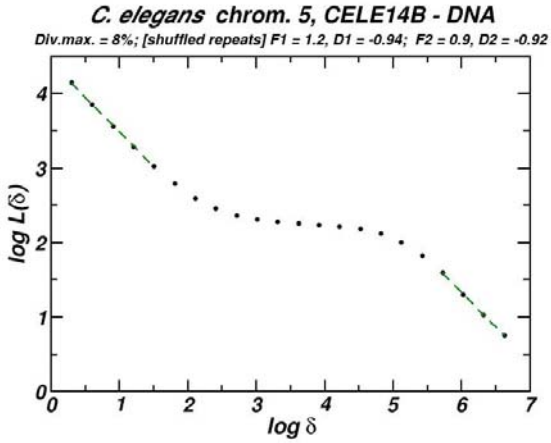

H. sapiens chrom. Y, L1M* - LINE/L1
Len.min. = 1300; F1 = 2.1, D1 = -0.98; F2 = 2.1, D2 = -0.62
log L(δ)
log δ

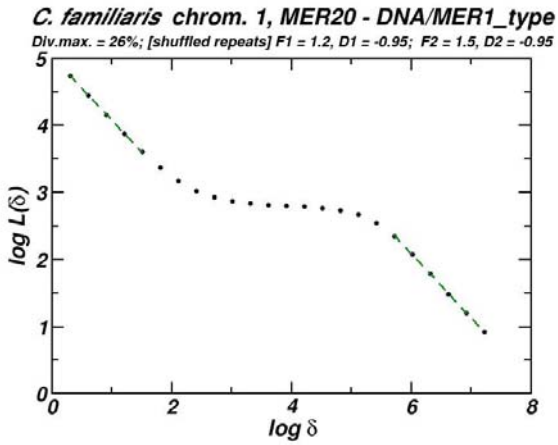

H. sapiens chrom.22, L1M* - LINE/L1
Len.min. =1400; F1 = 2.1, D1 = -0.98; F2 = 1.5, D2 = -0.48
log L(δ)
log δ

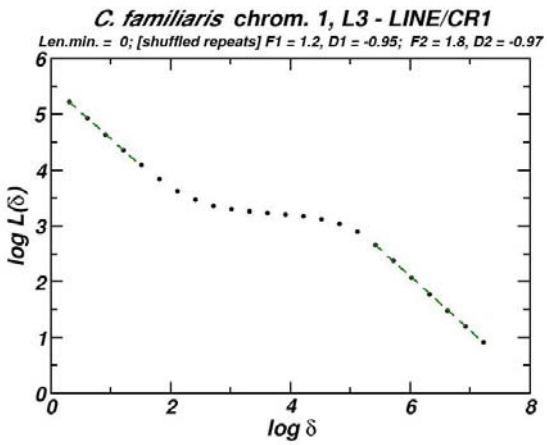
4. Shuffling (Random Rearrangement) of TE Population Inside the Initial Chromosome

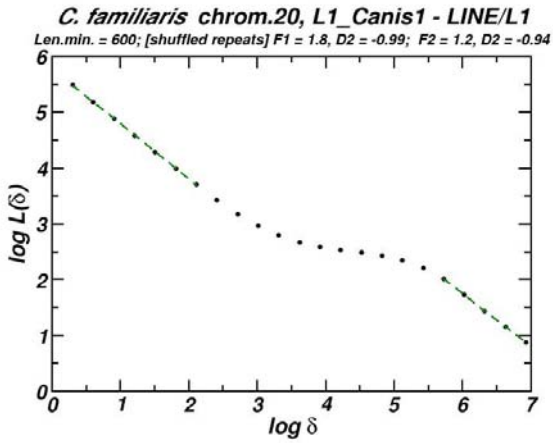


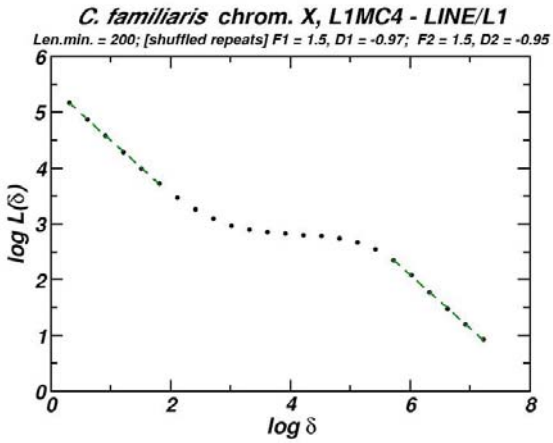

C. elegans chrom. 1, CERP3 - Unknown
Div.max. = 26%; [shuffled repeats] F1 = 1.2, D1 = -0.96; F2 = 1.5, D2 = -0.92
log L(δ)
log δ


C. elegans chrom. 5, CELE14B - DNA
Div.max. = 8%; [shuffled repeats] F1 = 1.2, D1 = -0.94; F2 = 0.9, D2 = -0.92
log L(δ)
log δ

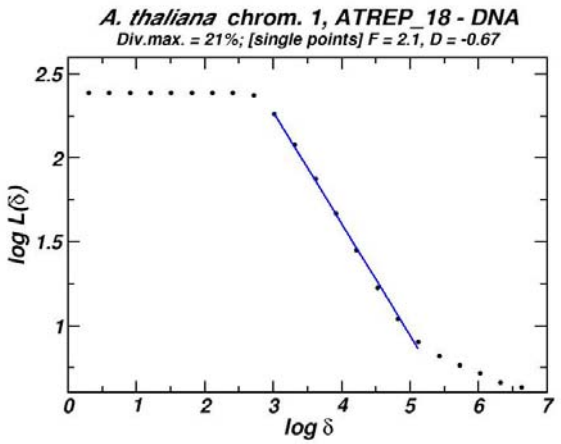

C. familiaris chrom. 1, MER20 - DNA/MER1_type
Div.max. = 26%; [shuffled repeats] F1 = 1.2, D1 = -0.95; F2 = 1.5, D2 = -0.95
log L(δ)
log δ

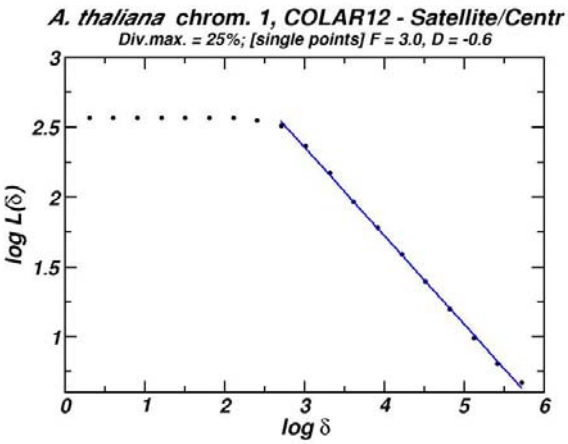

C. familiaris chrom. 1, L3 - LINE/CR1
Len.min. = 0; [shuffled repeats] F1 = 1.2, D1 = -0.95; F2 = 1.8, D2 = -0.97
log L(δ)
log δ

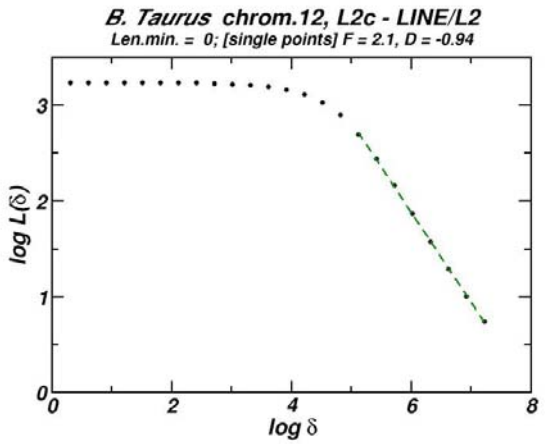

C. familiaris chrom.20, L1_Canis1 - LINE/L1
Len.min. = 600; [shuffled repeats] F1 = 1.8, D2 = -0.99; F2 = 1.2, D2 = -0.94
log L(δ)
log δ

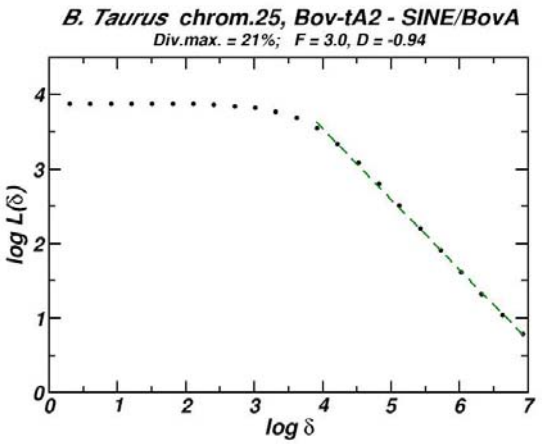

C. familiaris chrom. X, L1MC4 - LINE/L1
Len.min. = 200; [shuffled repeats] F1 = 1.5, D1 = -0.97; F2 = 1.5, D2 = -0.95
log L(δ)
log δ

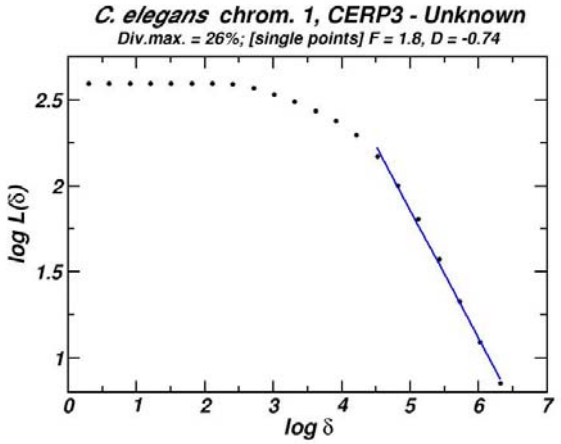
Replacement of Each Repeat by a Single "1" Symbol

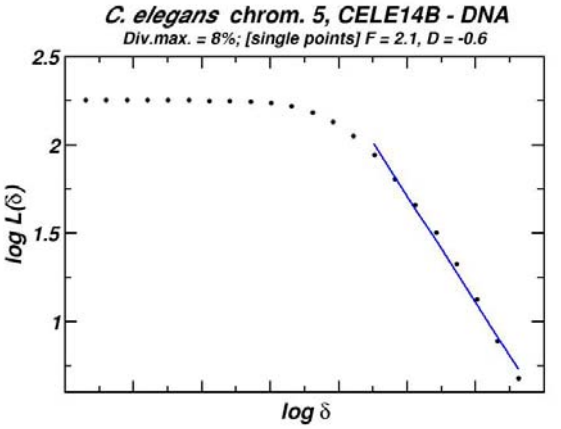


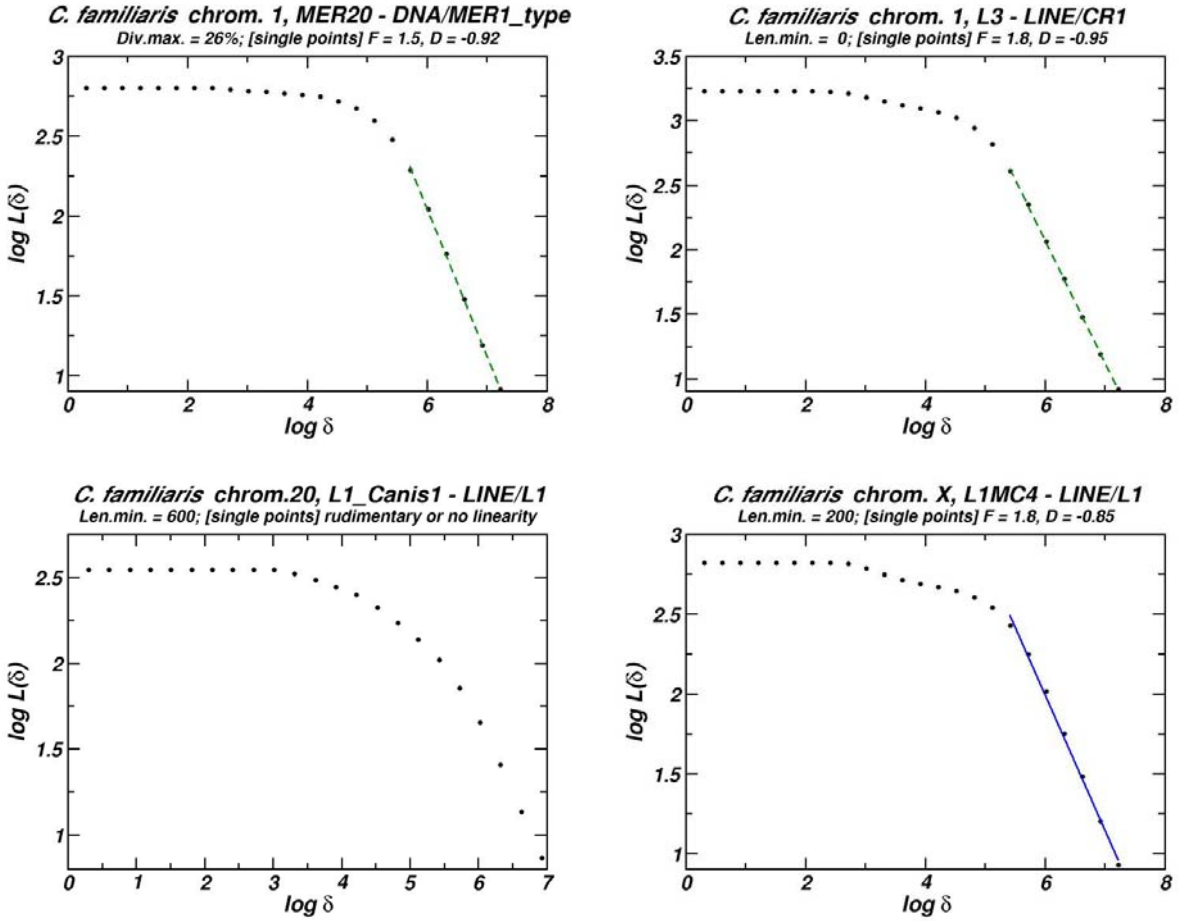




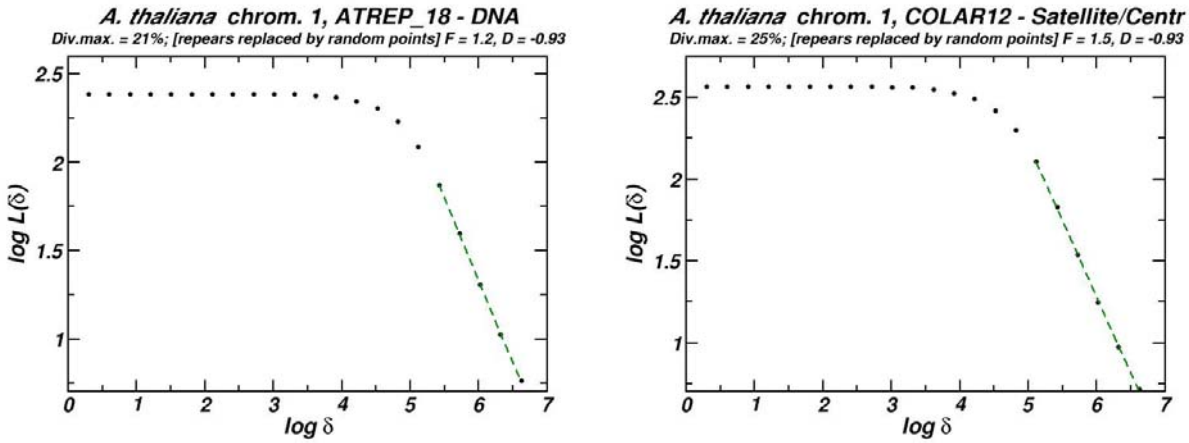


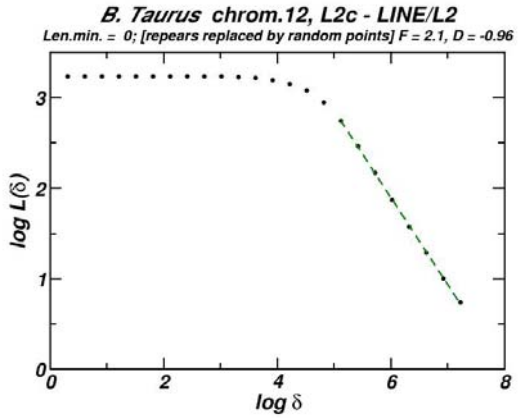


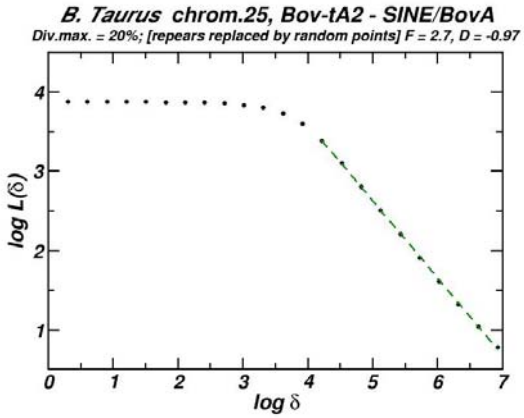


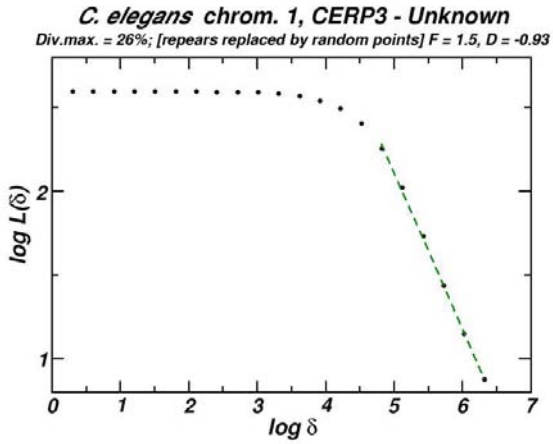


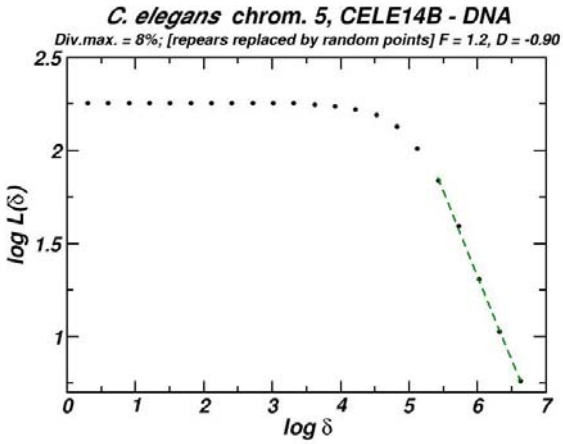
Replacement of Each Repeat by a Single "1" Symbol Followed by Shuffling (Random Rearrangement) of the Repeat Population Inside the Initial Chromosome

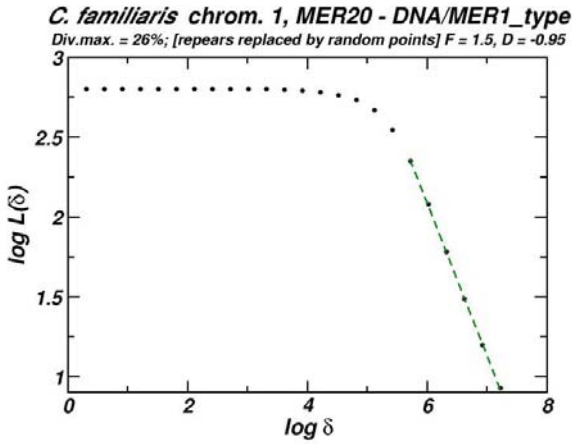


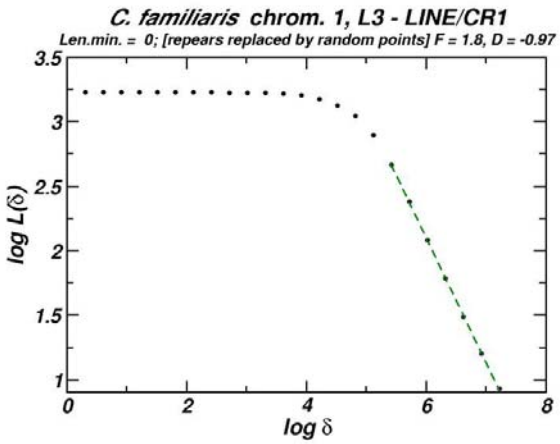

B. Taurus chrom.12, L2c - LINE/L2
Len.min. = 0; [repears replaced by random points] F = 2.1, D = -0.96
log L(δ)
log δ

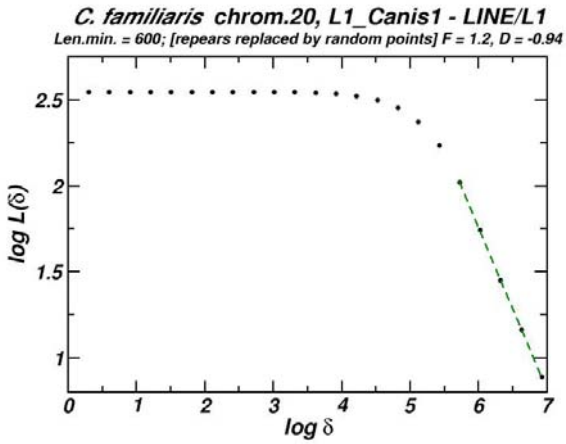

B. Taurus chrom.25, Bov-tA2 - SINE/BovA
Div.max. = 20%; [repears replaced by random points] F = 2.7, D = -0.97
log L(δ)
log δ

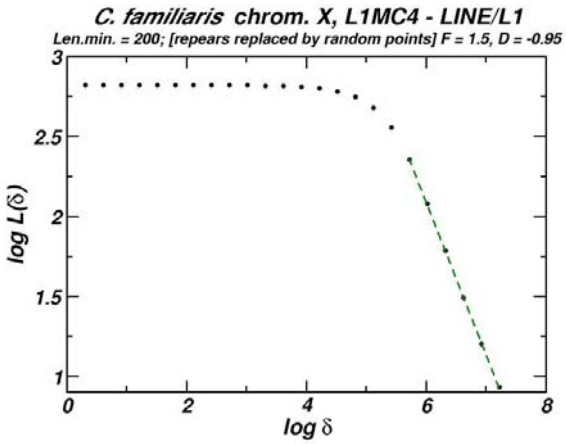

C. elegans chrom. 1, CERP3 - Unknown
Div.max. = 26%; [repears replaced by random points] F = 1.5, D = -0.93
log L(δ)
log δ


C. elegans chrom. 5, CELE14B - DNA
Div.max. = 8%; [repears replaced by random points] F = 1.2, D = -0.90
log L(δ)
log δ

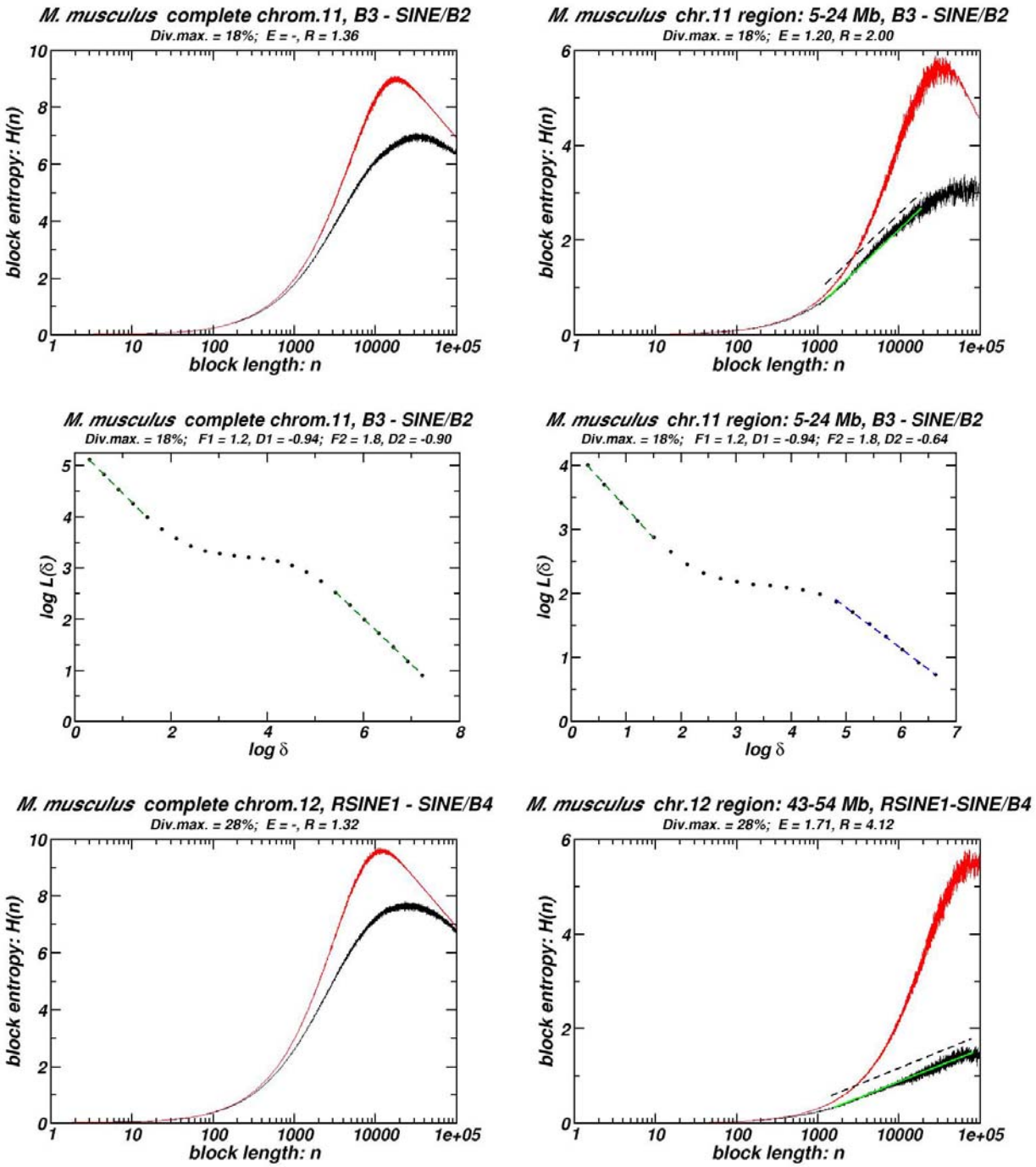

C. familiaris chrom. 1, MER20 - DNA/MER1_type
Div.max. = 26%; [repears replaced by random points] F = 1.5, D = -0.95
log L(δ)
log δ

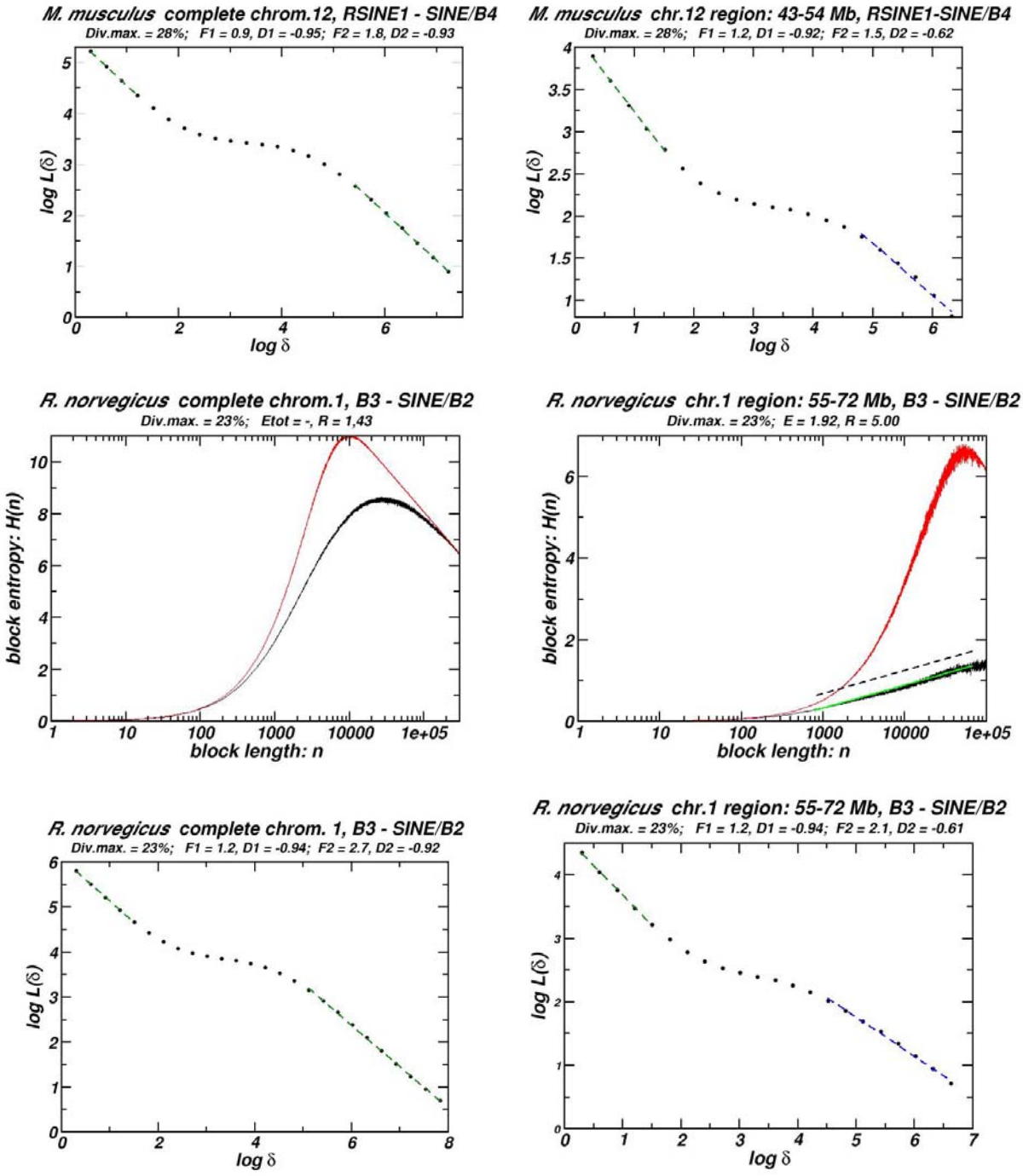

C. familiaris chrom. 1, L3 - LINE/CR1
Len.min. = 0; [repears replaced by random points] F = 1.8, D = -0.97
log L(δ)
log δ


C. familiaris chrom.20, L1_Canis1 - LINE/L1
Len.min. = 600; [repears replaced by random points] F = 1.2, D = -0.94
log L(δ)
log δ


C. familiaris chrom. X, L1MC4 - LINE/L1
Len.min. = 200; [repears replaced by random points] F = 1.5, D = -0.95
log L(δ)
log δ

5. Three Cases of Chromosomal Regions with High SIM *vs*. the Hosting Chromosomes, Not Exhibiting Fractality: Supplement to Figure 4





These are three cases where a chromosomal region is studied (entropic scaling and box-counting included) in conjunction with the whole chromosome. In all these cases along with the one presented in Figure 4, the chromosome as a whole does not show any trace of fractality. Chromosomal regions are selected with the criterion of having a particularly high SIM content. We observe that all these regions, when studied in isolation, present fractality quantified by both the slope D in box-counting, and a clear linearity in their entropic scaling combined with high R values. These findings are in favor of the hypothesis underlying the "insertion – elimination model" that high rates of insertions subsequent to the studied TE population contribute to the formation of a fractal-like structure.

6. Details of the Presented Simulations of the Insertion-Elimination Model

Examples of simulations using the insertion-elimination model are given in Figure 6. In these numerical experiments, in an artificial chromosome initially $20 \bullet 10^6$ nt (20 Mnt) long, 20,000 markers representing a TE population are randomly distributed. After consecutive rounds of marker eliminations and influx of external sequence segments, mostly representing insertions of TEs of more recent types (200 nt long in the presented simulations), a relatively small fraction of the initial number of markers is left. Their spatial arrangement is studied by means of entropic scaling and box-counting. In plots depicting entropic scaling, a (red or gray) curve corresponding to a sequence with randomly distributed markers is shown along with the black curve corresponding to the model generated artificial sequence. Both sequences are of the same length and host equal numbers of markers. In plots 6a,b we present the results for a model simulation experiment where 10 influx (insertion) events follow each marker-elimination (spacers' merging) event, until a population of 500 markers is left. In the model-simulation numerical experiment depicted in plots 6c,d 20 influx (insertion) events follow each marker-elimination (spacers' merging) event, until again a population of 500 markers is left. Thus, increase of the sequence material inserted during the "maturation" of a given TE population in a chromosome (named SIM in our analysis) leads to a higher degree of fractality, as quantified by both methods used herein. When we stop the action of the model before reducing to 500 the number of the markers left, fractality of the resulting sequences is lower (plots not shown). We conclude that either a more intense rate of younger repeat families' insertion or a longer time passed after the initial spreading of the repeat population are in favor of the emergence of a fractal pattern. As shown in Figure 2 of reference [10], where the same model settings were used, a more extended linearity in the power-law-like distribution of the inter-repeat distances is also attained for more "mature" TE populations.

In plots 6e,f the model-derived sequence whose features are depicted in c & d is exposed to 2000 events of random transpositions (genomic cut-transpose-and-paste events). These events, belonging to type (**c**) according to the model description, lead to a slow evolutionary shuffling of the genome. The plots show that this amount of shuffling is sufficient to completely destroy the fractality created after the action of the insertion-elimination model. The study of intermediate numbers of transposition events, through entropic scaling and box-counting, showed a progressive decomposition of the initial fractal pattern (figures not included).