



Comparative Microbial Genomics - Exercises

Dave Ussery
Biological Sequence Analysis
DTU course # 27803
15 April, 2005



Comparison Exercises

1. Simple comparisons

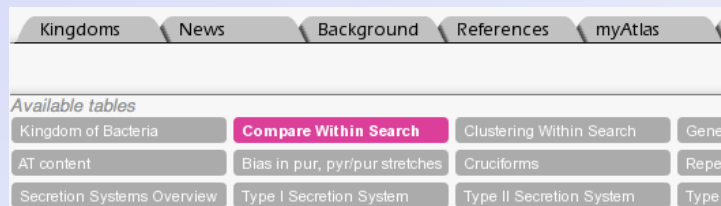
Length vs. AT content

AT content vs. Repeats

go to <http://www.cbs.dtu.dk/services/GenomeAtlas/>

click on "Bacteria"

click on "Compare within search"



[CBS](#) >> [CBS Prediction Servers](#) >> [GenomeAtlas](#) >> Overrep. Oligomers for strain ,



CBS GENOME ATLAS DATABASE

[Kingdoms](#)
[News](#)
[Background](#)
[References](#)
[myAtlas](#)
[Download](#)
[login](#)

Bacteria: 244 segments in 226 strains in 168 species in 105 genera selected

Available tables

[Kingdom of Bacteria](#)
[Compare Within Search](#)
[Clustering Within Search](#)
[Gene Prediction](#)
[Periodicity](#)
[Sigma Factors](#)
[AT content](#)
[Bias in pur, pyr/pur stretches](#)
[Cruciforms](#)
[Repeats](#)
[Protein Length](#)
[Origin of Replication](#)
[Secretion Systems Overview](#)
[Type I Secretion System](#)
[Type II Secretion System](#)
[Type III Secretion System](#)
[Type IV Secretion System](#)
[Type V Secretion System](#)

Compare Segments within your current search

The data comparison tool makes use of the CBS Genome Atlas Database. Please cite the references below if you are including these results in your scientific work

Graph title

Data for X axis LOG Transformation

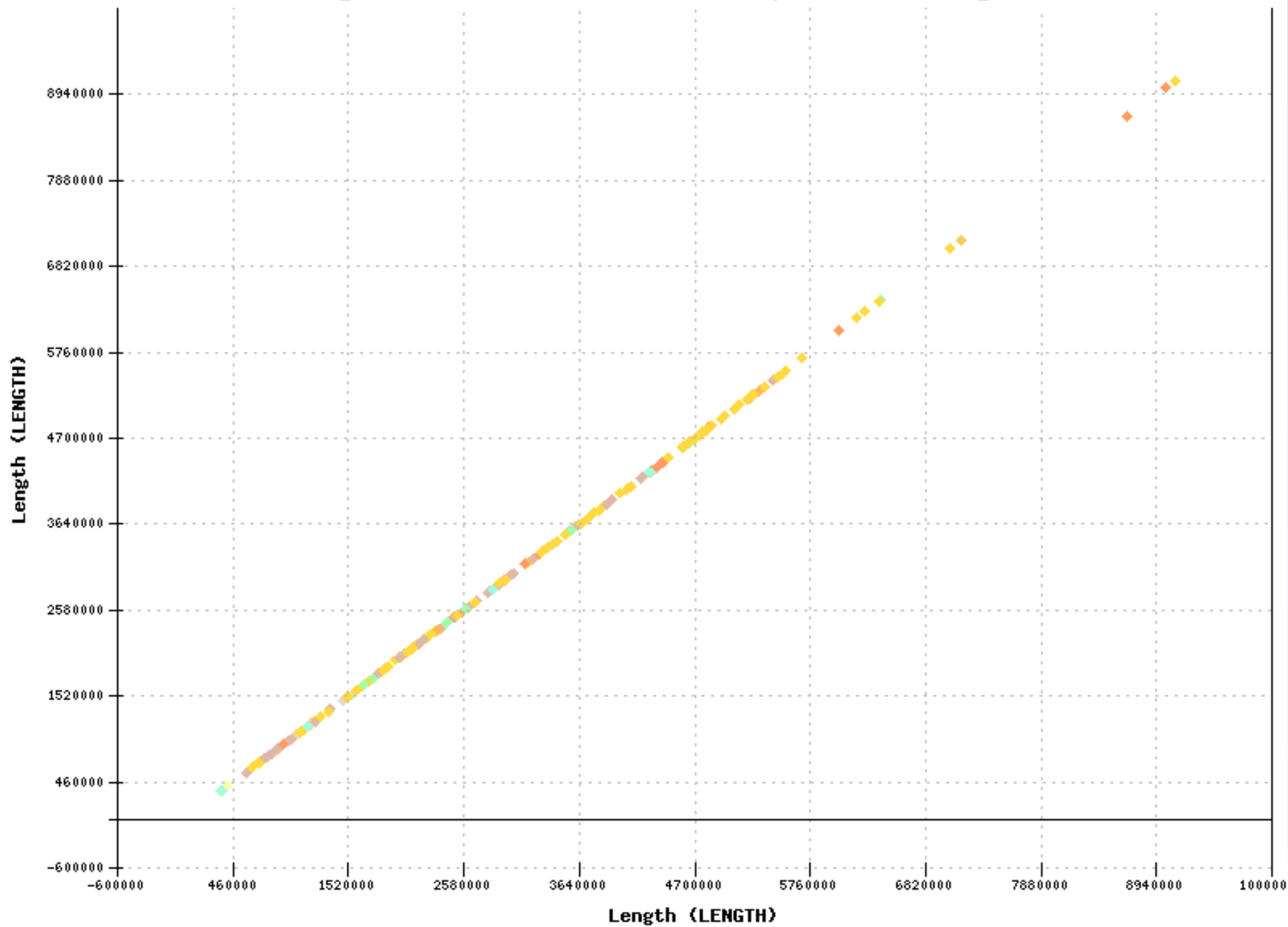
X axis title

Data for X axis LOG Transformation

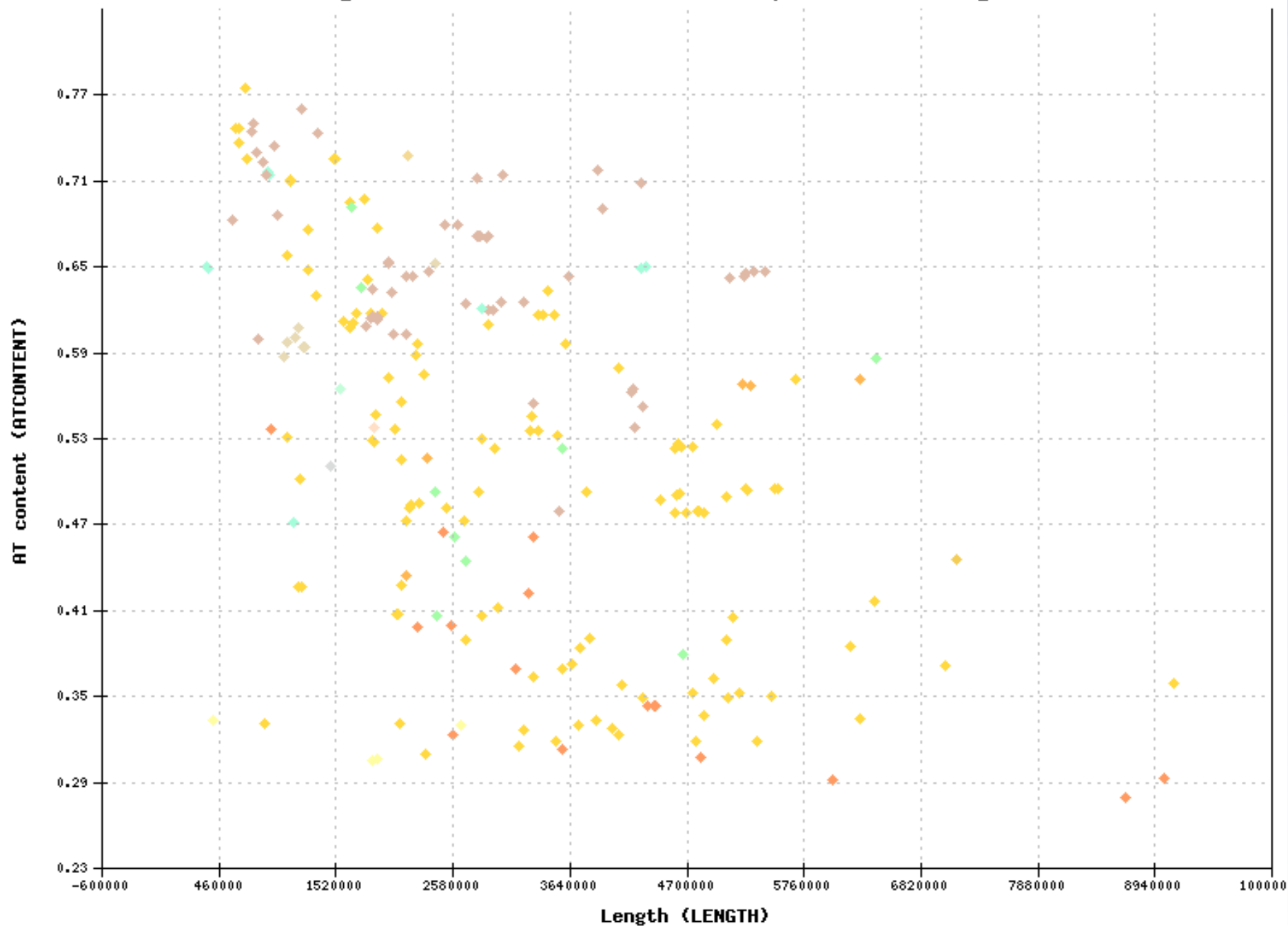
Y axis title

[Generate Scatter Plot](#)

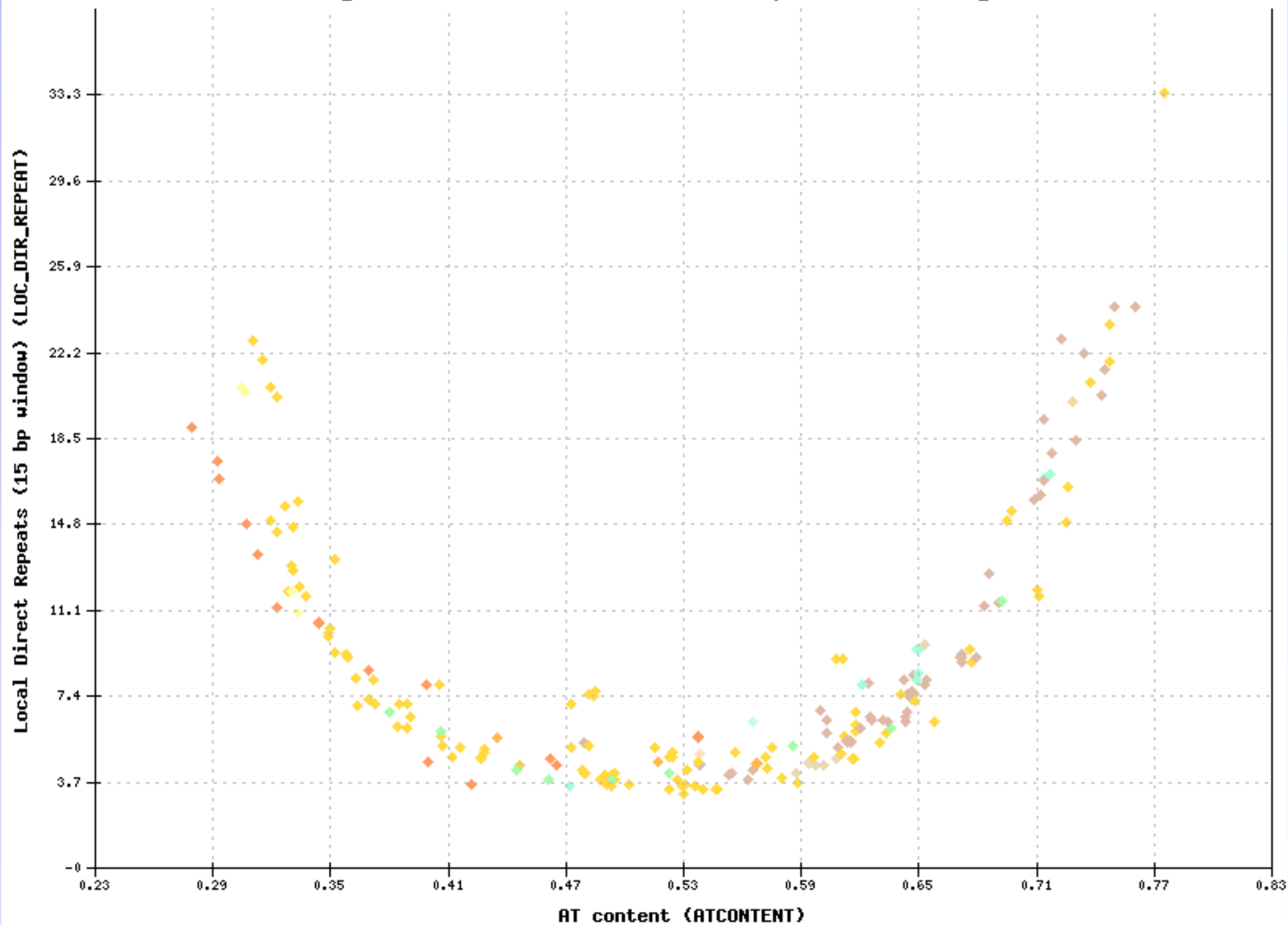
Bacteria : 243 segments in 225 strains in 168 species in 105 genera selected

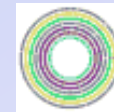


Bacteria : 243 segments in 225 strains in 168 species in 105 genera selected



Bacteria : 243 segments in 225 strains in 168 species in 105 genera selected





Comparison Exercises

<http://www.cbs.dtu.dk/services/GenomeAtlas/>

1. Simple comparisons

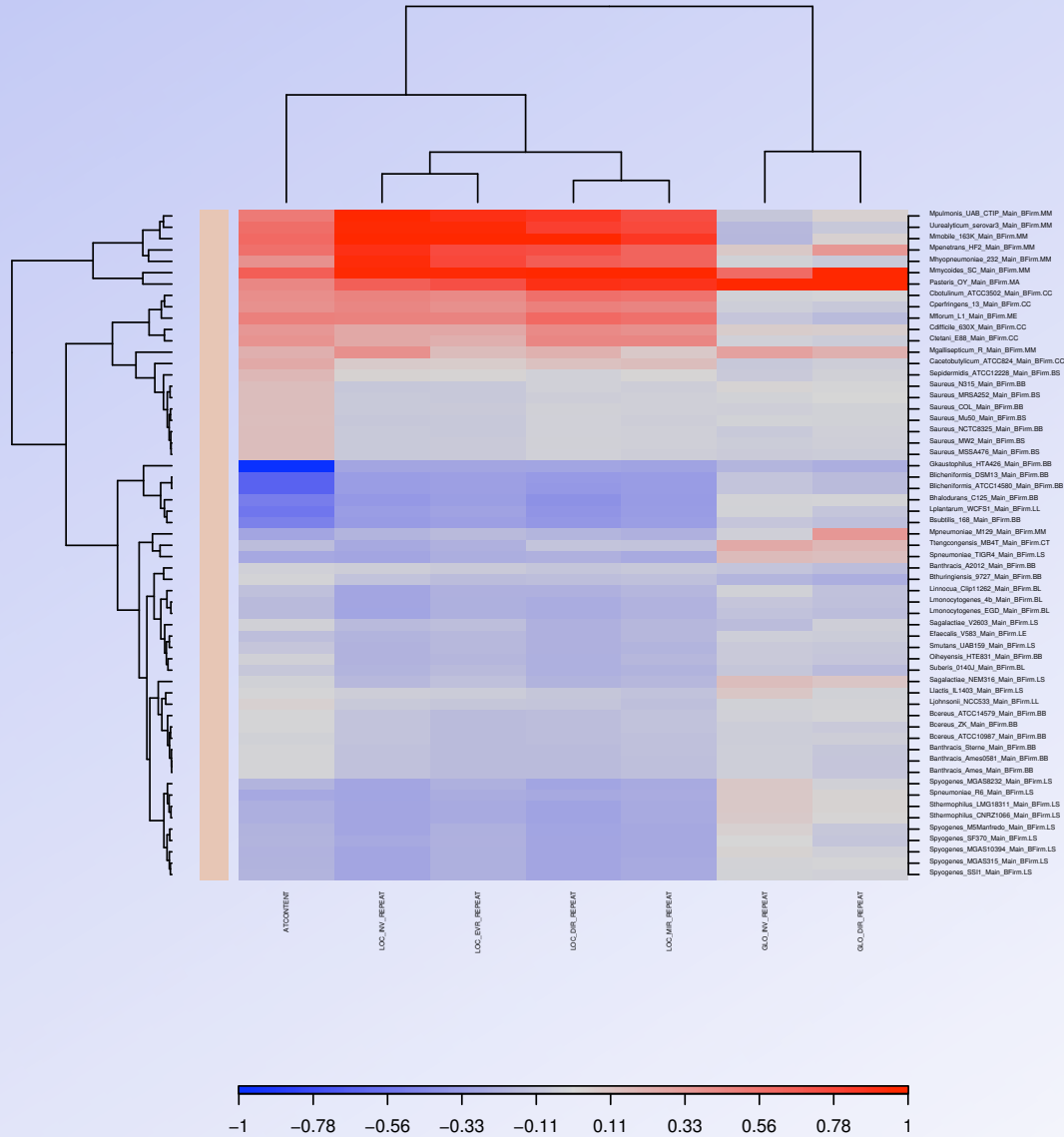
Length vs. AT content

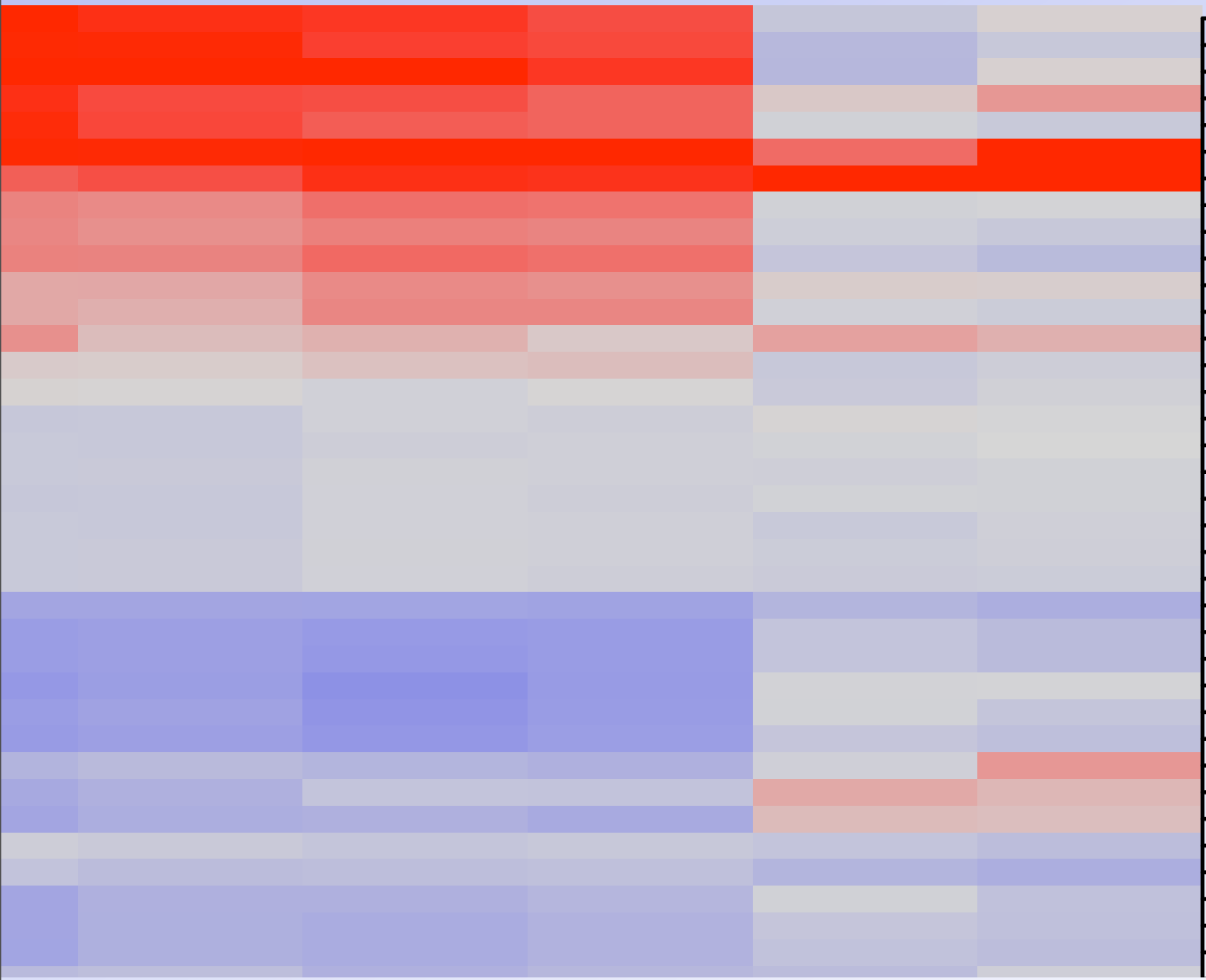
AT content vs. Repeats

2. 2-D Clustering Analysis



Two-way cluster analysis within your current search





- Mpulsimonis_UAB_CTIP_Main_BFirm.MM
- Uurealyticum_serovar3_Main_BFirm.MM
- Mmobile_163K_Main_BFirm.MM
- Mpenetrans_HF2_Main_BFirm.MM
- Mhyopneumoniae_232_Main_BFirm.MM
- Mmycoides_SC_Main_BFirm.MM
- Pasteris_OY_Main_BFirm.MA
- Cbotulinum_ATCC3502_Main_BFirm.CC
- Cperfringens_13_Main_BFirm.CC
- Mflorum_L1_Main_BFirm.ME
- Cdifficile_630X_Main_BFirm.CC
- Ctetani_E88_Main_BFirm.CC
- Mgallisepticum_R_Main_BFirm.MM
- Cacetobutylicum_ATCC824_Main_BFirm.CC
- Sepidermidis_ATCC12228_Main_BFirm.BS
- Saureus_N315_Main_BFirm.BB
- Saureus_MRSA252_Main_BFirm.BS
- Saureus_COL_Main_BFirm.BB
- Saureus_Mu50_Main_BFirm.BS
- Saureus_NCTC8325_Main_BFirm.BB
- Saureus_MW2_Main_BFirm.BS
- Saureus_MSSA476_Main_BFirm.BS
- Gkaustophilus_HTA426_Main_BFirm.BB
- Blicheniformis_DSM13_Main_BFirm.BB
- Blicheniformis_ATCC14580_Main_BFirm.BB
- Bhalodurans_C125_Main_BFirm.BB
- Lplantarum_WCFS1_Main_BFirm.LL
- Bsubtilis_168_Main_BFirm.BB
- Mpneumoniae_M129_Main_BFirm.MM
- Ttengcongensis_MB4T_Main_BFirm.CT
- Spneumoniae_TIGR4_Main_BFirm.LS
- Banthracis_A2012_Main_BFirm.BB
- Bthuringiensis_9727_Main_BFirm.BB
- Linnocua_Clip11262_Main_BFirm.BL
- Lmonocytogenes_4b_Main_BFirm.BL
- Lmonocytogenes_EGD_Main_BFirm.BL
- Sanalactiae_V2603_Main_BFirm.IS



Summary

1. Local repeats are less likely to occur by chance for genomes with around 50% AT content
2. Clustering analysis can be used to compare many genomic properties.