ORGANIZACIÓN EN ISOCOROS DEL GENOMA HUMANO

Héctor Musto

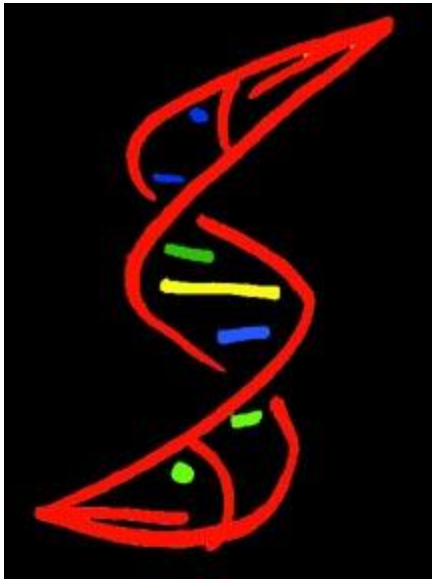
Laboratorio Genómica Evolutiva

Departamento Biología Celular y Molecular

Facultad de Ciencias

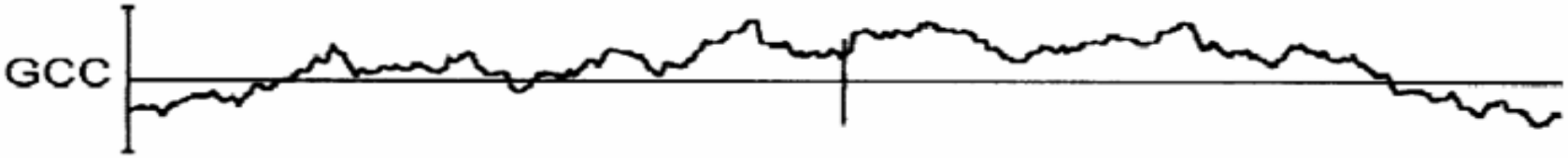
Uruguay

hmusto@gmail.com

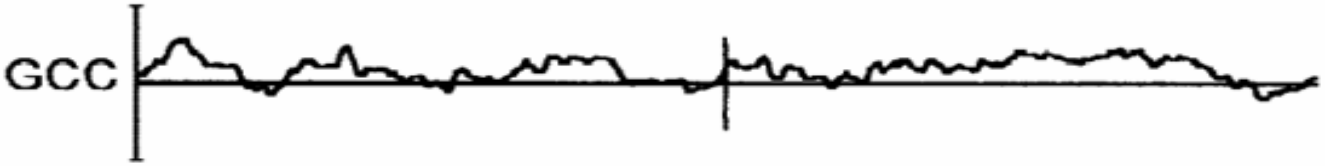


Giorgio Bernardi (1930-2021)

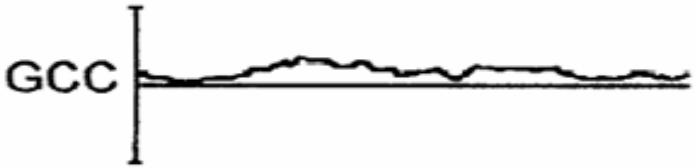
Variación en el contenido en GC% en procariotas



E. coli

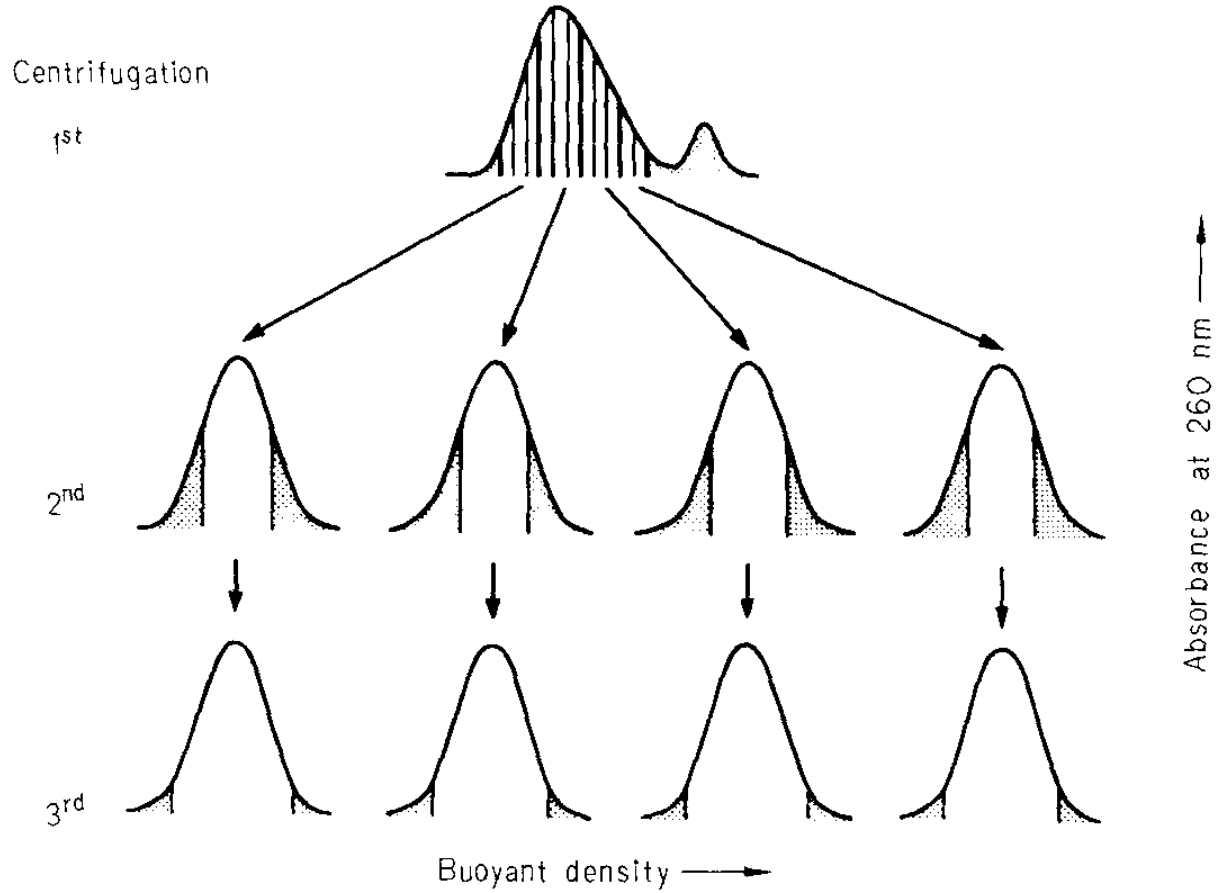


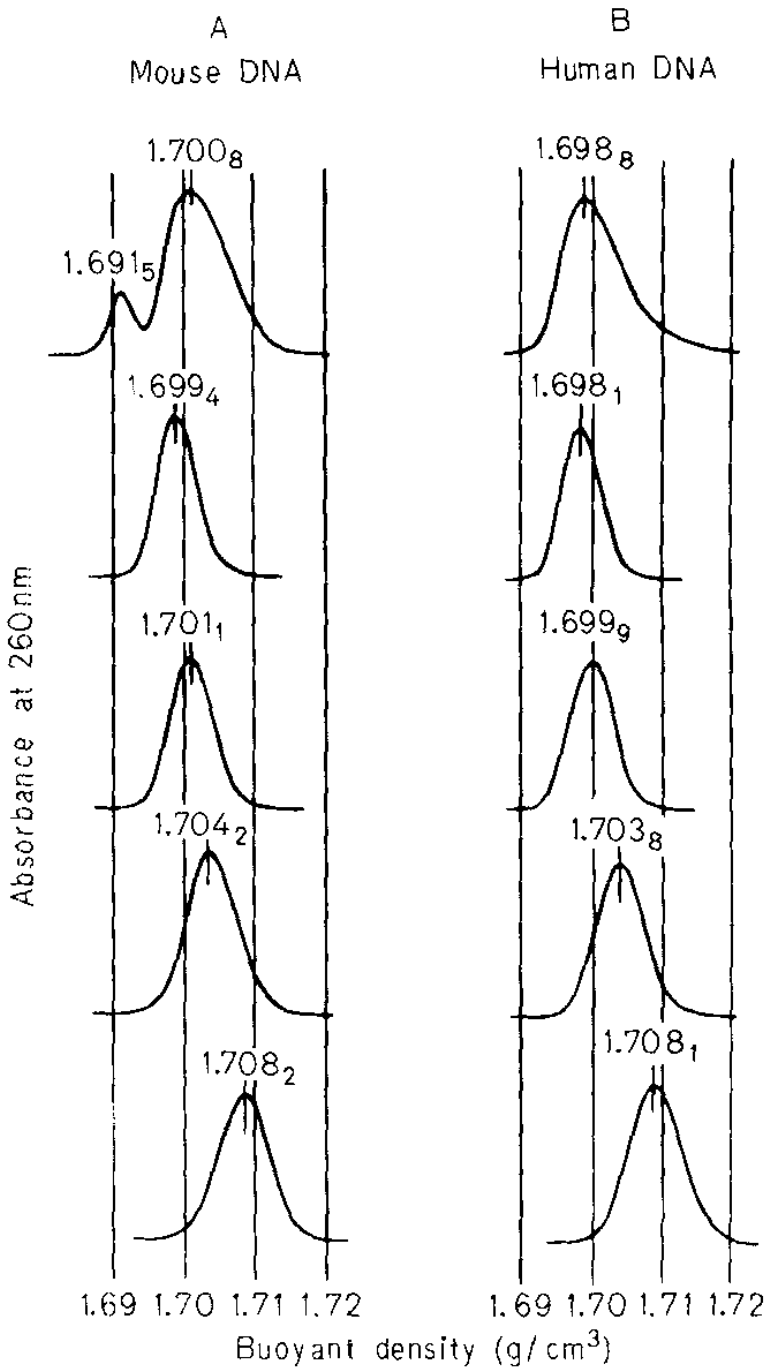
Synechocystis sp.

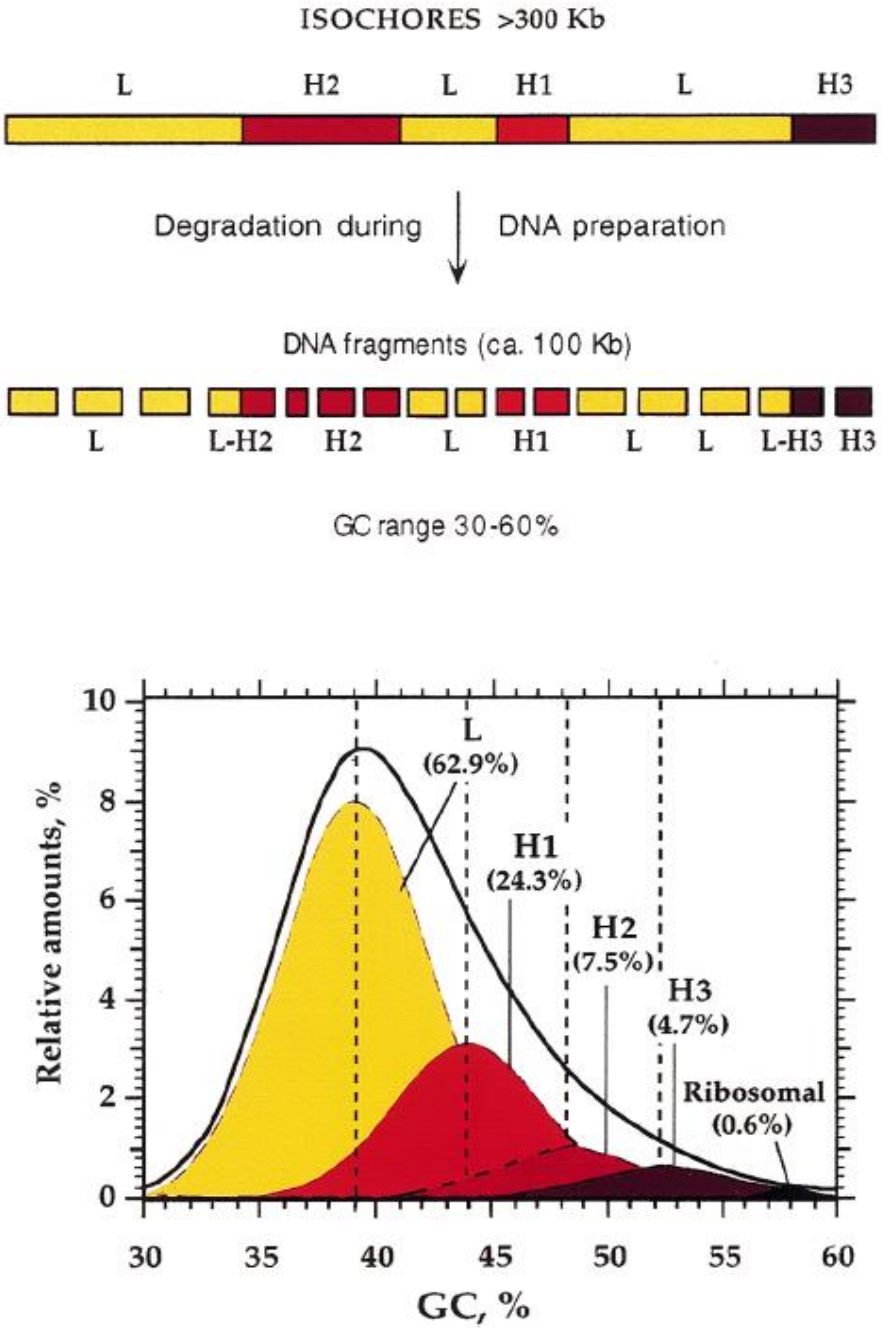


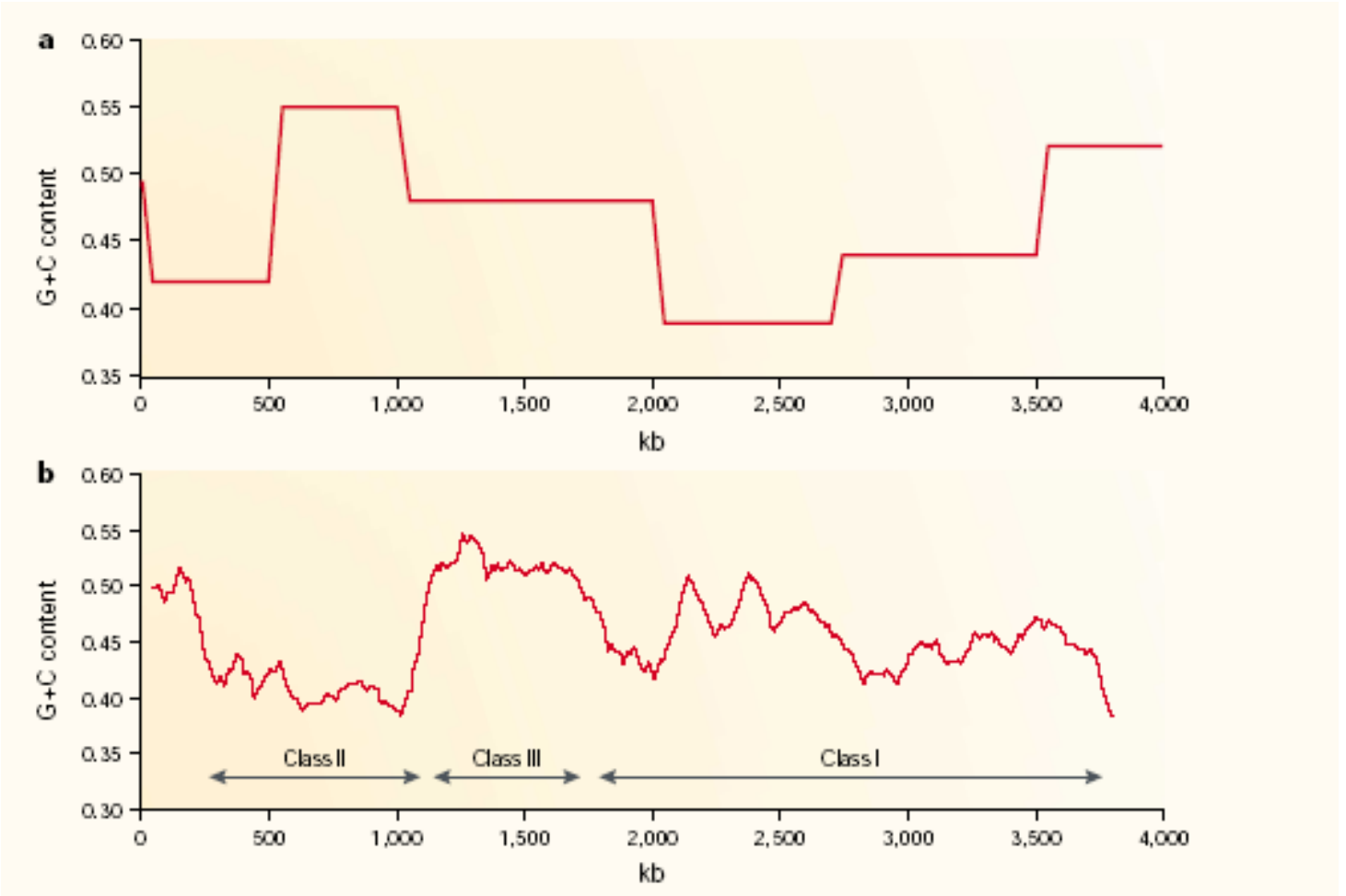
M. jannaschii

Preparación de componentes mayores del genoma de ratón

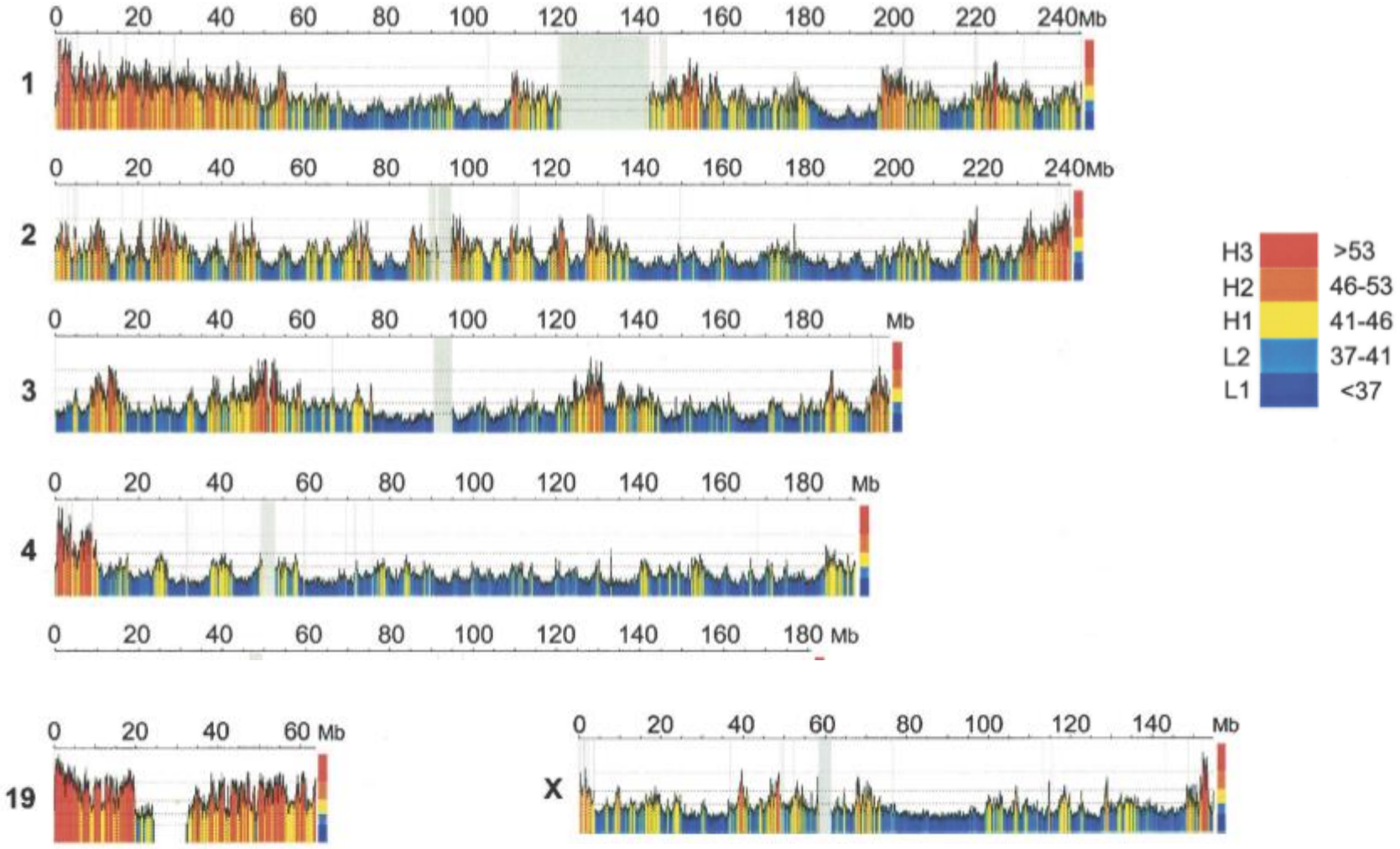



A
Mouse DNA
B
Human DNA
1.700 8
1.691 5
1.699 4
1.701 1
1.704 2
1.708 2
1.698 8
1.698 1
1.699 9
1.703 8
1.708 1
Absorbance at 260nm
1.69 1.70 1.71 1.72
1.69 1.70 1.71 1.72
Buoyant density (g/cm3)


ISOCHORES >300 Kb
L
H2
L
H1
L
H3
Degradation during
DNA preparation
DNA fragments (ca. 100 Kb)
L
L-H2
H2
L
H1
L
L
L-H3
H3
GC range 30-60%
Relative amounts, %
GC, %
L
(62.9%)
H1
(24.3%)
H2
(7.5%)
H3
(4.7%)
Ribosomal
(0.6%)


a
0.60
0.55
0.50
0.45
0.40
0.35
G+C content
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
kb
b
0.60
0.55
0.50
0.45
0.40
0.35
0.30
G+C content
Class II
Class III
Class I
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
kb

Visualización de los isocoros por métodos computacionales



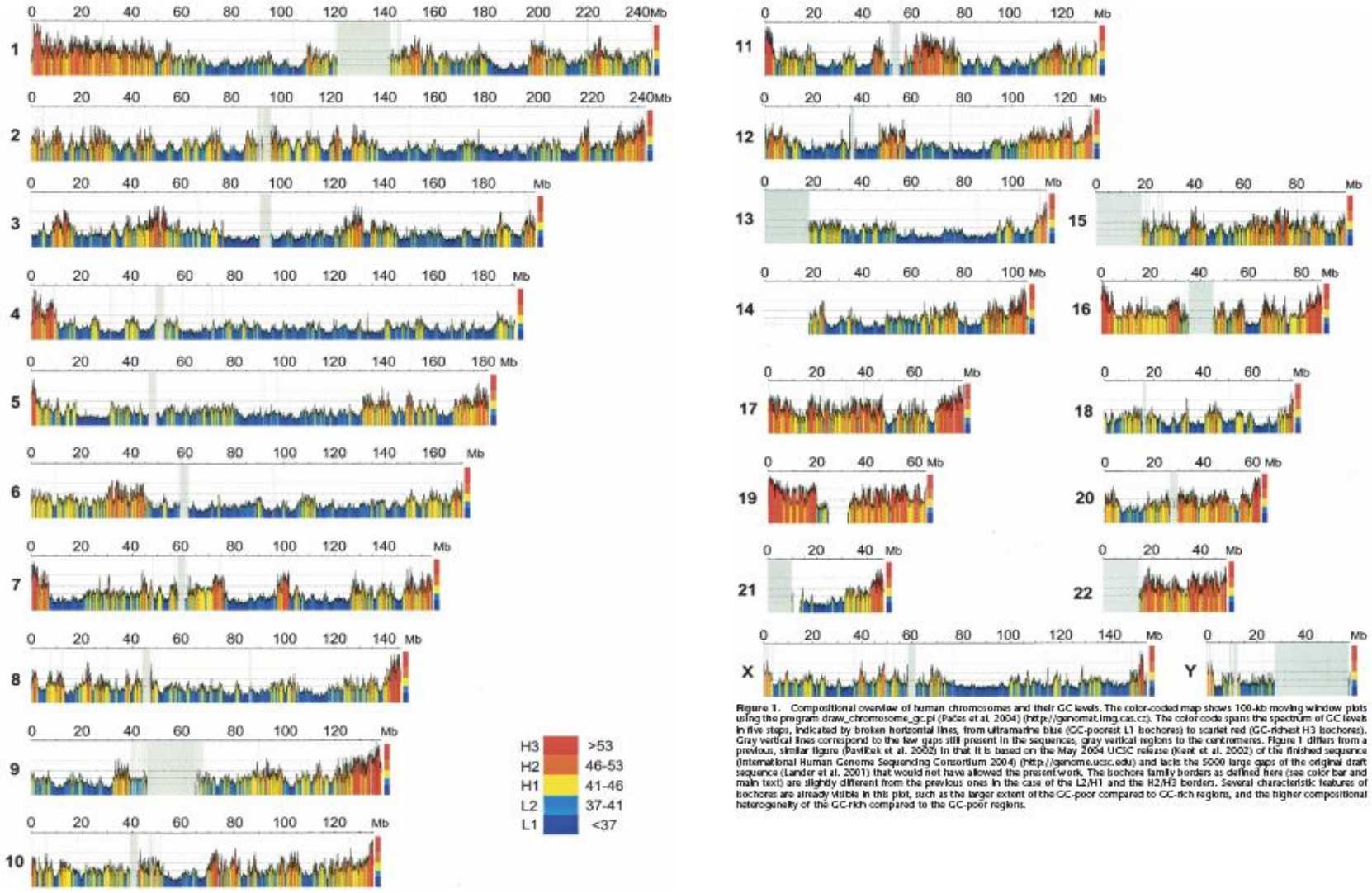
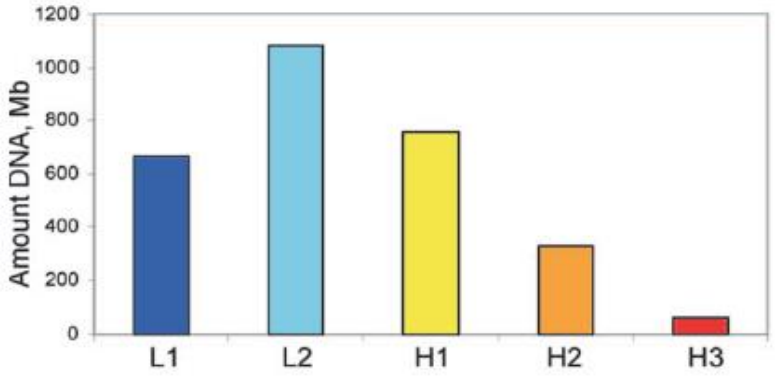
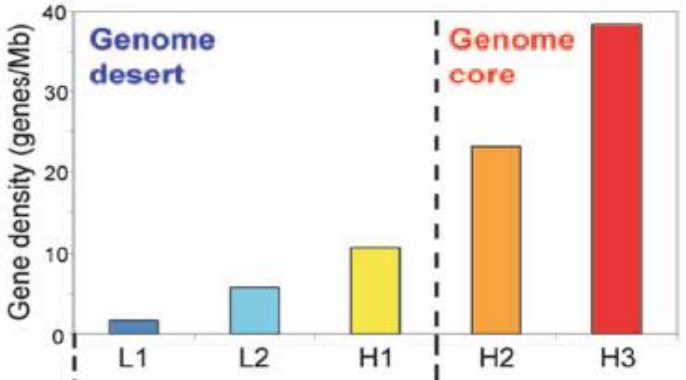


Figure 1. Compositional overview of human chromosomes and their GC levels. The color-coded map shows 100-kb moving window plots using the program draw_chromosome_gc.pl (Paces et al. 2004) (http://genomat.img.cas.cz). The color code spans the spectrum of GC levels in five steps, indicated by broken horizontal lines, from ultramarine blue (GC-poorest L1 isochores) to scarlet red (GC-richest H3 isochores). Gray vertical lines correspond to the few gaps still present in the sequences, gray vertical regions to the centromeres. Figure 1 differs from a previous, similar figure (Pavlicek et al. 2002) in that it is based on the May 2004 UCSC release (Kent et al. 2002) of the finished sequence (International Human Genome Sequencing Consortium 2004) (http://genome.ucsc.edu) and lacks the 5000 large gaps of the original draft sequence (Lander et al. 2001) that would not have allowed the present work. The isochore family borders as defined here (see color bar and main text) are slightly different from the previous ones in the case of the L2/H1 and the H2/H3 borders. Several characteristic features of isochores are already visible in this plot, such as the larger extent of the GC-poor compared to GC-rich regions, and the higher compositional heterogeneity of the GC-rich compared to the GC-poor regions.

Isochore families



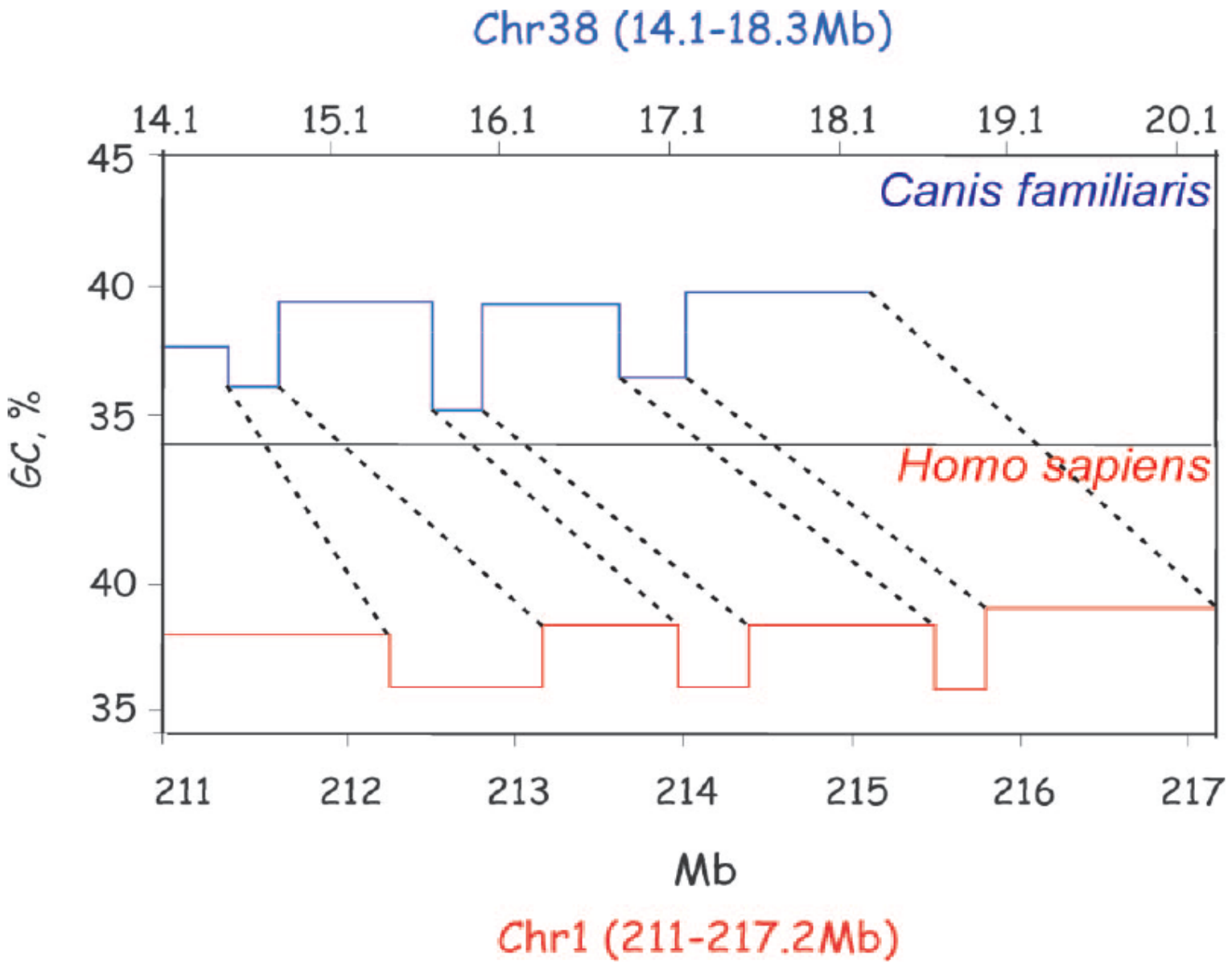
Gene distribution

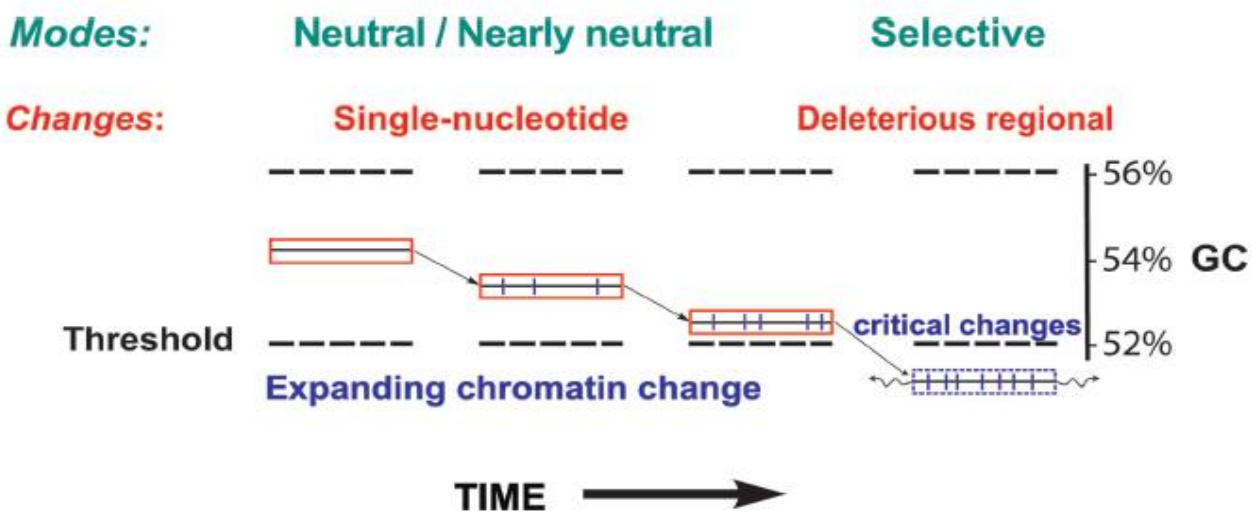


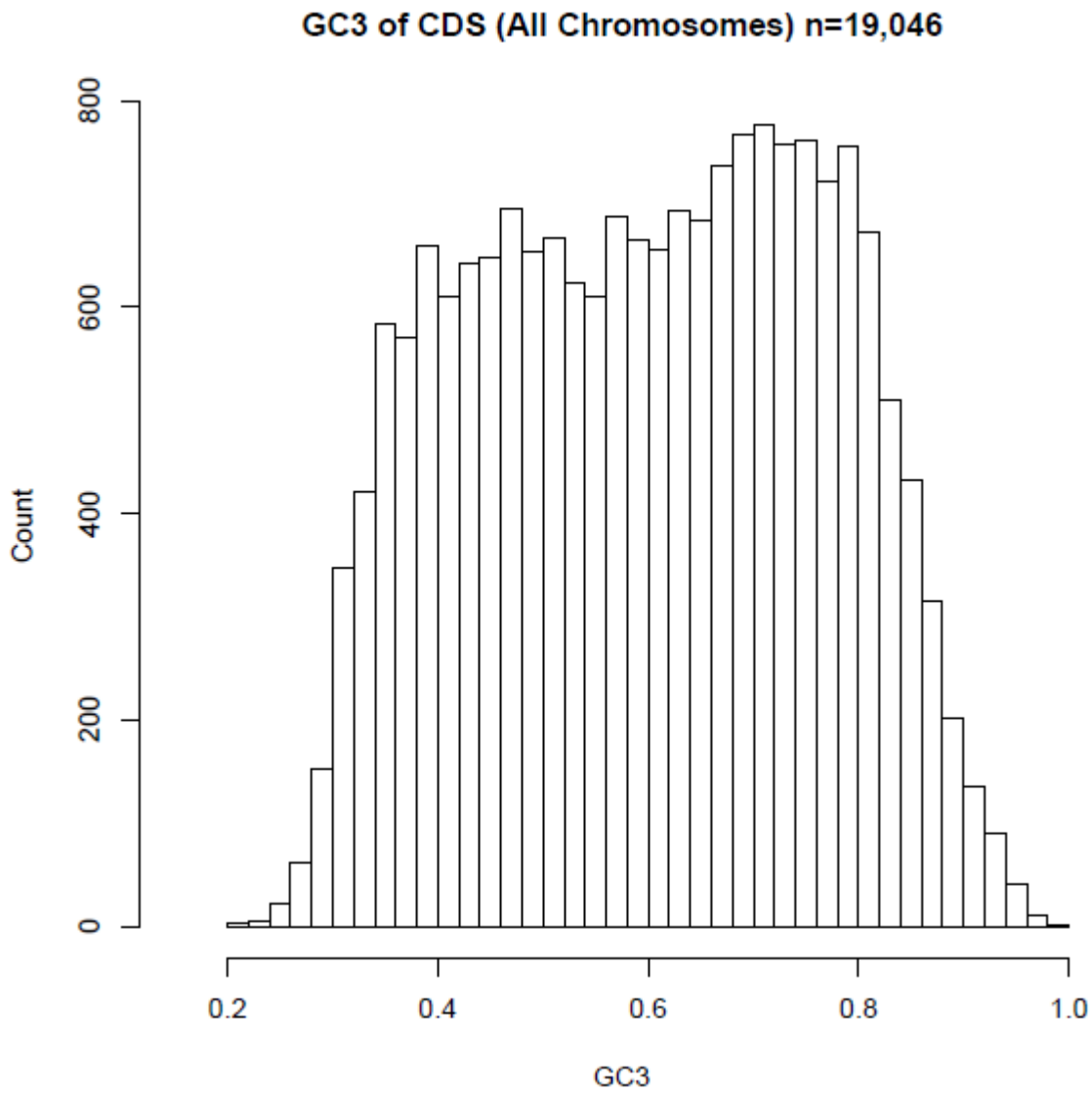
Increasing GC →

Correlations of gene spaces with structures and functions

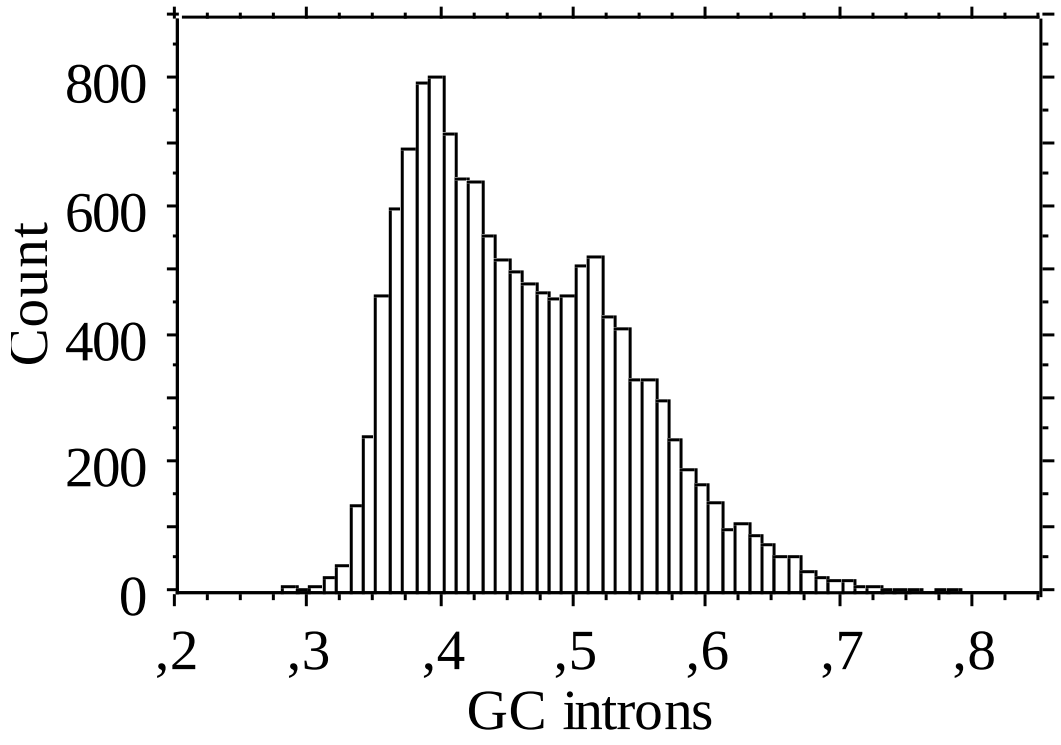
Intron, UTR size	Large	Small
Chromatin structure	Closed	Open
GC heterogeneity	Low	High
SINEs	Low	High
LINEs	High	Low
Gene expression	Low	High
Replication timing	Late	Early
Recombination	Low	High


Chr38 (14.1-18.3Mb)
14.1
15.1
16.1
17.1
18.1
19.1
20.1
45
40
35
40
35
GC, %
Canis familiaris
Homo sapiens
211
212
213
214
215
216
217
Mb
Chr1 (211-217.2Mb)

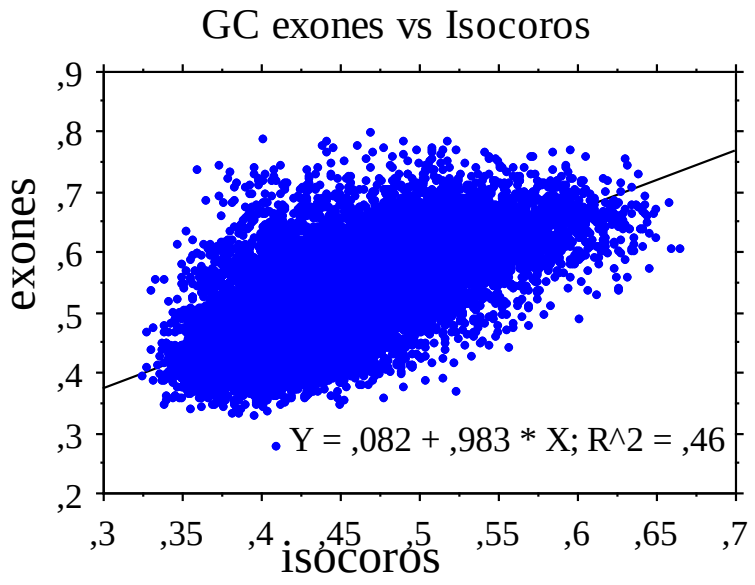

Modes:
Neutral / Nearly neutral
Selective
Changes:
Single-nucleotide
Deleterious regional
56%
54%
GC
52%
Threshold
critical changes
Expanding chromatin change
TIME

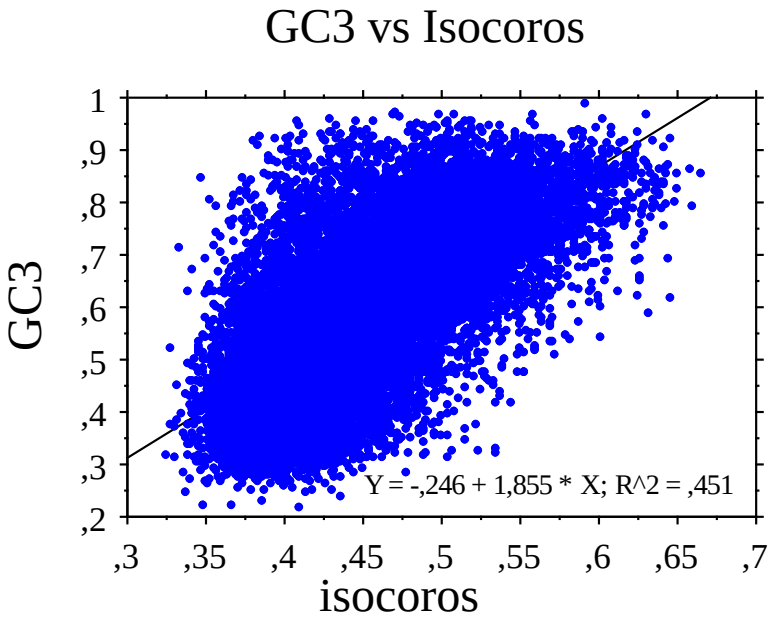

GC3 of CDS (All Chromosomes) n=19,046
Count
0
200
400
600
800
0.2
0.4
0.6
0.8
1.0
GC3

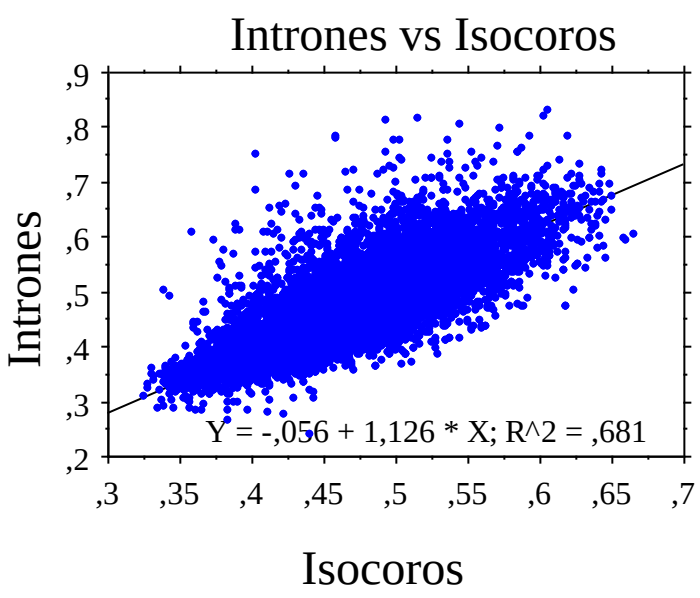
Contenido en GC de los intrones humanos

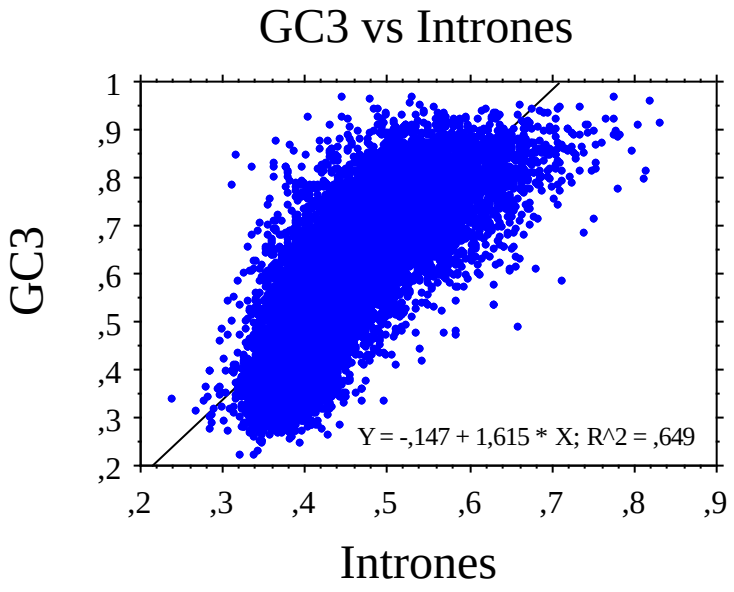


CORRELACIONES COMPOSICIONALES

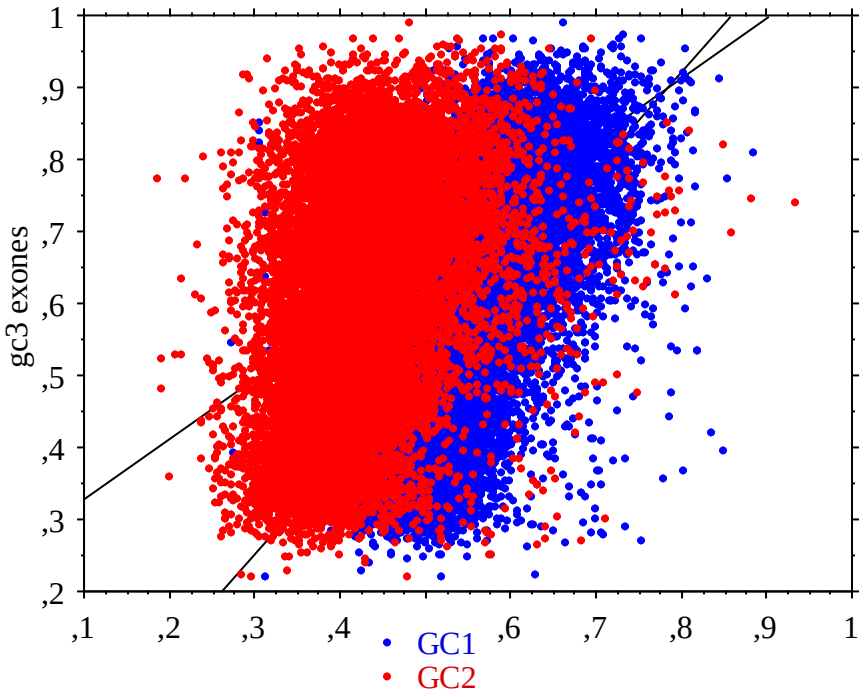




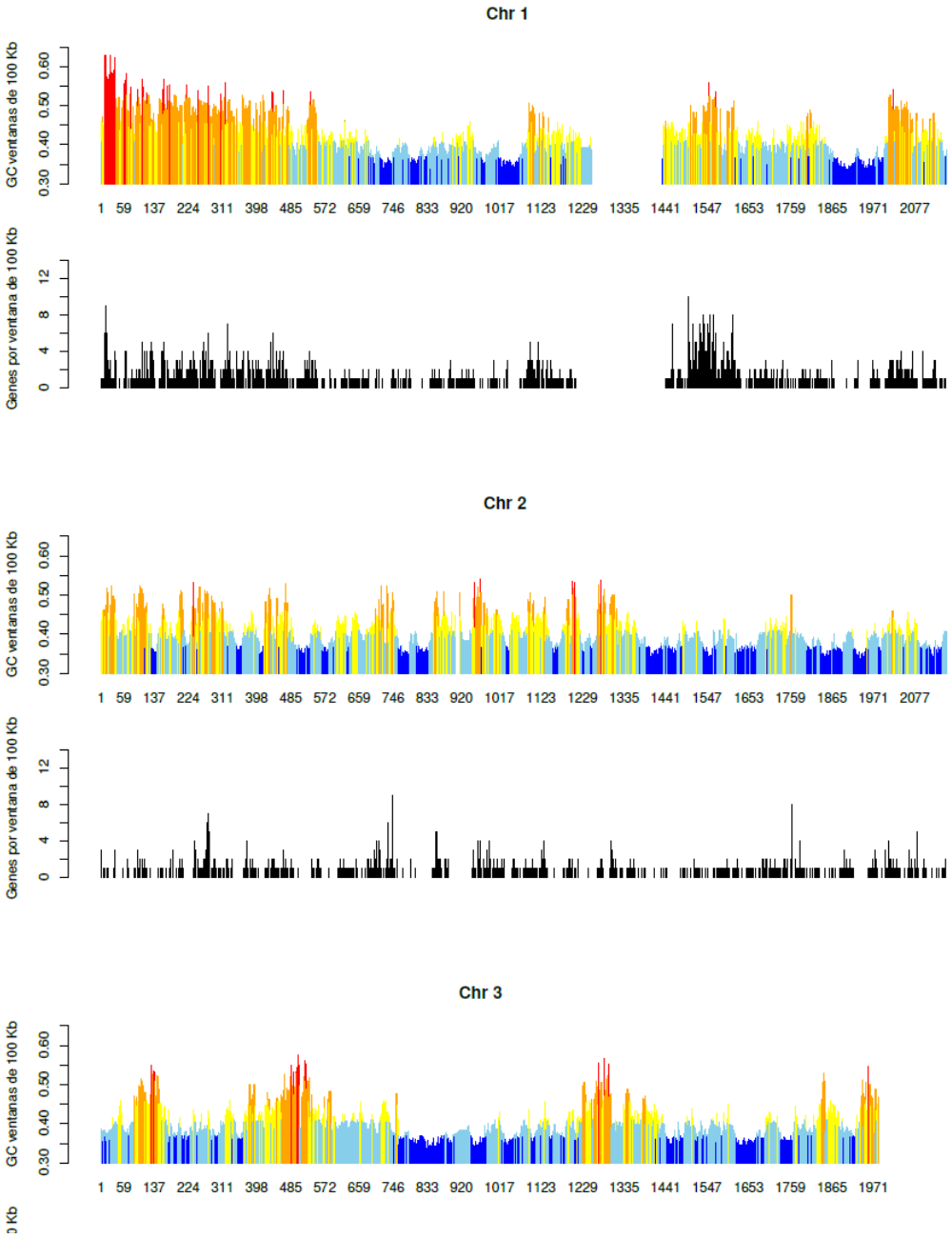

Intrones vs Isocoros
Intrones
,9
,8
,7
,6
,5
,4
,3
,2
,3
,35
,4
,45
,5
,55
,6
,65
,7
Isocoros
Y = -,056 + 1,126 * X; R^2 = ,681

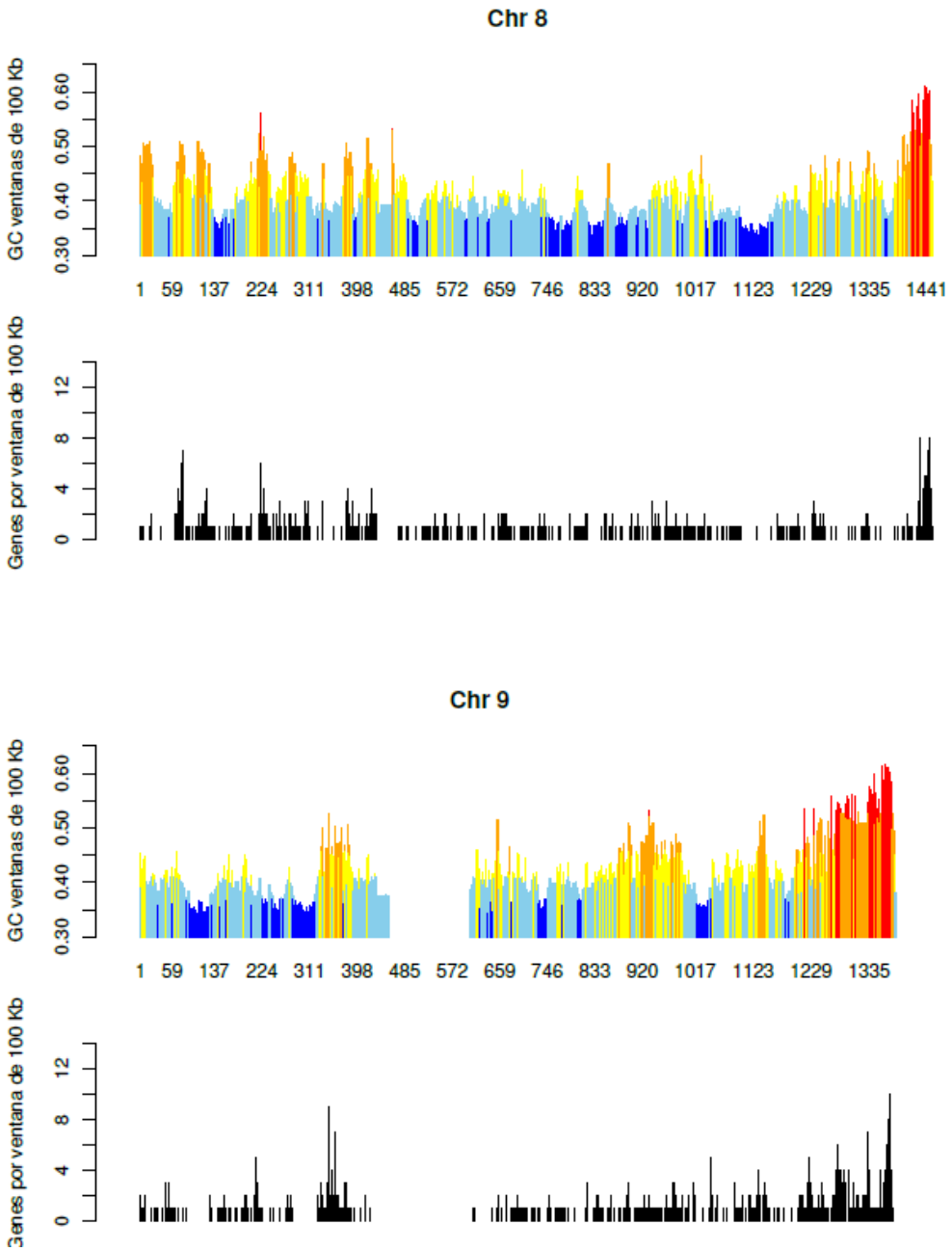

GC3 vs Intrones
GC3
1
,9
,8
,7
,6
,5
,4
,3
,2
,2
,3
,4
,5
,6
,7
,8
,9
Intrones
Y = -,147 + 1,615 * X; R^2 = ,649

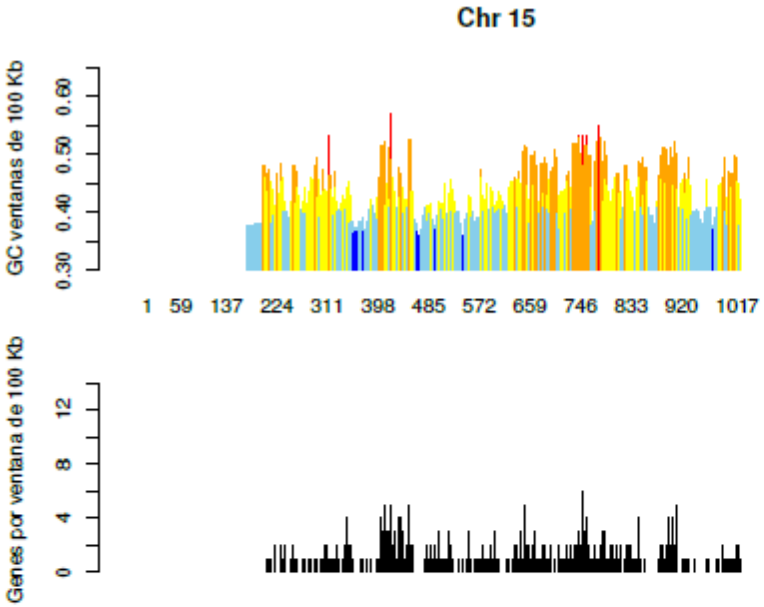
Correlaciones entre GC3 vs GC2 y GC1

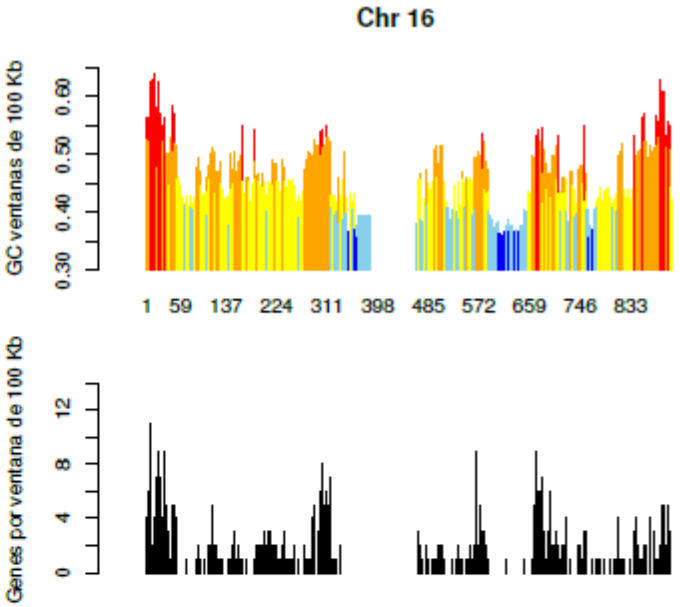


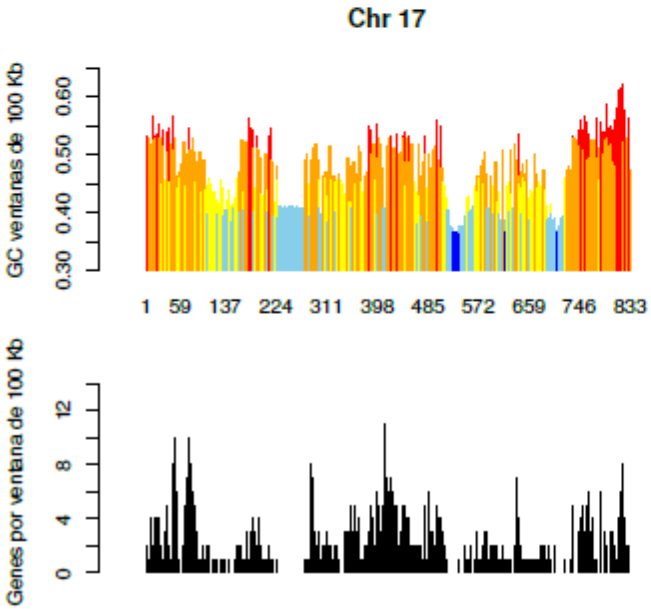
gc3 exones = ,244 + ,837 * GC2; R^2 = ,142

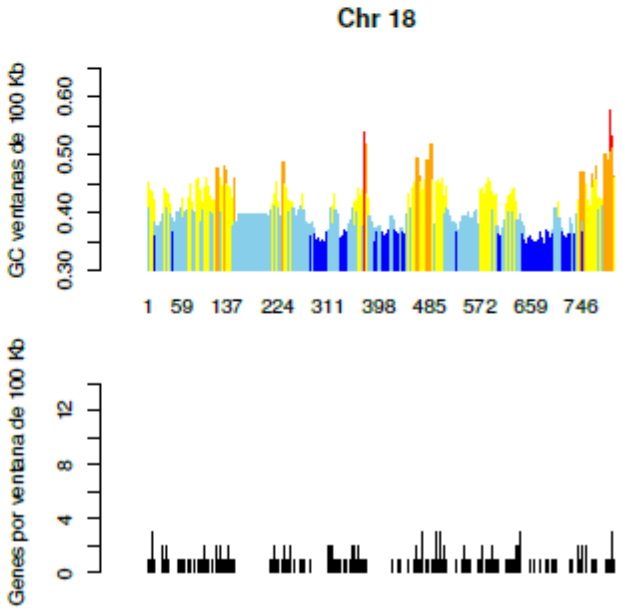

Chr 1
GC ventanas de 100 Kb
0.30 0.40 0.50 0.60
1 59 137 224 311 398 485 572 659 746 833 920 1017 1123 1229 1335 1441 1547 1653 1759 1865 1971 2077
Genes por ventana de 100 Kb
0 4 8 12
Chr 2
GC ventanas de 100 Kb
0.30 0.40 0.50 0.60
1 59 137 224 311 398 485 572 659 746 833 920 1017 1123 1229 1335 1441 1547 1653 1759 1865 1971 2077
Genes por ventana de 100 Kb
0 4 8 12
Chr 3
GC ventanas de 100 Kb
0.30 0.40 0.50 0.60
1 59 137 224 311 398 485 572 659 746 833 920 1017 1123 1229 1335 1441 1547 1653 1759 1865 1971

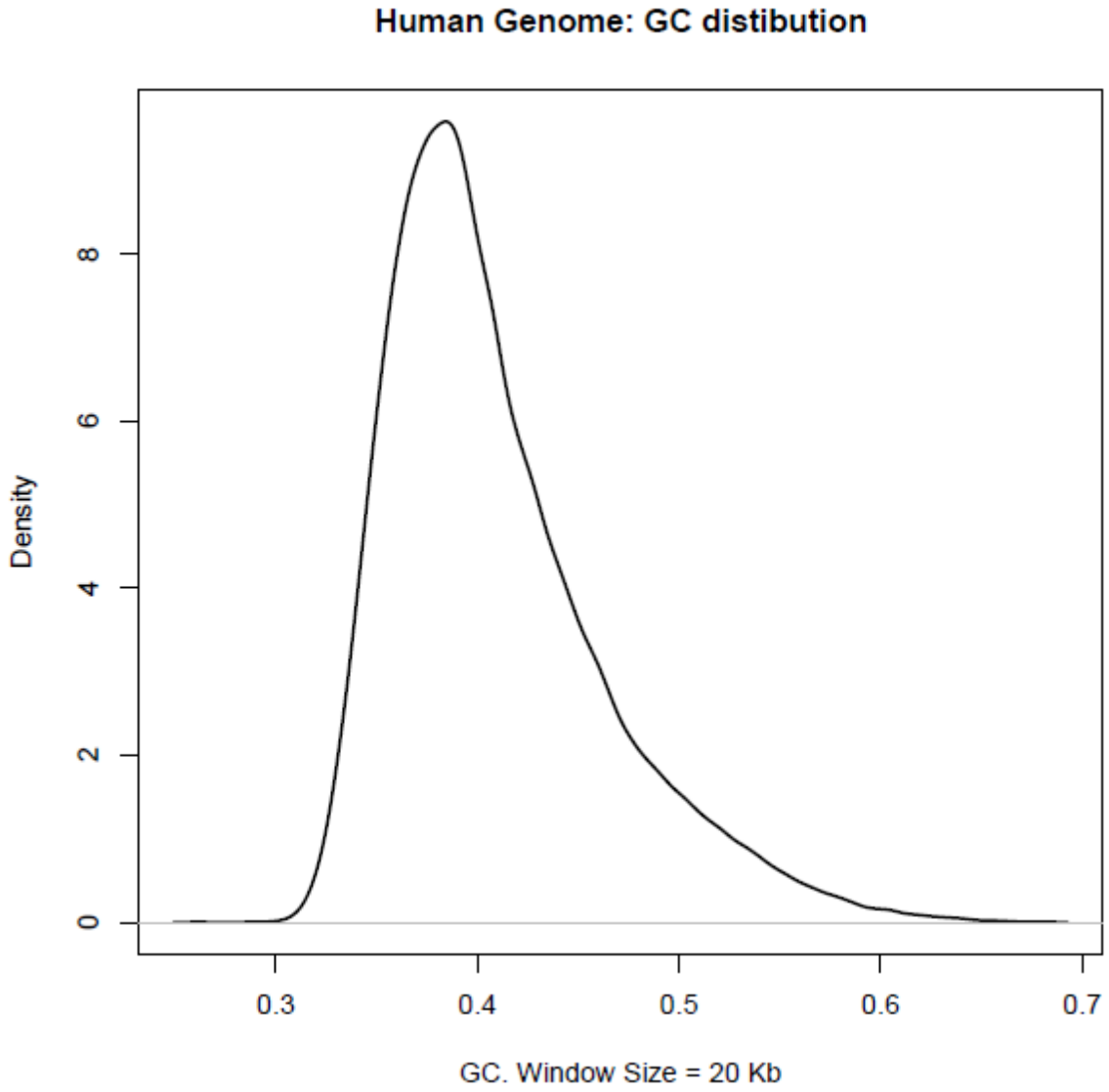

Chr 8
GC ventanas de 100 Kb
0.30
0.40
0.50
0.60
1
59
137
224
311
398
485
572
659
746
833
920
1017
1123
1229
1335
1441
Genes por ventana de 100 Kb
0
4
8
12
Chr 9
GC ventanas de 100 Kb
0.30
0.40
0.50
0.60
1
59
137
224
311
398
485
572
659
746
833
920
1017
1123
1229
1335
Genes por ventana de 100 Kb
0
4
8
12

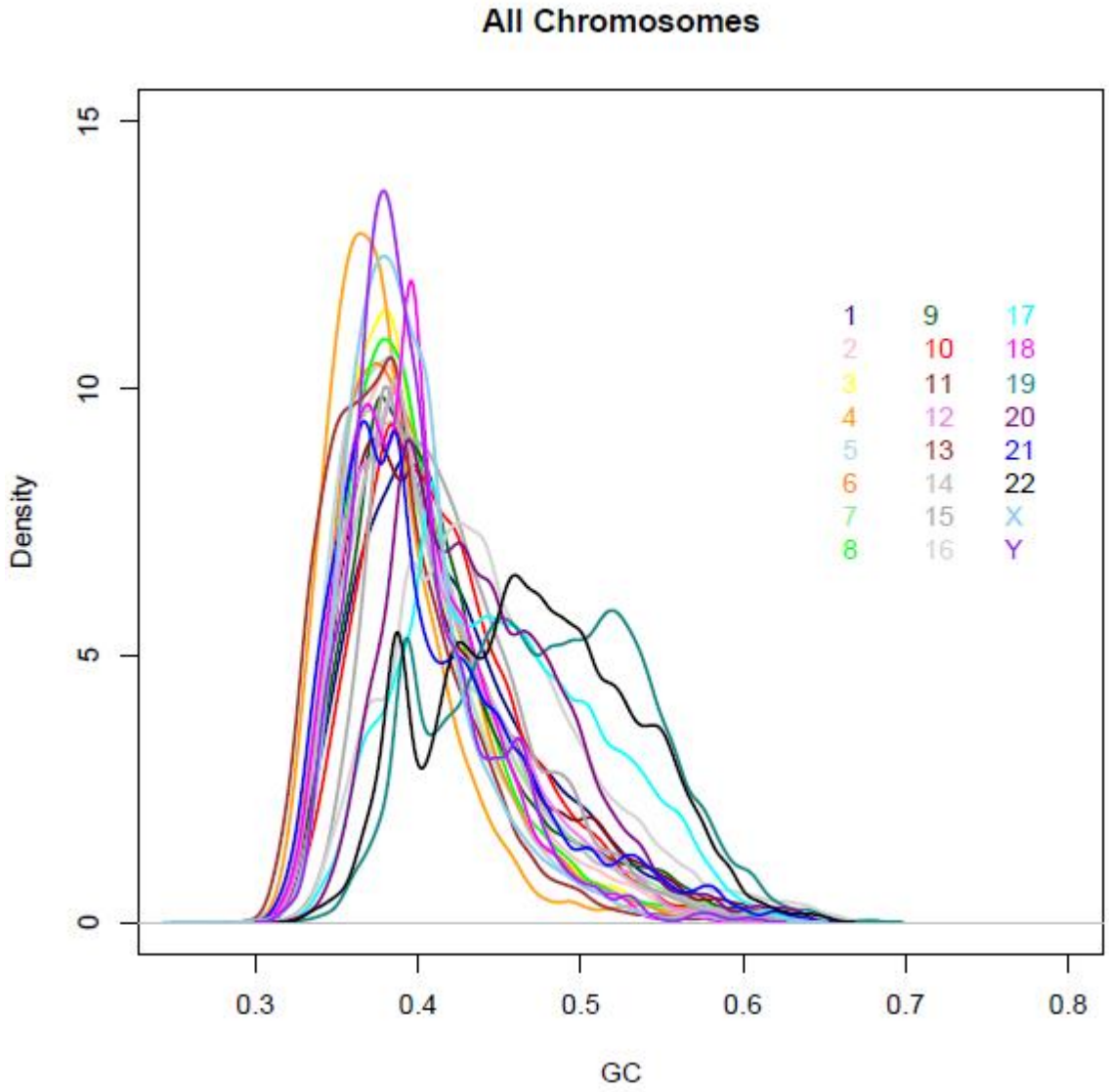

Chr 15
GC ventanas de 100 Kb
0.30
0.40
0.50
0.60
1 59 137 224 311 398 485 572 659 746 833 920 1017
Genes por ventana de 100 Kb
0
4
8
12


Chr 16
GC ventanas de 100 Kb
0.30
0.40
0.50
0.60
1 59 137 224 311 398 485 572 659 746 833
Genes por ventana de 100 Kb
0
4
8
12

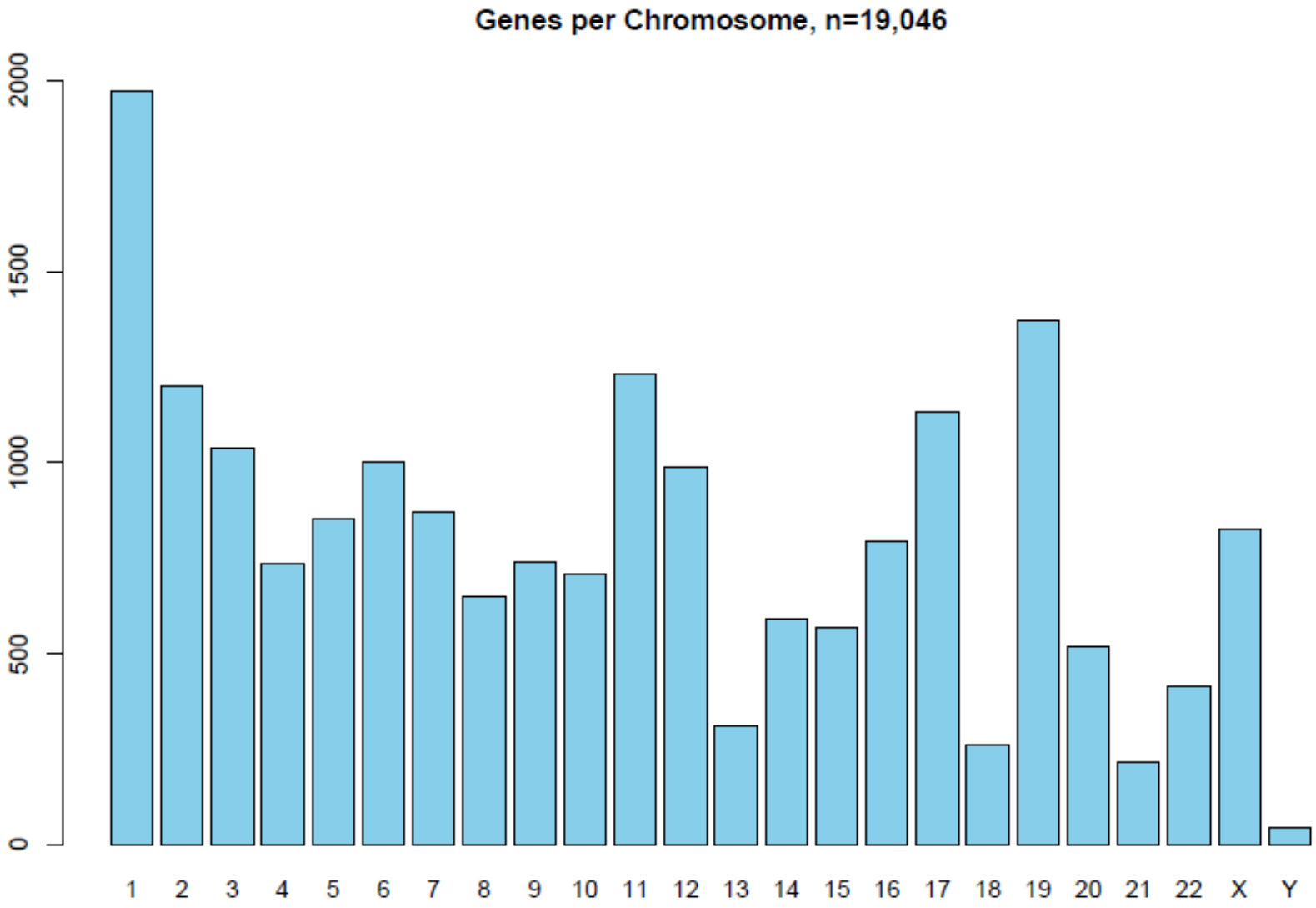

Chr 17
GC ventanas de 100 Kb
0.30
0.40
0.50
0.60
1 59 137 224 311 398 485 572 659 746 833
Genes por ventana de 100 Kb
0
4
8
12

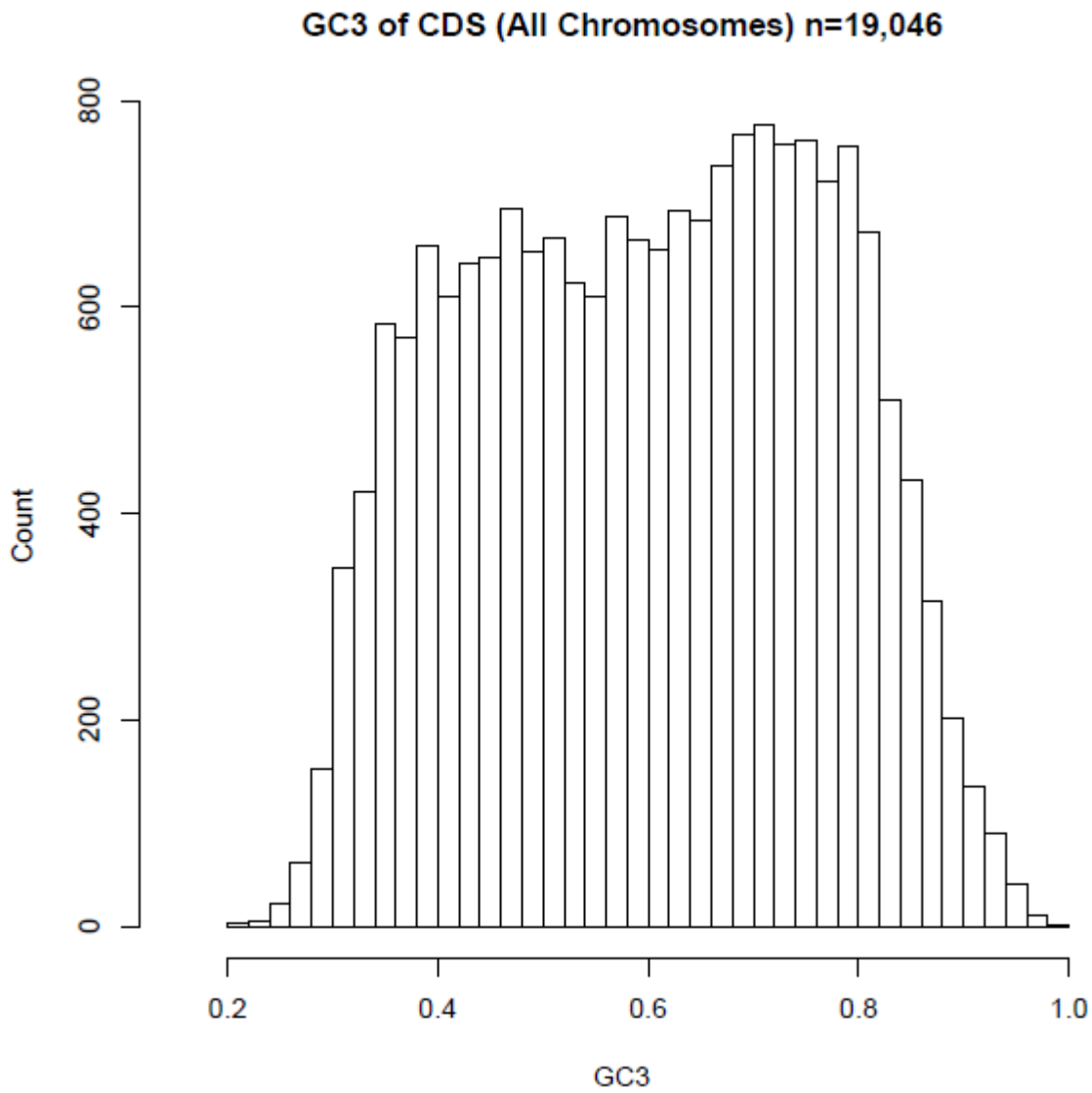

Chr 18
GC ventanas de 100 Kb
0.30
0.40
0.50
0.60
1 59 137 224 311 398 485 572 659 746
Genes por ventana de 100 Kb
0
4
8
12

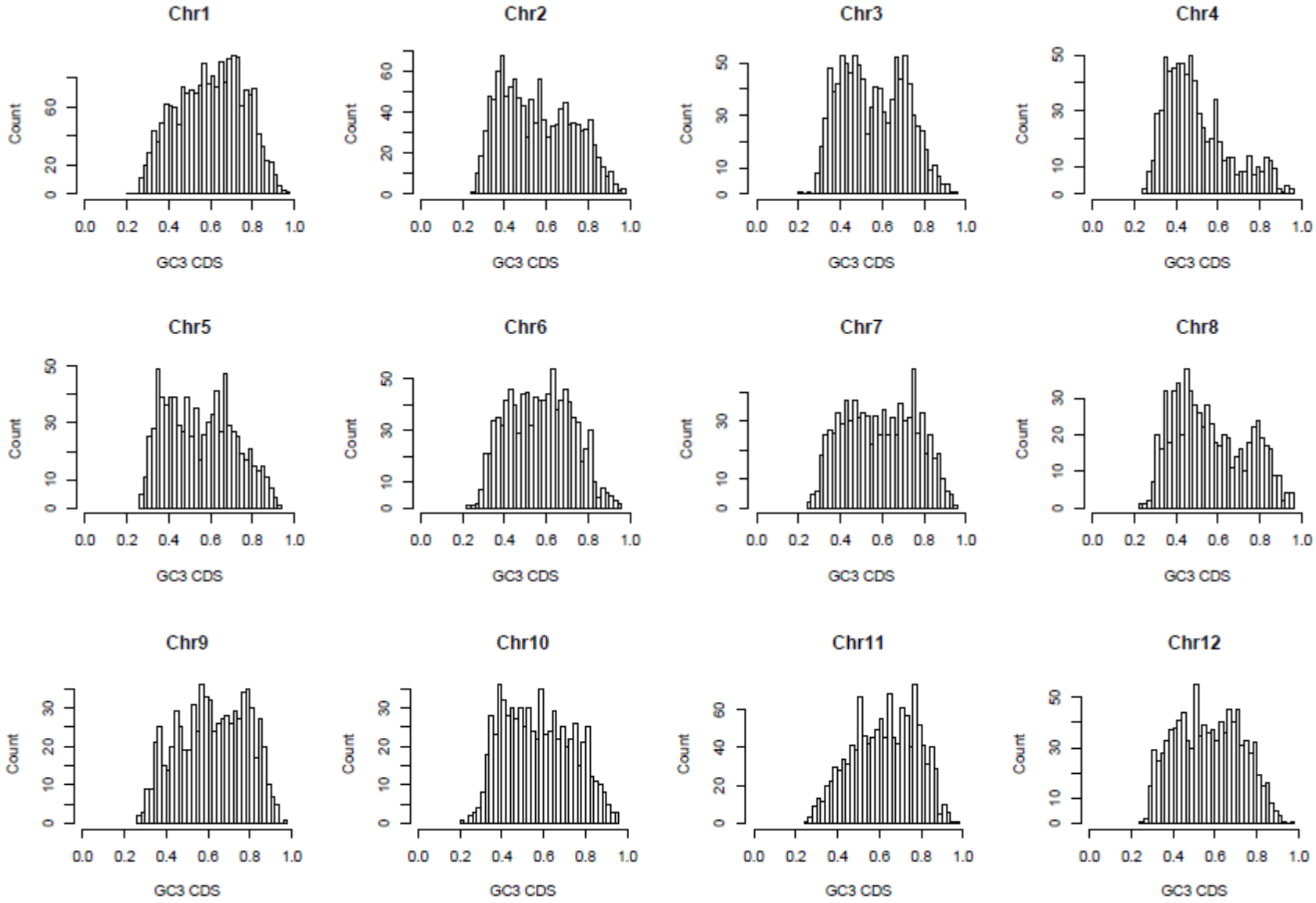

Human Genome: GC distibution
Density
0
2
4
6
8
0.3
0.4
0.5
0.6
0.7
GC. Window Size = 20 Kb

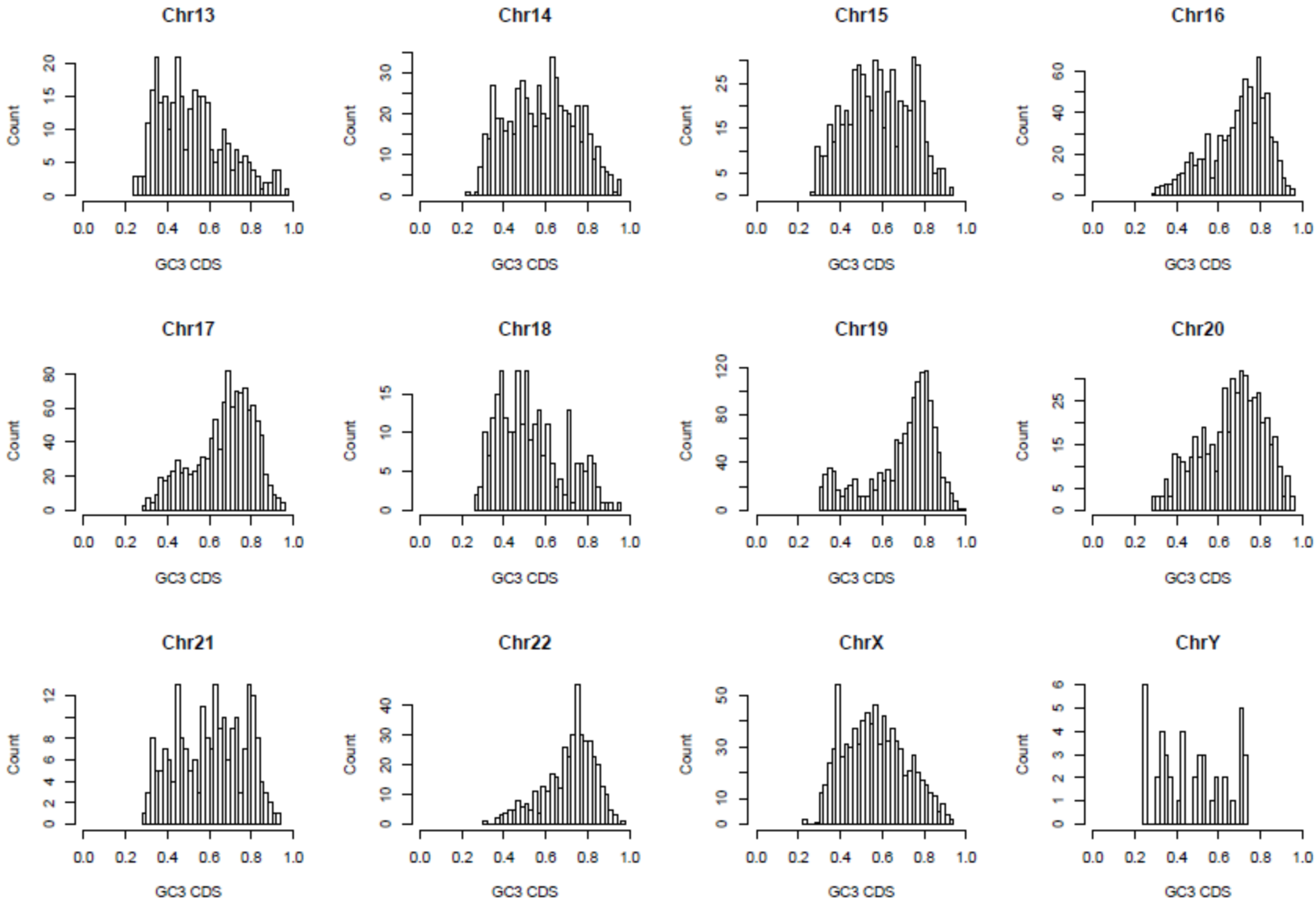

All Chromosomes
Density
0
5
10
15
0.3
0.4
0.5
0.6
0.7
0.8
GC
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

Genes por cromosoma

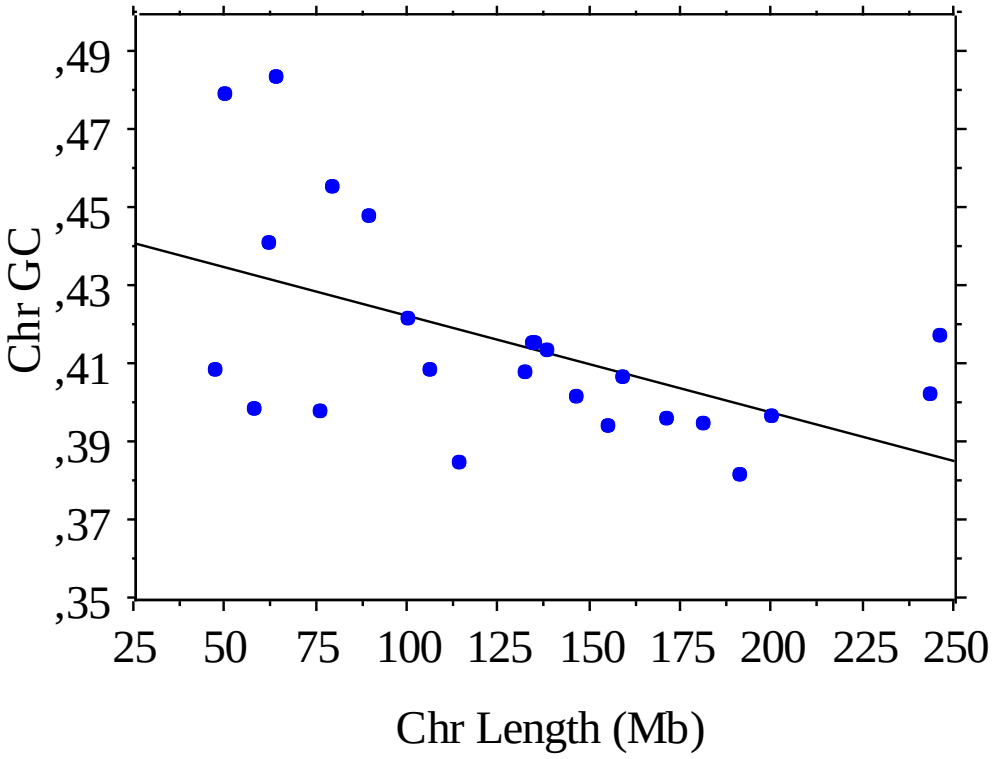



GC3 of CDS (All Chromosomes) n=19,046
Count
0
200
400
600
800
0.2
0.4
0.6
0.8
1.0
GC3


Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Count
GC3 CDS
0.0
0.2
0.4
0.6
0.8
1.0

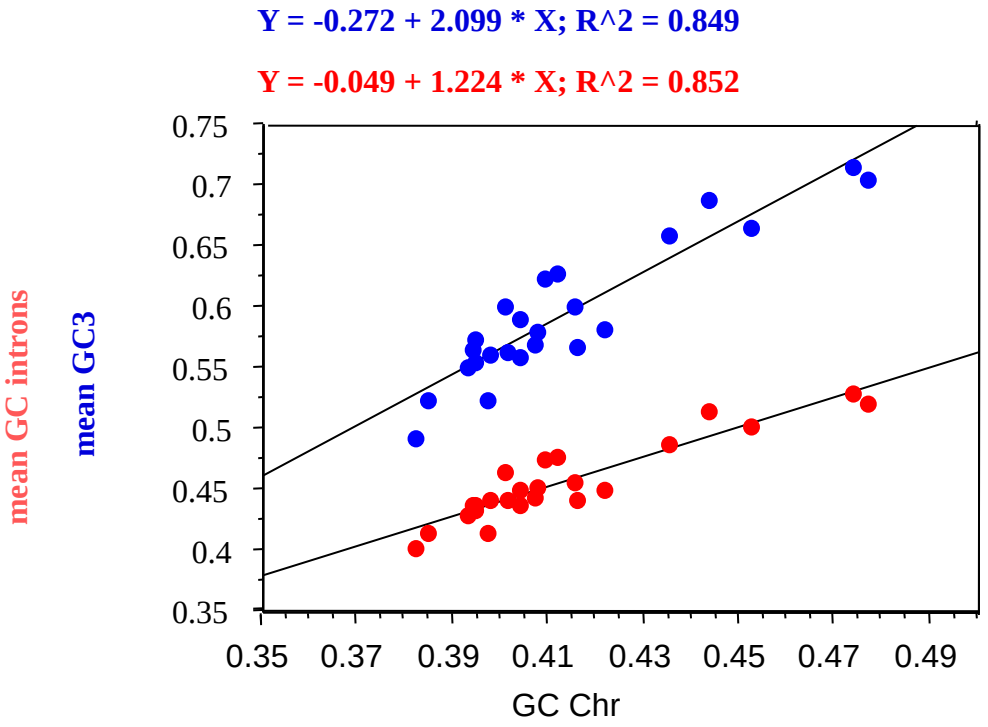

Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
ChrY
Count
GC3 CDS

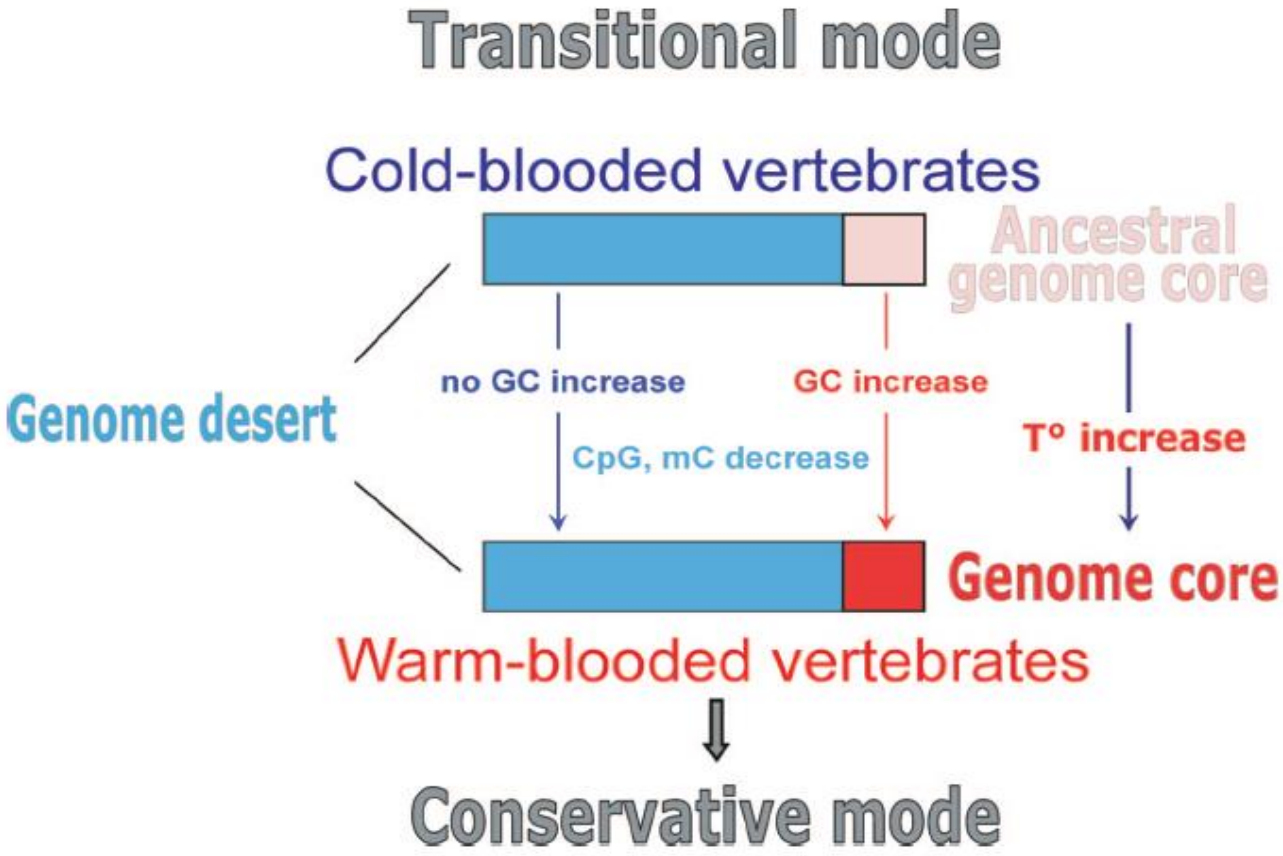
Correlación entre el contenido en GC de los cromosomas y la respectiva longitud

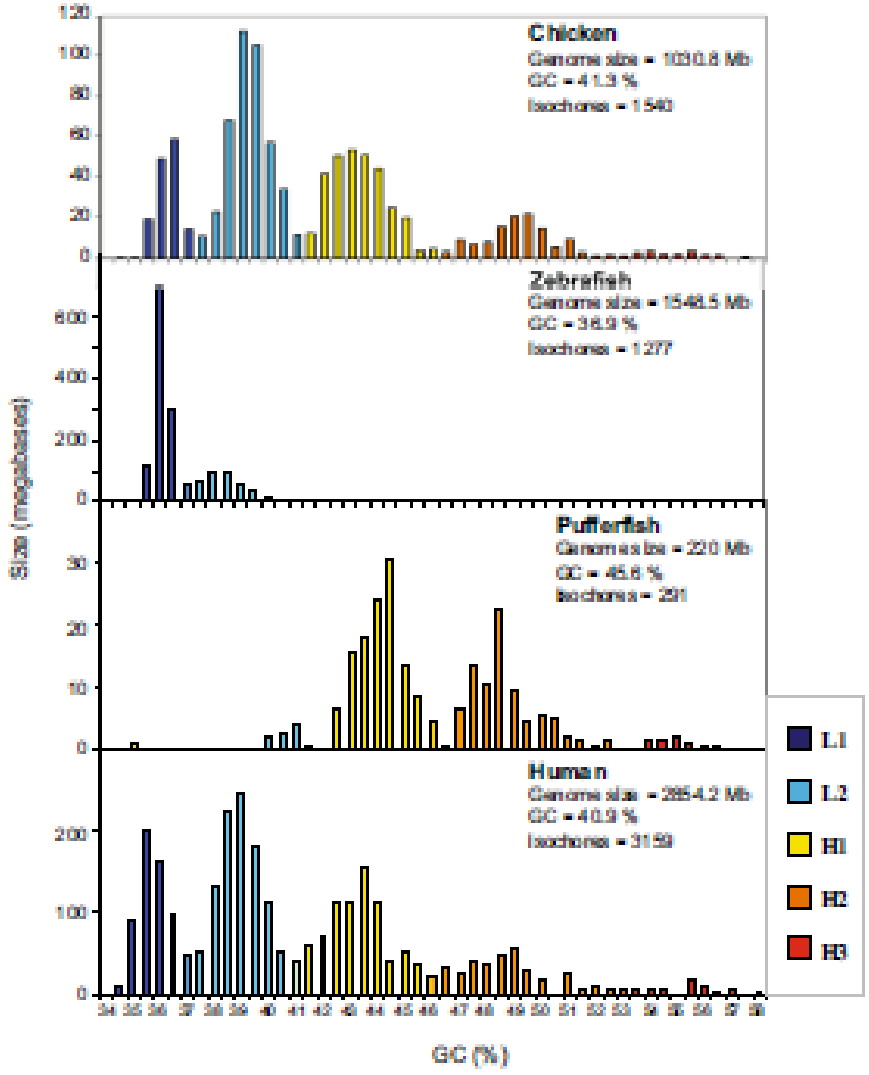


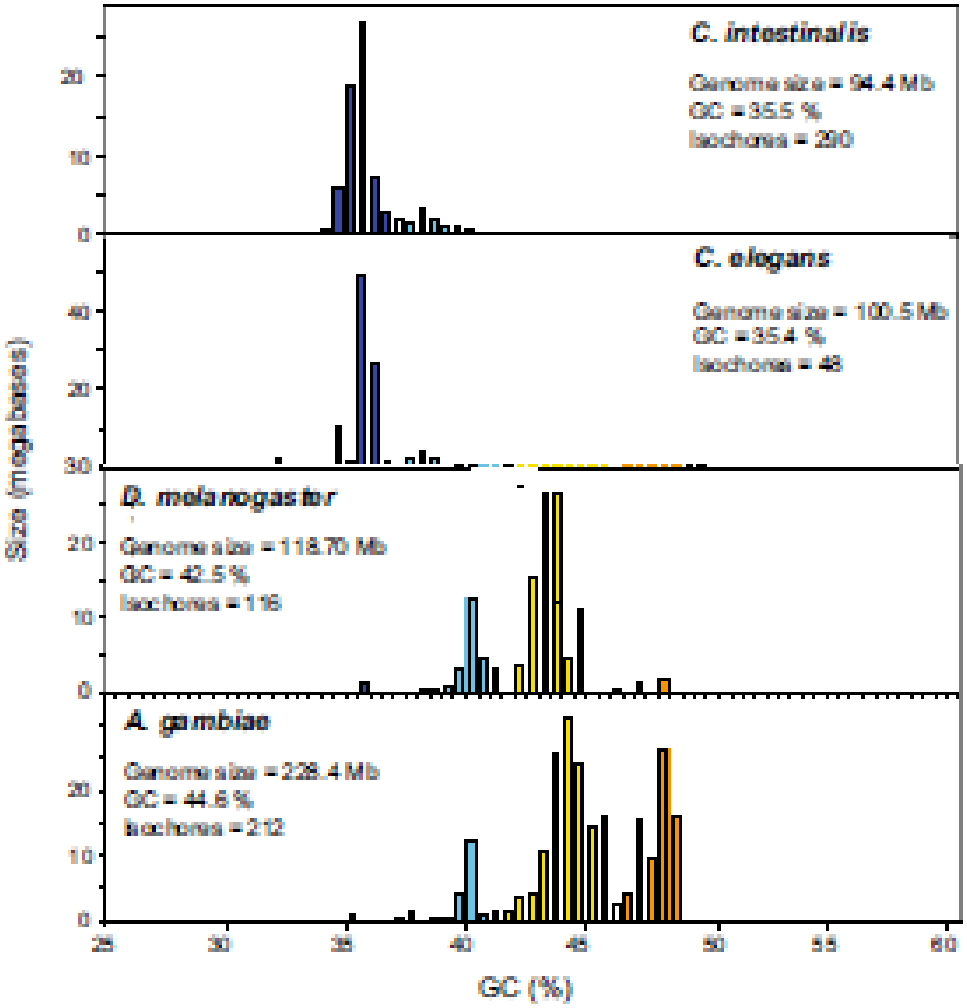
Y = ,447 - 2,462E-4 * X; R^2 = ,277; *P* = 0,008

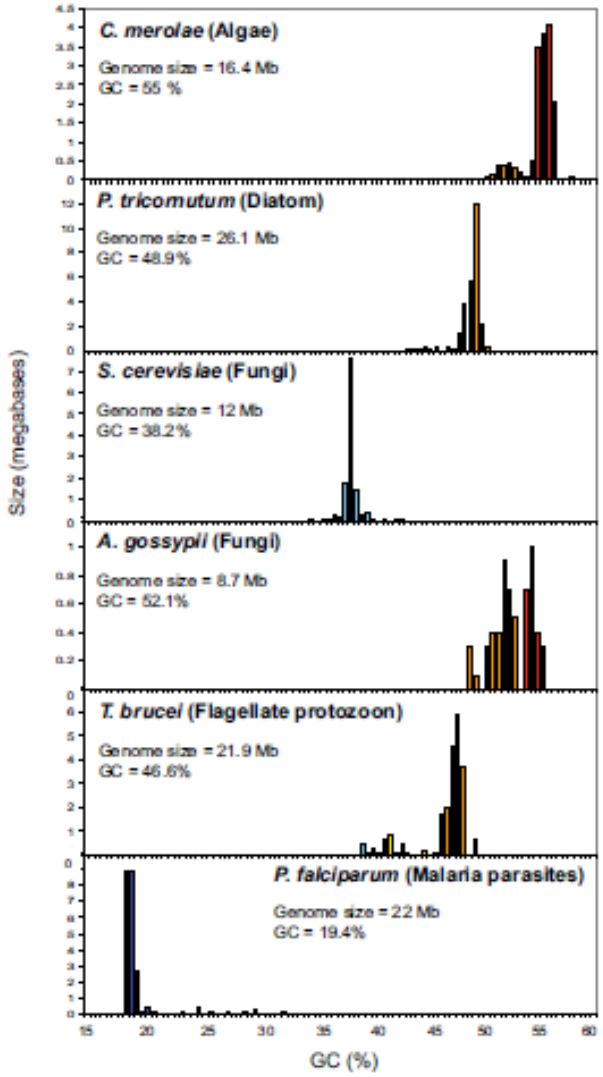
Correlaciones entre el GC de cada cromosoma y la media de GC3 y de los intrones




Transitional mode
Cold-blooded vertebrates
Ancestral genome core
no GC increase
GC increase
Genome desert
CpG, mC decrease
T° increase
Genome core
Warm-blooded vertebrates
Conservative mode


Chicken
Genome size = 1030.8 Mb
GC = 41.3 %
Isochores = 1540
Zebrafish
Genome size = 1548.5 Mb
GC = 36.9 %
Isochores = 1277
Pufferfish
Genome size = 220 Mb
GC = 45.6 %
Isochores = 291
Human
Genome size = 2854.2 Mb
GC = 40.9 %
Isochores = 3159
L1
L2
H1
H2
H3
Size (megabases)
GC (%)


C. intestinalis
Genome size = 94.4 Mb
GC = 35.5 %
Isochores = 290
C. elegans
Genome size = 100.5 Mb
GC = 35.4 %
Isochores = 48
D. melanogaster
Genome size = 118.70 Mb
GC = 42.5 %
Isochores = 116
A. gambiae
Genome size = 228.4 Mb
GC = 44.6 %
Isochores = 212
Size (megabases)
GC (%)


C. merolae (Algae)
Genome size = 16.4 Mb
GC = 55 %
P. tricornutum (Diatom)
Genome size = 26.1 Mb
GC = 48.9 %
S. cerevisiae (Fungi)
Genome size = 12 Mb
GC = 38.2 %
A. gossypii (Fungi)
Genome size = 8.7 Mb
GC = 52.1%
T. brucei (Flagellate protozoon)
Genome size = 21.9 Mb
GC = 46.6%
P. falciparum (Malaria parasites)
Genome size = 2.2 Mb
GC = 19.4%
Size (megabases)
GC (%)