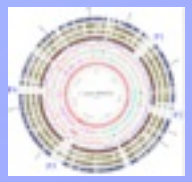


CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

Methods to Compare Genomes

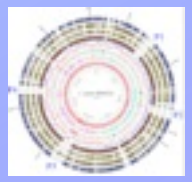
with special emphasis on AT content

Dave Ussery

MBV-INF4410
Bioinformatics for Molecular Biology
First talk

Friday, 11 September, 2009

DTU



_computational BIOLOGY

PERSPECTIVE

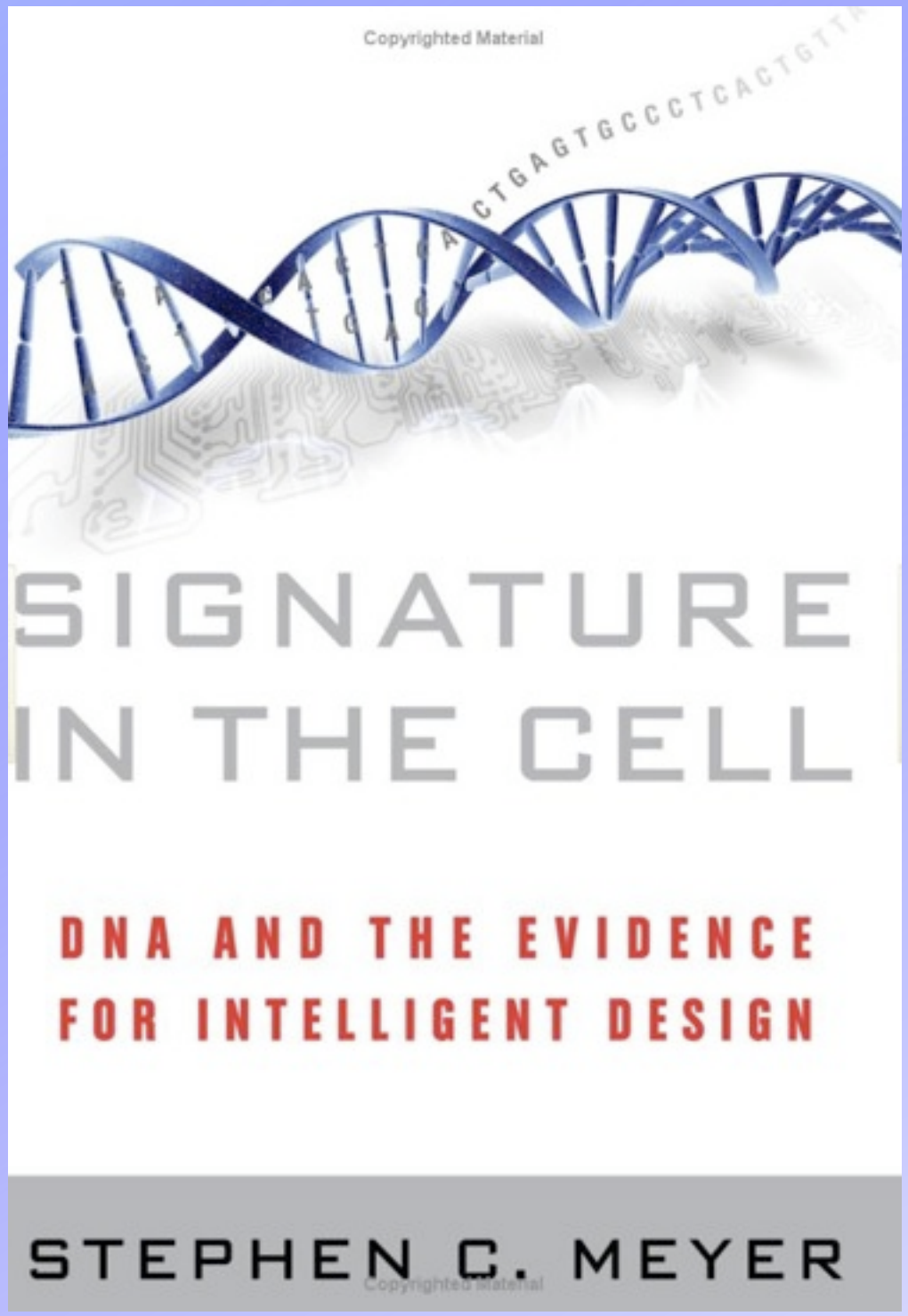
Fifteen years of microbial genomics: meeting the challenges and fulfilling the dream

Nikos C Kyrpides

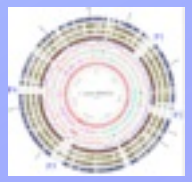
Nikos C Kyrpides

As we approach the completed sequencing of 1,000 microbial genomes, the field of microbial genomics is poised at a crossroads. The future holds great promise for far-reaching advancements in microbiology as well as in diverse, related sciences. But realizing that potential will require meeting the challenges that have accompanied the rapid development of the underlying technology and the exponential growth of data. New technologies provide unprecedented opportunities but also call for conceptual shifts. Experience gained in the first decade of genomics can guide the improved approaches now needed for the selection of genome sequencing projects and their funding, for genome publication and annotation, as well as for data analysis and access. Equipped with these new tools and policies, microbiologists will have a unique opportunity for unprecedented exploration of our microbial planet.

Table 1 Estimating the magnitude of microbial diversity	
Number of bacteriophages on Earth	10^{31}
Number of microbes on Earth	5×10^{30}
Number of stars in the universe	7×10^{21}
Number of microbes in all humans	6×10^{23}
Number of humans	6×10^{9}
Number of microbial cells in one human gut	10^{14}
Number of human cells in one human	10^{13}
Number of microbial genes in one human gut	3×10^{6}
Number of genes in the human genome	2.5×10^{4}
Combined length of all bacteriophages on Earth	10^{8} Ly
Diameter of the Milky Way	10^{5} Ly



"An inordinate fondness of bacteriophages..."



What have we learned from ~1000 sequenced bacterial genomes?

1. Genetic diversity is greater than we thought.

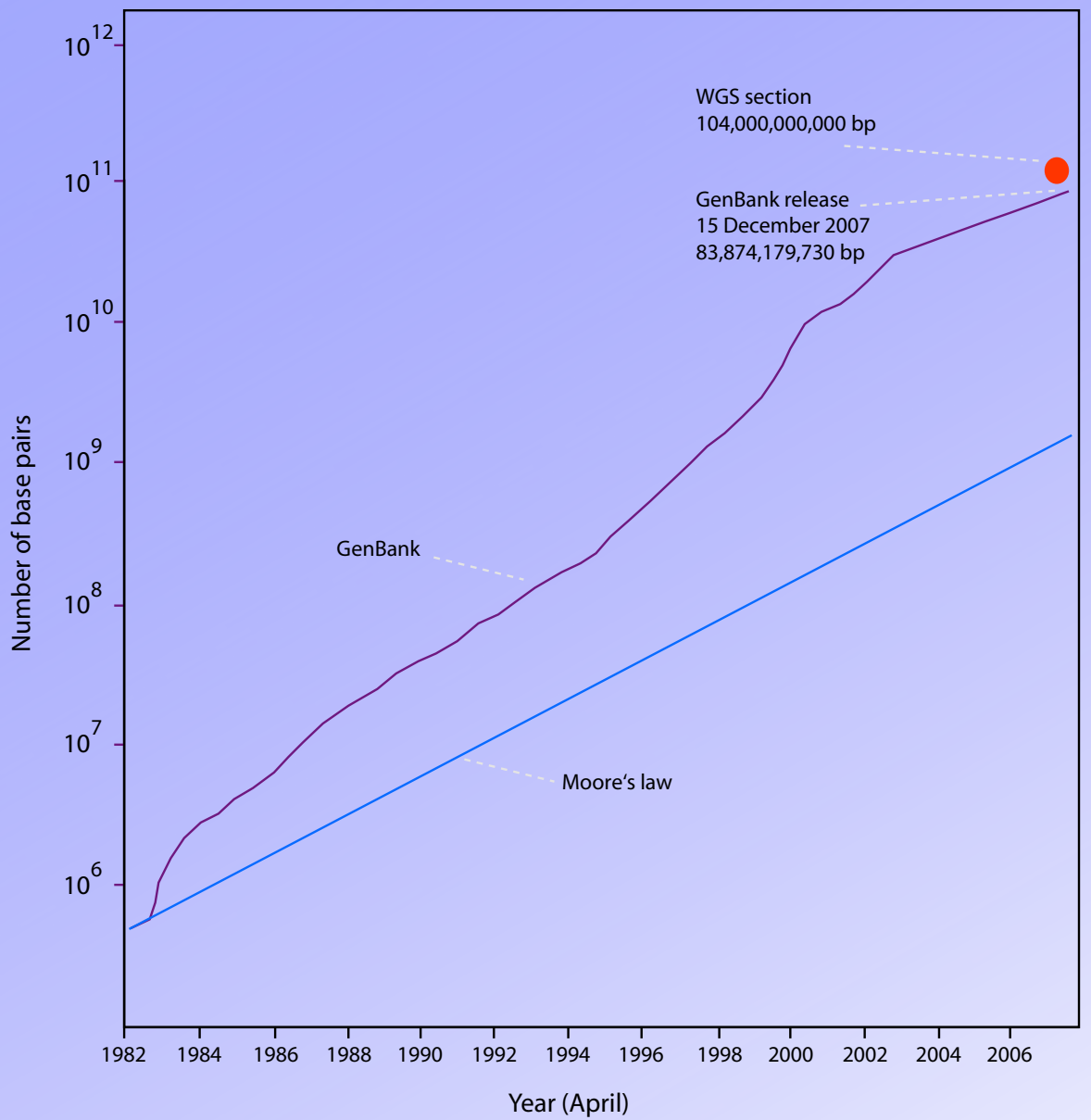
2. Very large genomes tend to be GC-rich, whilst small genomes tend to be AT-rich.

3. There are patterns, both locally and globally of AT-richness in bacterial chromosomes.

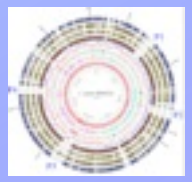
4. Generally G's are biased towards the leading strand, but the strand bias of A's is tax-specific.

Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark

The problem....



D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009



20 Methods to Compare Bacterial Genomes

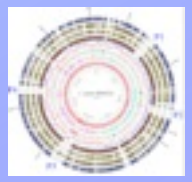
Genome

1. Length
2. Number genes [coding density]
3. AT content
4. Oligomer skews
5. Chromosome alignment
6. Repeats
7. Periodicity

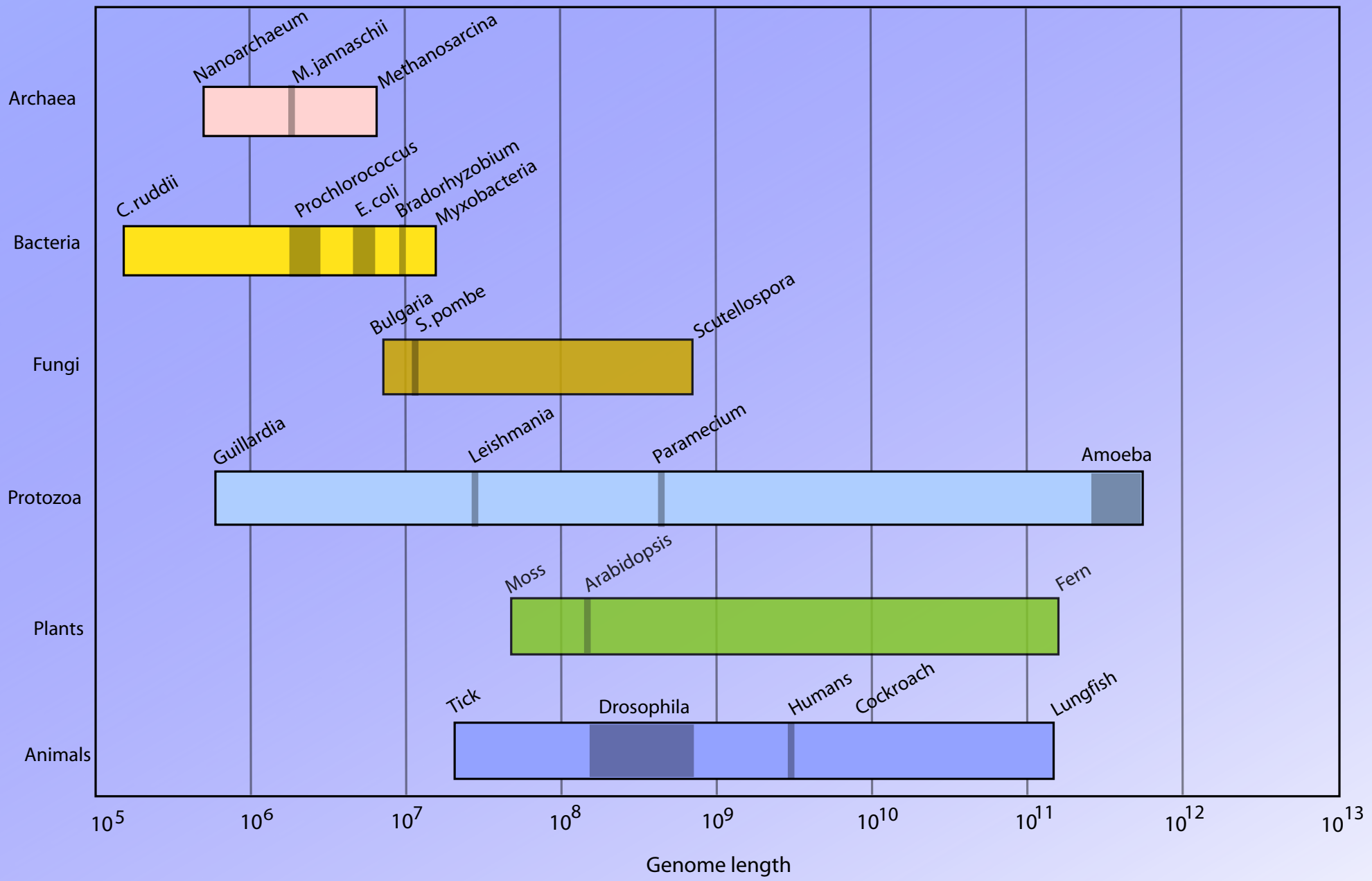
Coding

8. tRNAs and codon usage
9. Bias in codon usage
10. Amino acid usage

Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark

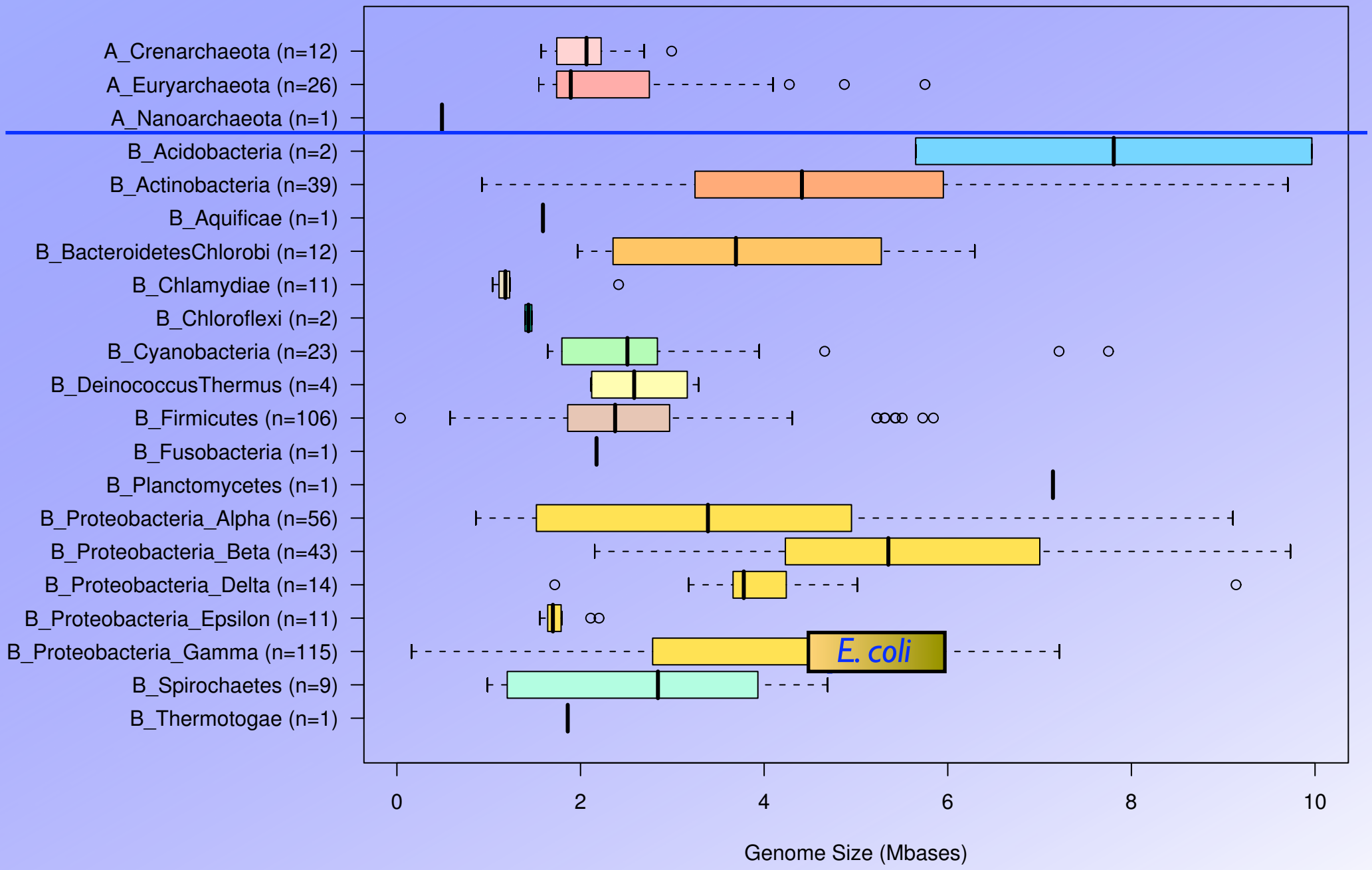


11. Promoters
12. Annotation quality
13. Blast atlases
14. Proteome comparisons
15. 2-D correlation of properties
16. Sigma Factors
17. Two-component systems
18. Transcription Factors
19. Membrane Proteins
20. Secreted Proteins



D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009

1. Length



D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009

1. Length

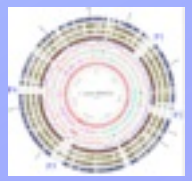
Peter

Available Tables

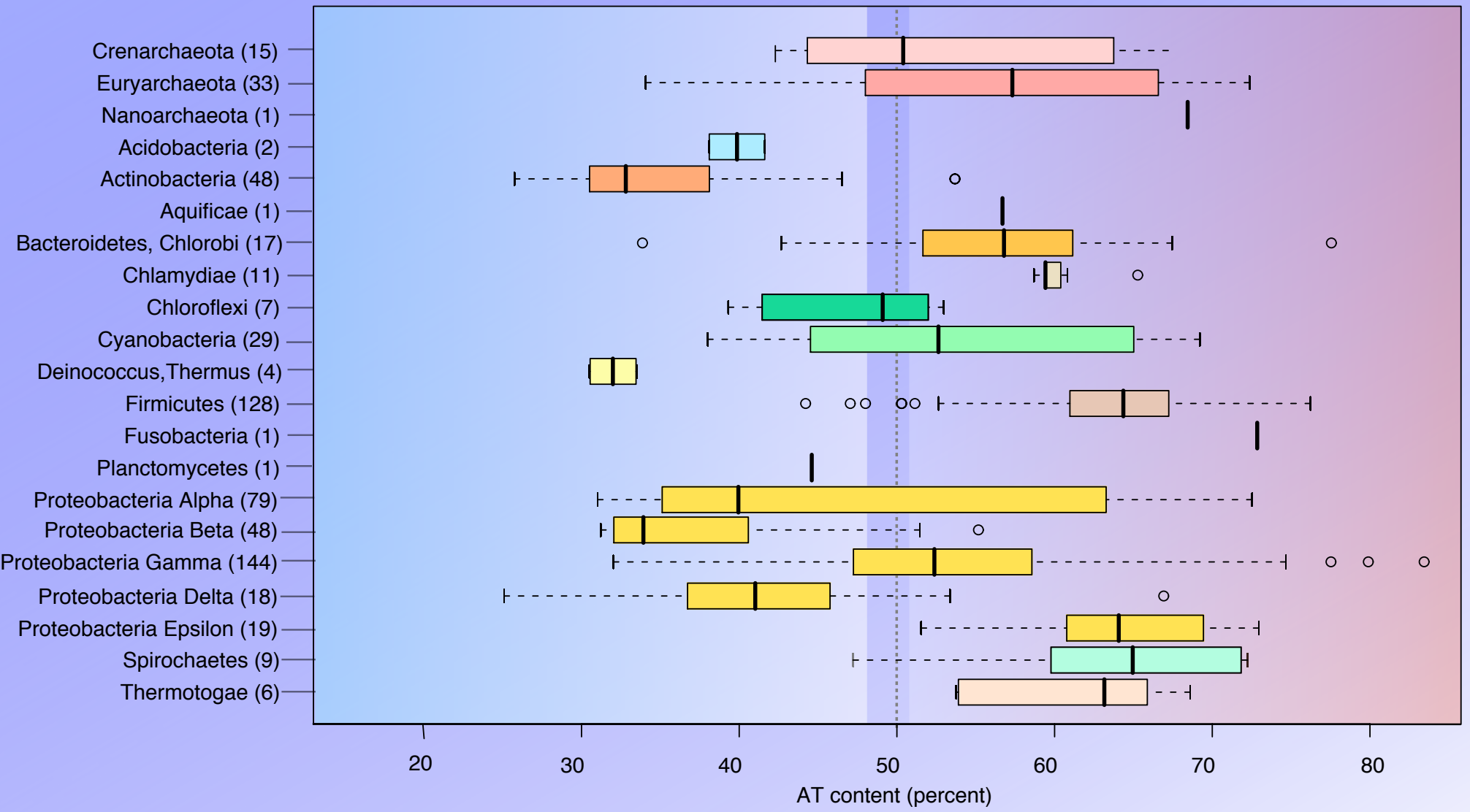
General | Sigma Factors | Two-Component systems | Repeats | Protein length | A/Z-DNA

Row	Organism	Tax Group	NCBI Project ID	Replicons	Total Size (bp)	Number of genes	5S rRNA count	16S rRNA count	23S rRNA count		
1	Sorangium cellulosum So ce 56	BProt DM	28111	1	13,033,779	9,384	4	4	4	60	28.6
2	Rhodococcus jostii RHA1	BActin AA	13693	4	9,702,737	9,145	4	4	4	52	33.0
3	Burkholderia xenovorans LB400	BProt BB	254	3	9,731,138	8,702	6	6	6	64	37.4
4	Acaryochloris marina MBIC11017	BCyano A	12997	10	8,361,599	8,383	2	2	2	69	53.0
5	Bradyrhizobium japonicum USDA 110	BProt ARh	17	1	9,105,828	8,317	1	1	1	51	35.9
6	Methylobacterium nodulans ORS 2060	BProt ARh	20477	8	8,839,022	8,309	7	7	7	73	31.6
7	Streptomyces coelicolor A3(2)	BActin AA	242	3	9,054,847	8,215	6	6	6	65	28.0
8	Candidatus Solibacter usitatus Ellin6076	BASS	12638	1	9,965,640	7,826	2	2	2	52	38.1
9	Burkholderia sp. 383	BProt BB	10695	3	8,676,277	7,717	6	6	6	67	33.7
10	Streptomyces avermitilis MA-4680	BActin AA	189	2	9,119,895	7,673	6	6	6	68	29.3
11	Bradyrhizobium sp. BTAi1	BProt ARh	16137	2	8,493,513	7,622	2	2	2	52	35.2
12	Burkholderia vietnamiensis G4	BProt BB	10696	8	8,391,070	7,617	6	6	6	67	34.3
13	Burkholderia phymatum STM815	BProt BB	17409	4	8,676,562	7,496	6	6	6	62	37.7
14	Myxococcus xanthus DK 1622	BProt DM	1421	1	9,139,763	7,331	4	4	4	65	31.1
15	Rhodopirellula baltica SH 1	BPPP	413	1	7,145,576	7,325	1	1	1	76	44.6
16	Rhizobium leguminosarum bv. viciae 3841	BProt ARh	344	7	7,751,309	7,285	3	3	3	52	39.1
17	Mesorhizobium loti MAFF303099	BProt ARh	18	3	7,596,297	7,281	2	2	2	54	37.5
18	Rhodococcus opacus B4	BActin AA	34839	1	7,913,450	7,252	4	4	4	49	32.1
19	Burkholderia phytofirmans PsJN	BProt BB	17463	3	8,214,658	7,241	6	6	6	63	37.7
20	Burkholderia pseudomallei 668	BProt BB	13953	2	7,040,403	7,230	4	4	4	59	31.7
21	Saccharopolyspora erythraea NRRL 2338	BActin AA	18489	1	8,212,805	7,198	4	4	4	50	28.9
22	Frankia sp. EAN1pec	BActin AA	13915	1	8,982,042	7,191	3	3	3	47	28.9
23	Burkholderia pseudomallei 1106a	BProt BB	16182	2	7,089,249	7,183	4	4	4	59	31.7
24	Streptomyces griseus subsp. griseus NBRC 13350	BActin AA	20085	1	8,545,929	7,138	6	6	6	66	27.8
25	Burkholderia cenocepacia MC0-3	BProt BB	17929	3	7,971,389	7,008	6	6	6	67	33.4
26	Rhizobium leguminosarum bv. trifolii WSM1325	BProt ARh	20097	6	7,418,122	7,001	3	3	3	51	39.2
27	Burkholderia cenocepacia HI2424	BProt BB	13918	4	7,702,840	6,919	6	6	6	67	33.2
28	Hahella chejuensis KCTC 2396	BProt GO	16064	1	7,215,267	6,782	5	5	5	67	46.1
29	Bradyrhizobium sp. ORS278	BProt ARh	19575	1	7,456,587	6,752	2	2	2	50	34.5
30	Frankia alni ACN14a	BActin AA	17403	1	7,497,934	6,723	2	2	2	46	27.2
31	Mycobacterium smegmatis str. MC2 155	BActin AA	92	1	6,988,209	6,716	2	2	2	47	32.6
32	Burkholderia ambifaria MC40-6	BProt BB	17411	4	7,642,536	6,697	6	6	6	68	33.6
33	Methylobacterium sp. 4-46	BProt ARh	18809	3	7,737,025	6,692	6	6	6	63	28.5
34	Nostoc punctiforme PCC 73102	BCyano NN	216	6	9,059,191	6,690	4	4	4	77	58.7
35	Agrobacterium radiobacter K84	BProt ARh	13402	5	7,273,300	6,684	3	3	3	52	40.1
36	Ralstonia eutropha H16	BProt BB	13603	3	7,416,678	6,626	5	5	5	61	33.7
37	Burkholderia ambifaria AMMD	BProt BB	13490	4	7,528,567	6,617	6	6	6	69	33.2
38	Pseudomonas fluorescens SBW25	BProt GPs	31229	2	7,147,633	6,487	6	5	5	66	39.9
39	Burkholderia cenocepacia J2315	BProt BB	339	4	8,055,782	6,485	6	6	6	73	33.1
40	Burkholderia cenocepacia AU 1054	BProt BB	13919	3	7,279,116	6,477	6	6	6	67	33.1
41	Ralstonia eutropha JMP134	BProt BB	10646	4	7,255,290	6,446	6	6	6	66	35.6
42	Rhodococcus erythropolis PR4	BActin AA	20395	4	6,895,538	6,437	5	5	5	54	37.7
43	Methylobacterium radiotolerans JCM 2831	BProt ARh	18817	9	6,899,110	6,431	6	6	6	60	29.0
44	Rhizobium leguminosarum bv. trifolii WSM2304	BProt ARh	20179	5	6,872,702	6,415	3	3	3	53	38.8
45	Rhizobium sp. NGR234	BProt ARh	21101	3	6,891,900	6,366	3	3	3	54	37.6

837	Wolbachia endosymbiont strain TRS of Brugia malayi	BProt ARi	12475	1	1,080,084	805	1	1	1	34	65.8
838	Tropheryma whipplei TW08/27	BActin AA	354	1	925,938	784	1	1	1	51	53.7
839	Mycoplasma pulmonis UAB CTIP	BFirm MM	100	1	963,879	782	2	1	1	29	73.4
840	uncultured Termite group 1 bacterium phylotype Rs-D17	BCE	20871	4	1,148,570	776	1	1	1	45	64.8
841	Onion yellows phytoplasma OY-M	BFirm MA	9615	1	860,631	754	2	2	2	32	72.3
842	Mycoplasma agalactiae PG2	BFirm MM	16095	1	877,438	751	2	2	2	34	70.3
843	Mycoplasma gallisepticum R	BFirm MM	409	1	996,422	726	2	2	2	32	68.6
844	Mycoplasma conjunctivae	BTMM	32285	1	846,214	694	1	1	1	28	71.3
845	Aster yellows witches-broom phytoplasma AYWB	BFirm MA	13478	5	723,970	693	2	2	2	32	73.2
846	Mycoplasma hyopneumoniae 232	BFirm MM	13120	1	892,758	691	1	1	1	30	71.4
847	Mycoplasma pneumoniae M129	BFirm MM	99	1	816,394	688	1	1	1	37	60.0
848	Mesoplasma florum L1	BFirm ME	10650	1	793,224	683	2	2	2	29	73.0
849	Mycoplasma hyopneumoniae J	BFirm MM	10675	1	897,405	674	1	1	1	30	71.5
850	Mycoplasma synoviae 53	BFirm MM	10676	1	799,476	672	3	2	2	34	71.5
851	Mycoplasma hyopneumoniae 7448	BFirm MM	10639	1	920,079	663	1	1	1	30	71.5
852	Ureaplasma urealyticum serovar 10 str. ATCC 33699	BTMM	20247	1	874,478	646	2	2	2	30	74.2
853	Mycoplasma mobile 163K	BFirm MM	10697	1	777,079	635	1	1	1	28	75.1
854	Mycoplasma arthritidis 158L3-1	BFirm MM	1422	1	820,453	631	1	1	1	33	69.3
855	Ureaplasma parvum serovar 3 str. ATCC 700970	BFirm MM	101	1	751,719	611	2	2	2	30	74.5
856	Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	BProt GE	274	2	703,004	611	2	2	2	34	77.5
857	Candidatus Blochmannia pennsylvanicus str. BPEN	BProt GE	13875	1	791,654	610	1	1	1	40	70.4
858	Ureaplasma parvum serovar 3 str. ATCC 27815	BFirm MM	19087	1	751,679	609	2	2	2	30	74.5
859	Baumannia cicadellinicola str. Hc (Homalodisca coagulata)	BProt GC	12513	1	686,194	595	2	2	2	39	66.8
860	Candidatus Blochmannia floridanus	BProt GE	443	1	705,557	589	1	1	1	37	72.6
861	Buchnera aphidicola str. APS (Acyrthosiphon pisum)	BProt GE	245	3	655,725	574	1	1	1	32	73.6
862	Buchnera aphidicola str. 5A (Acyrthosiphon pisum)	BProt GE	31225	1	642,122	555	1	1	1	32	73.7
863	Buchnera aphidicola str. Tuc7 (Acyrthosiphon pisum)	BProt GE	31223	1	641,895	553	1	1	1	32	73.7
864	Buchnera aphidicola str. Sg (Schizaphis graminum)	BProt GE	312	1	641,454	545	1	1	1	32	74.7
865	Buchnera aphidicola str. Bp (Baizongia pistaciae)	BProt GE	256	2	618,379	507	1	1	1	32	74.7
866	Candidatus Phytoplasma mali	BTMA	25335	1	601,943	497	2	2	2	32	78.6
867	Mycoplasma genitalium G37	BFirm MM	97	1	580,076	476	1	1	1	36	68.3
868	Buchnera aphidicola str. Cc (Cinara cedri)	BProt GE	16372	2	422,434	365	1	1	1	31	79.8
869	Candidatus Sulcia muelleri GWSS	BBFF	19617	1	245,530	227	1	1	1	31	77.6
870	Candidatus Carsonella ruddii PV	BProt GC	17977	1	159,662	182	0	1	1	28	83.4
871	Candidatus Hodgkinia cicadicola Dsem	BProt AC	32135	1	143,795	169	1	1	1	15	41.6



3. AT content

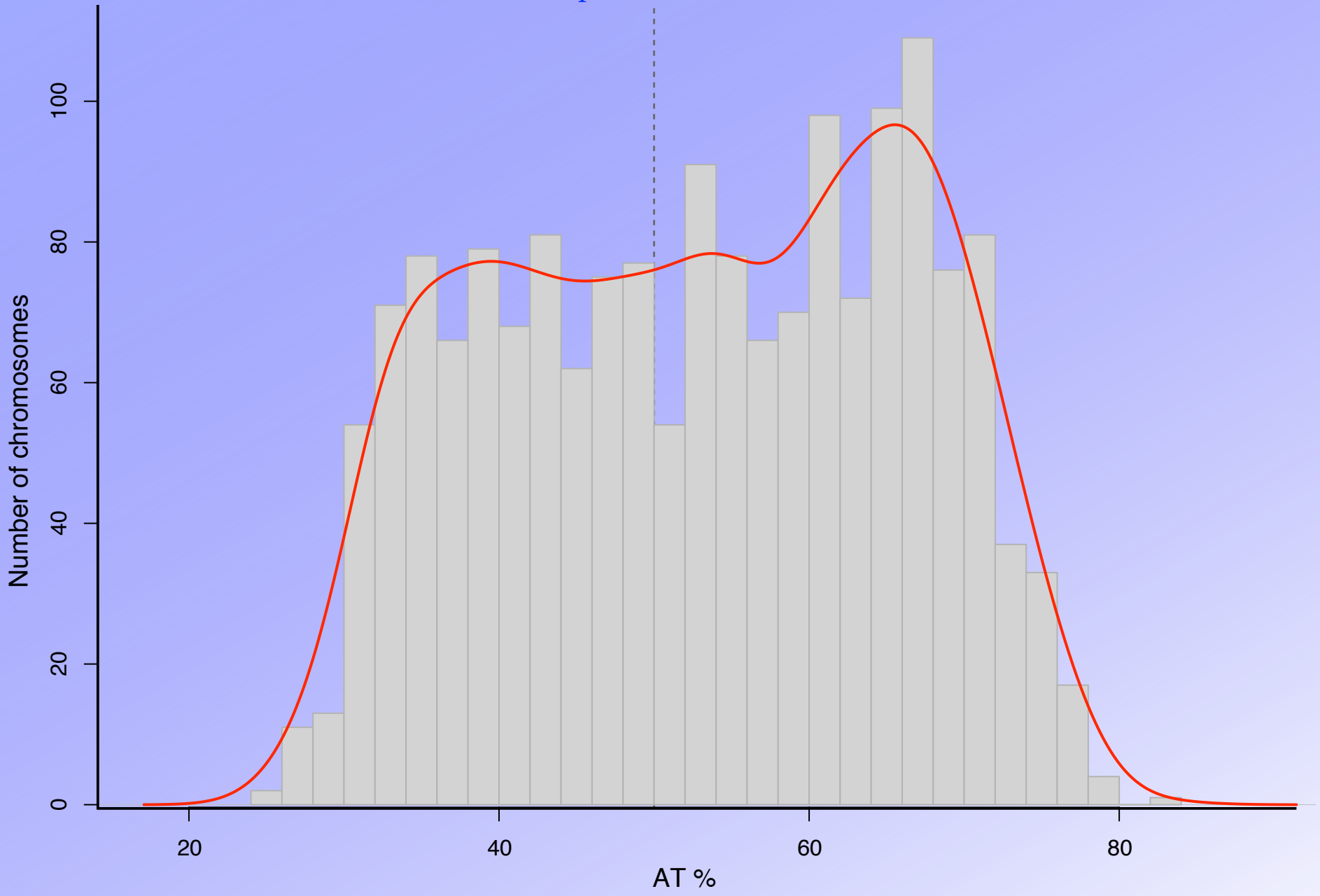


D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009

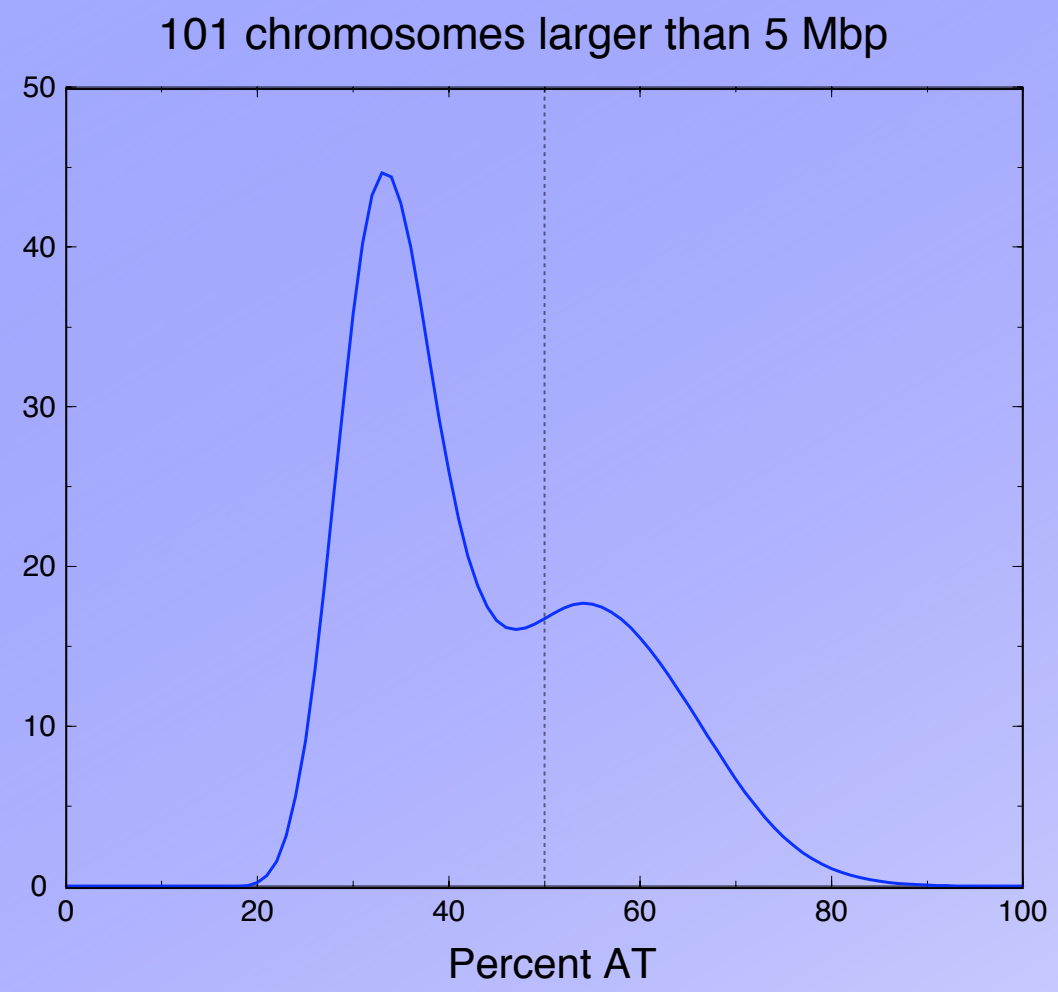
Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark

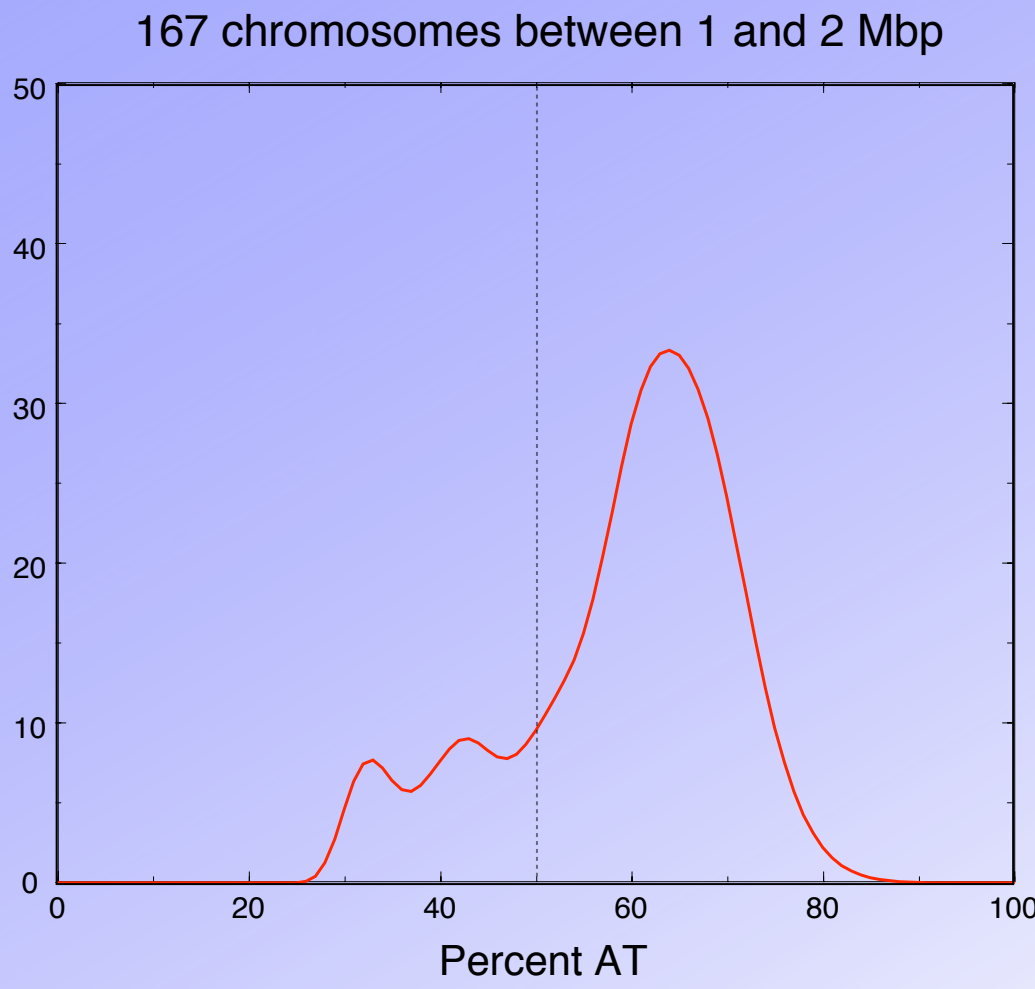
AT content in 1723 Prokaryotic Genomes

All chromosomes / plasmids in GenBank (Dec 2007)



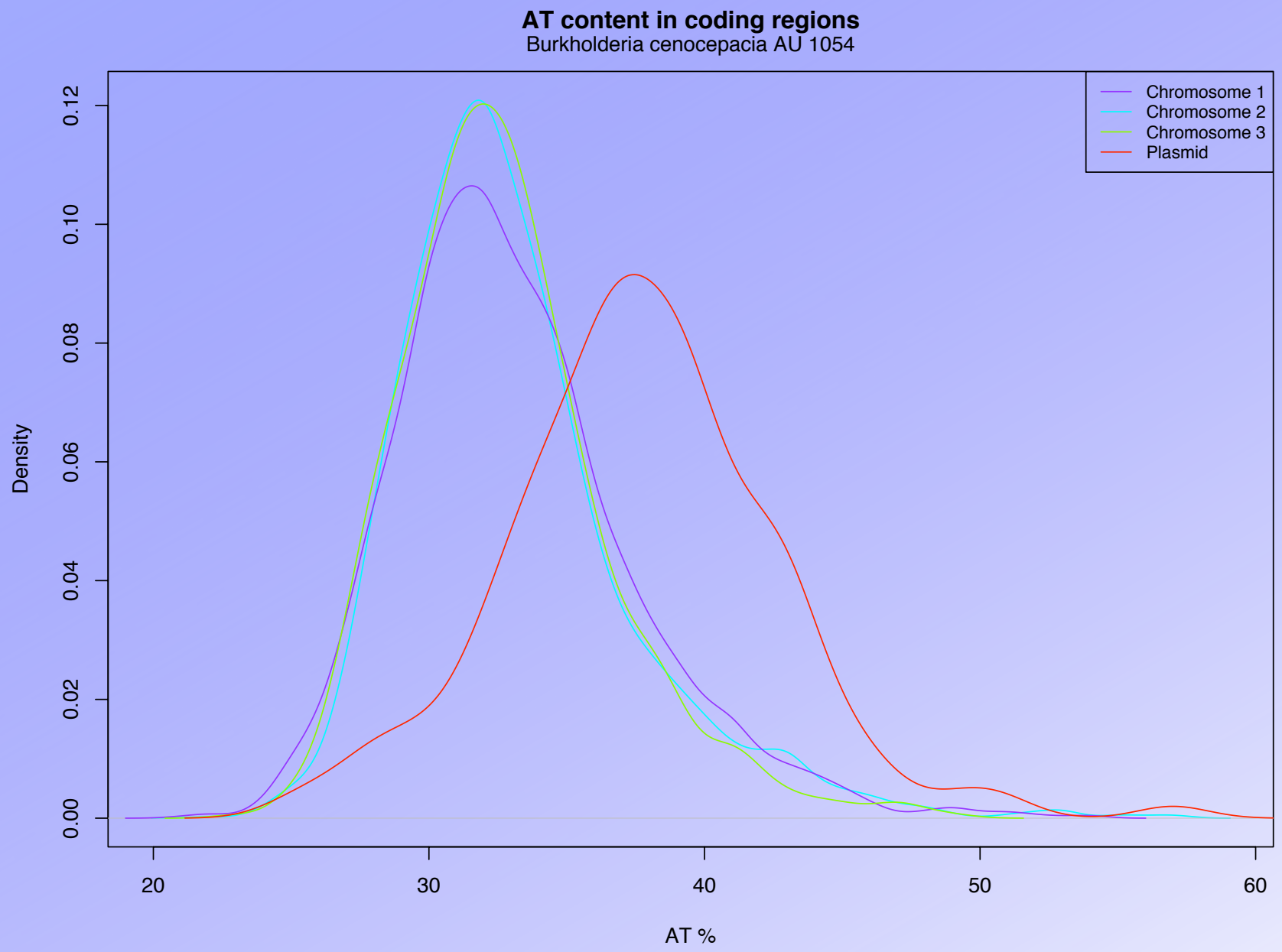
D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009





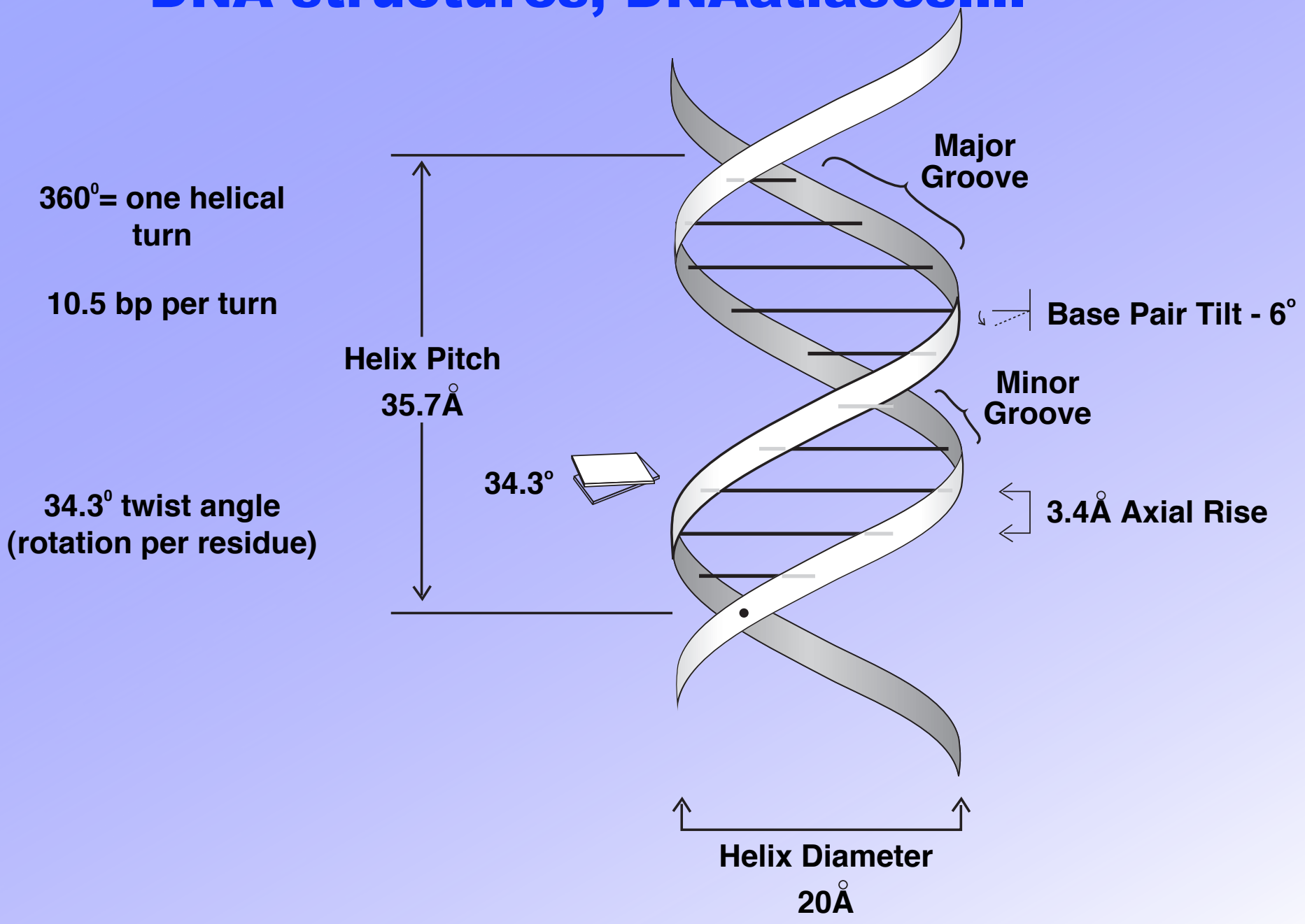
D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009

AT content varies amongst different chromosomes

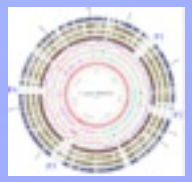


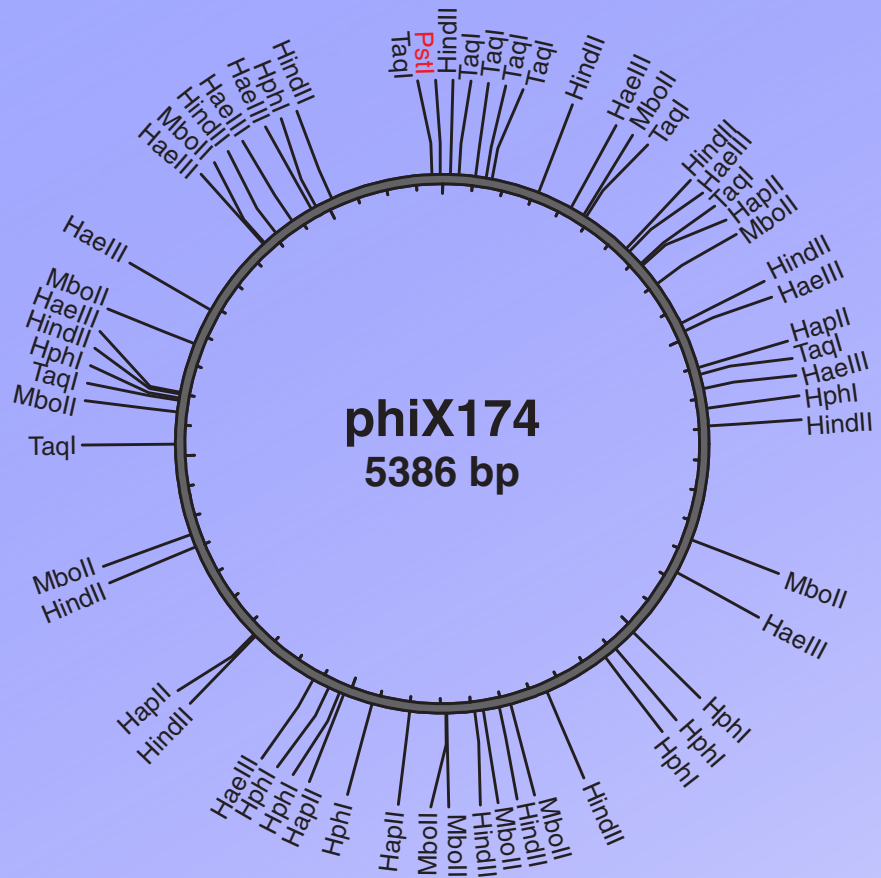
D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009

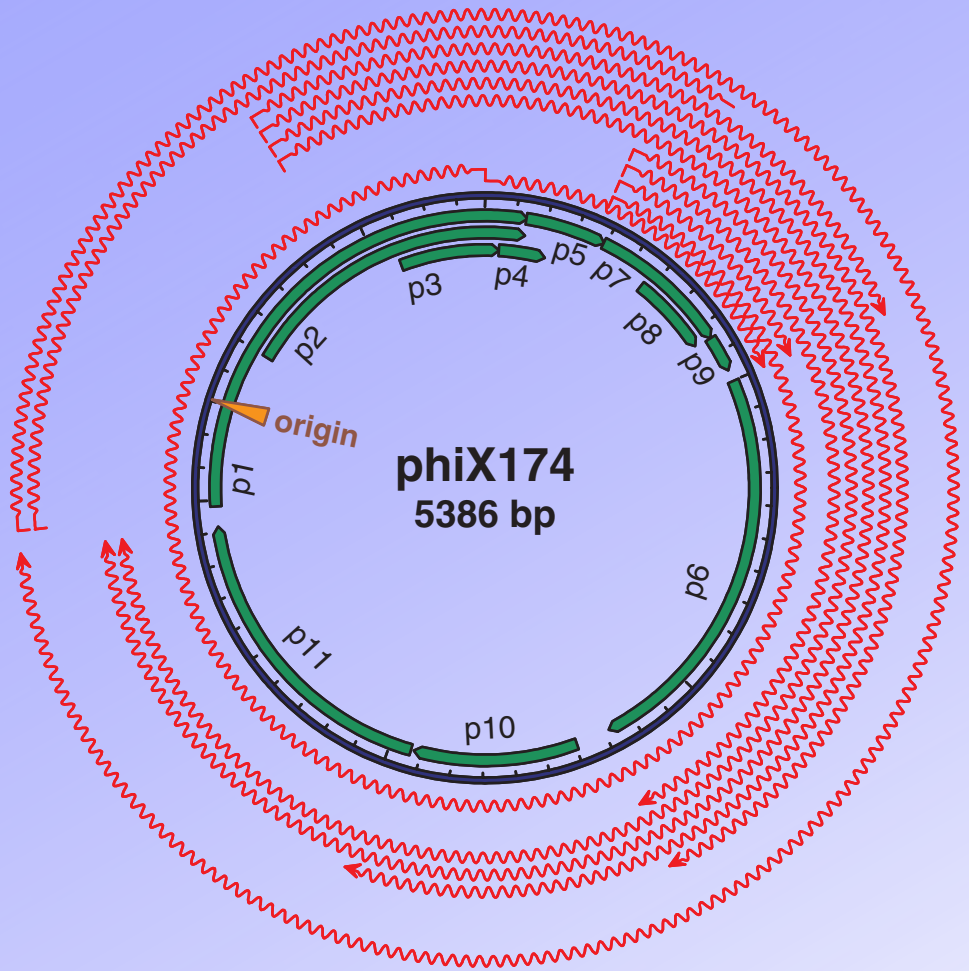
a digression - a few words about DNA structures, DNAatlases....



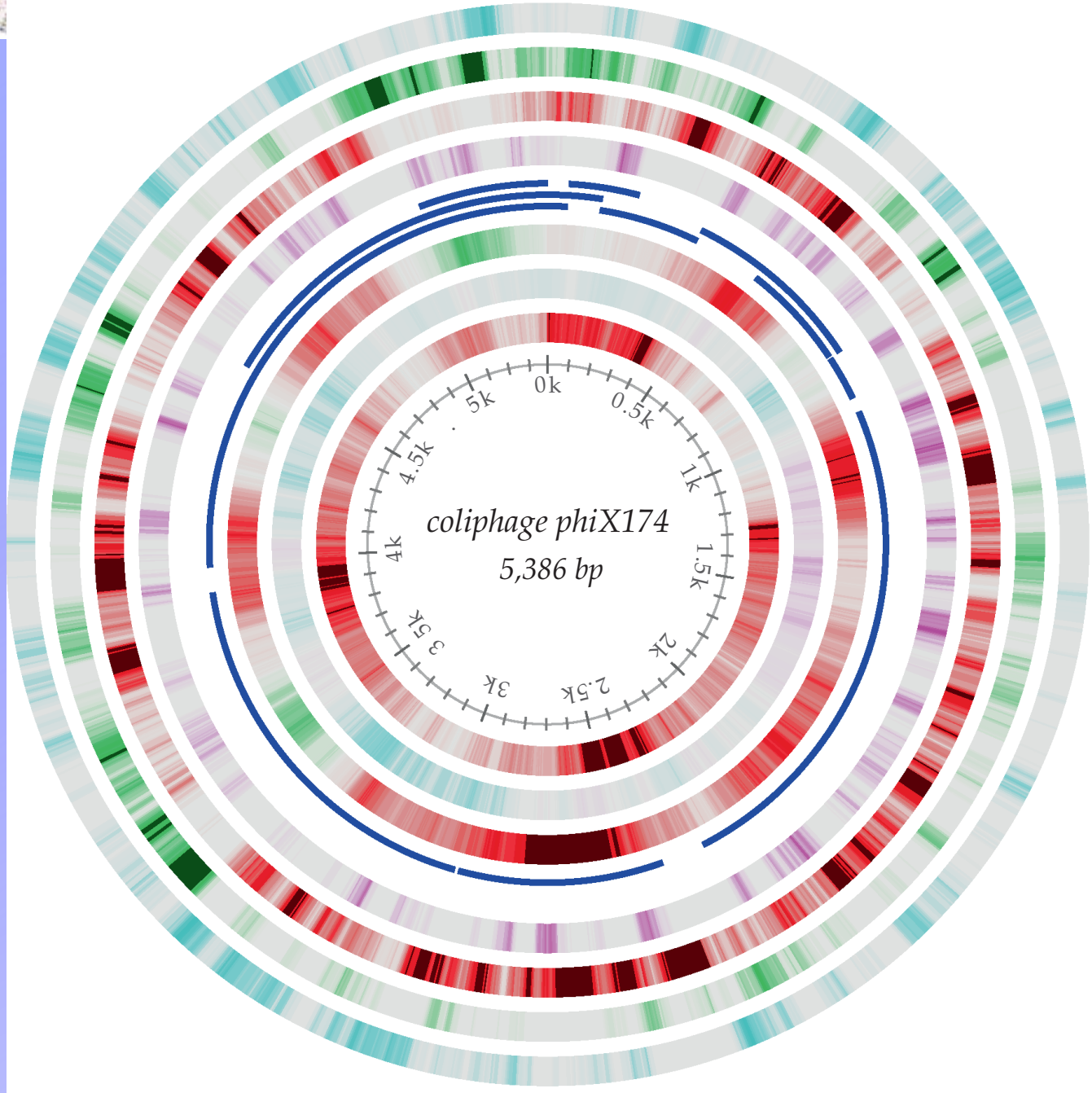
gagttttatc gcttccatga cgcagaagtt aacactttcg gatatttctg atgagtcgaa aaattatctt gataaagcag gaattactac tgcttgttta cgaattaaat cgaagtggac
tgctggcgga aaatgagaaa attcgaccta tccttgcgca gctcgagaag ctcttacttt gcgacctttc gccatcaact aacgattctg tcaaaaactg acgcgttgga tgaggagaag
tggcttaata tgcttggcac gttcgtcaag gactggttta gatatgagtc acattttgtt catggtagag attctcttgt tgacatttta aaagagcgtg gattactatc tgagtccgat
gctgttcaac cactaatagg taagaaatca tgagtcaagt tactgaacaa tccgtacgtt tccagaccgc tttggcctct attaagctca ttcaggcttc tgccgttttg gatttaaccg
aagatgattt cgattttctg acgagtaaca aagtttggat tgctactgac cgctctcgtg ctcgtcgctg cgttgaggct tgcgtttatg gtacgctgga ctttgtggga taccctcgct
ttcctgctcc tgttgagttt attgctgccg tcattgctta ttatgttcat cccgtcaaca ttcaaacggc ctgtctcatc atggaaggcg ctgaatttac ggaaaacatt attaatggcg
tcgagcgtcc ggttaaagcc gctgaattgt tcgcgtttac cttgcgtgta cgcgcaggaa acactgacgt tcttactgac gcagaagaaa acgtgcgtca aaaattacgt gcggaaggag
tgatgtaatg tctaaaggta aaaaacgttc tggcgctcgc cctggtcgtc cgcagccgtt gcgaggtact aaaggcaagc gtaaaggcgc tcgtctttgg tatgtaggtg gtcaacaatt
ttaattgcag gggcttcggc cccttacttg aggataaatt atgtctaata ttcaaactgg cgccgagcgt atgccgcatg acctttccca tcttggcttc cttgctggtc agattggtcg
tcttattacc atttcaacta ctccggttat cgctggcgac tccttcgaga tggacgccgt tggcgctctc cgtctttctc cattgcgtcg tggccttgct attgactcta ctgtagacat
ttttactttt tatgtccctc atcgtcacgt ttatggtgaa cagtggatta agttcatgaa ggatggtgtt aatgccactc ctctcccgac tgttaacact actggttata ttgaccatgc
cgcttttctt ggcacgatta accctgatac caataaaatc cctaagcatt tgtttcaggg ttatttgaat atctataaca actattttaa agcgccgtgg atgcctgacc gtaccgaggc
taaccctaat gagcttaatc aagatgatgc tcgttatggt ttccgttgct gccatctcaa aaacatttgg actgctccgc ttcctcctga gactgagctt tctcgccaaa tgacgacttc
taccacatct attgacatta tgggtctgca agctgcttat gctaatttgc atactgacca agaacgtgat tacttcatgc agcgttacca tgatgttatt tcttcatttg gaggtaaaac
ctcttatgac gctgacaacc gtcctttact tgtcatgcgc tctaatctct gggcatctgg ctatgatgtt gatggaactg accaaacgtc gttaggccag ttttctggtc gtgttcaaca
gacctataaa cattctgtgc cgcgtttctt tgttcctgag catggcacta tgtttactct tgcgcttgtt cgttttccgc ctactgcgac taaagagatt cagtacctta acgctaaagg
tgctttgact tataccgata ttgctggcga ccctgttttg tatggcaact tgccgccgcg tgaaatttct atgaaggatg ttttccgttc tggtgattcg tctaagaagt ttaagattgc
tgagggtcag tggtatcgtt atgcgccttc gtatgtttct cctgcttatc accttcttga aggcttccca ttcattcagg aaccgccttc tggtgatttg caagaacgcg tacttattcg
ccaccatgat tatgaccagt gtttccagtc cgttcagttg ttgcagtgga atagtcaggt taaatttaat gtgaccgttt atcgcaatct gccgaccact cgcgattcaa tcatgacttc
gtgataaaag attgagtgtg aggttataac gccgaagcgg taaaaatttt aatttttgcc gctgaggggt tgaccaagcg aagcgcggta ggttttctgc ttaggagttt aatcatgttt
cagactttta tttctcgcca taattcaaac tttttttctg ataagctggt tctcacttct gttactccag cttcttcggc acctgtttta cagacaccta aagctacatc gtcaacgtta
tattttgata gtttgacggt taatgctggt aatggtggtt ttcttcattg cattcagatg gatacatctg tcaacgccgc taatcaggtt gtttctgttg gtgctgatat tgcttttgat
gccgacccta aatttttttgc ctgtttggtt cgctttgagt cttcttcggt tccgactacc ctcccgactg cctatgatgt ttatcctttg aatggtcgcc atgatggtgg ttattatacc
gtcaaggact gtgtgactat tgacgtcctt ccccgtacgc cgggcaataa cgtttatgtt ggtttcatgg tttggtctaa ctttaccgct actaaatgcc gcggattggt ttcgctgaat
aagagattat ttgtctccag ccacttaagt gaggtgattt atgtttggtg ctattgctgg cggtattgct tctgctcttg ctggtggcgc catgtctaaa ttgtttggag gcggtcaaaa
agccgcctcc ggtggcattc aaggtgatgt gcttgctacc gataacaata ctgtaggcat gggtgatgct ggtattaaat ctgccattca aggctctaat gttcctaacc ctgatgaggc
cgcccctagt tttgtttctg gtgctatggc taaagctggt aaaggacttc ttgaaggtac gttgcaggct ggcacttctg ccgtttctga taagttgctt gatttggttg gacttggtgg
caagtctgcc gctgataaag gaaaggatac tcgtgattat cttgctgctg catttcctga gcttaatgct tgggagcgtg ctggtgctga tgcttcctct gctggtatgg ttgacgccgg
atttgagaat caaaaagagc ttactaaaat gcaactggac aatcagaaag agattgccga gatgcaaaat gagactcaaa aagagattgc tggcattcag tcggcgactt cacgccagaa
tacgaaagac caggtatatg cacaaaatga gatgcttgct tatcaacaga aggagtctac tgctcgcgtt gcgtctatta tggaaaacac caatctttcc aagcaacagc aggtttccga
gattatgcgc caaatgctta ctcaagctca aacggctggt cagtatttta ccaatgacca aatcaaagaa atgactcgca aggttagtgc tgaggttgac ttagttcatc agcaaacgca
gaatcagcgg tatggctctt ctcatattgg cgctactgca aaggatattt ctaatgtcgt cactgatgct gcttctggtg tggttgatat ttttcatggt attgataaag ctgttgccga
tacttggaac aatttctgga aagacggtaa agctgatggt attggctcta atttgtctag gaaataaccg tcaggattga caccctccca attgtatgtt ttcatgcctc caaatcttgg
aggctttttt atggttcgtt cttattaccc ttctgaatgt cacgctgatt attttgactt tgagcgtatc gaggctctta aacctgctat tgaggcttgt ggcatttcta ctctttctca
atccccaatg cttggcttcc ataagcagat ggataaccgc atcaagctct tggaagagat tctgtctttt cgtatgcagg gcgttgagtt cgataatggt gatatgtatg ttgacggcca
taaggctgct tctgacgttc gtgatgagtt tgtatctgtt actgagaagt taatggatga attggcacaa tgctacaatg tgctccccca acttgatatt aataacacta tagaccaccg
ccccgaaggg gacgaaaaat ggtttttaga gaacgagaag acggttacgc agttttgccg caagctggct gctgaacgcc ctcttaagga tattcgcgat gagtataatt accccaaaaa
gaaaggtatt aaggatgagt gttcaagatt gctggaggcc tccactatga aatcgcgtag aggctttgct attcagcgtt tgatgaatgc aatgcgacag gctcatgctg atggttggtt
tatcgttttt gacactctca cgttggctga cgaccgatta gaggcgtttt atgataatcc caatgctttg cgtgactatt ttcgtgatat tggtcgtatg gttcttgctg ccgagggtcg
caaggctaat gattcacacg ccgactgcta tcagtatttt tgtgtgcctg agtatggtac agctaatggc cgtcttcatt tccatgcggt gcactttatg cggacacttc ctacaggtag
cgttgaccct aattttggtc gtcgggtacg caatcgccgc cagttaaata gcttgcaaaa tacgtggcct tatggttaca gtatgcccat cgcagttcgc tacacgcagg acgctttttc
acgttctggt tggttgtggc ctgttgatgc taaaggtgag ccgcttaaag ctaccagtta tatggctgtt ggtttctatg tggctaaata cgttaacaaa aagtcagata tggaccttgc
tgctaaaggt ctaggagcta aagaatggaa caactcacta aaaaccaagc tgtcgctact tcccaagaag ctgttcagaa tcagaatgag ccgcaacttc gggatgaaaa tgctcacaat
gacaaatctg tccacggagt gcttaatcca acttaccaag ctgggttacg acgcgacgcc gttcaaccag atattgaagc agaacgcaaa aagagagatg agattgaggc tgggaaaagt
tactgtagcc gacgttttgg cggcgcaacc tgtgacgaca aatctgctca aatttatgcg cgcttcgata aaaatgattg gcgtatccaa cctgca







Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark



G Content
0.00 fix avg 0.40

A Content
0.00 fix avg 0.40

T Content
0.00 fix avg 0.40

C Content
0.00 fix avg 0.40

Annotations:
CDS +

AT Skew
–0.20 fix avg 0.20

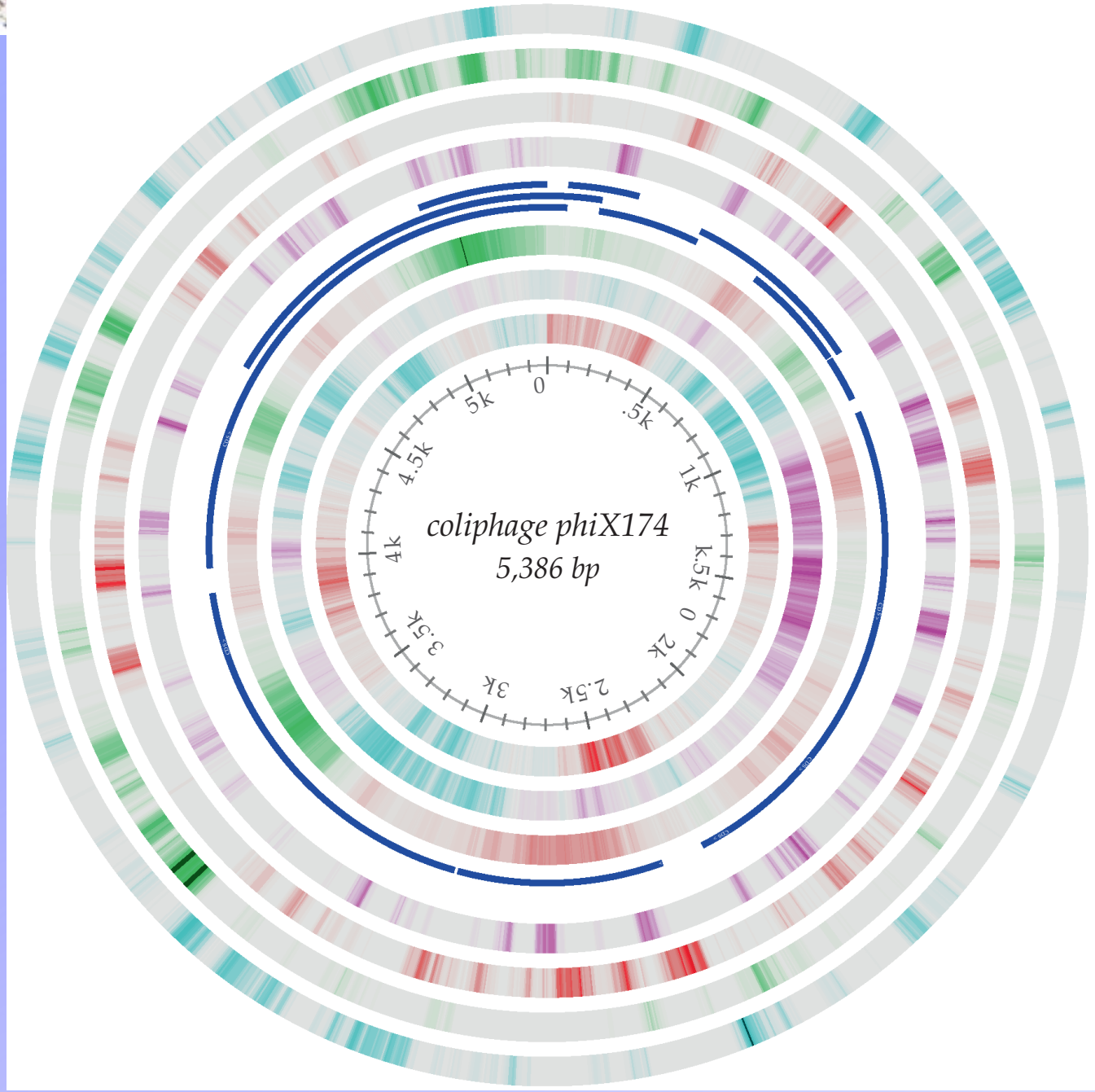
GC Skew
–0.20 fix avg 0.20

Percent AT
0.40 fix avg 0.60

Resolution: 3

BASE ATLAS

Center for Biological Sequence analysis
Department of Systems Biology, Technical University of Denmark



G Content
dev avg
0.07 0.39

A Content
dev avg
0.01 0.47

T Content
dev avg
0.10 0.53

C Content
dev avg
0.04 0.39

Annotations:
CDS +

AT Skew
dev avg
–0.33 0.18

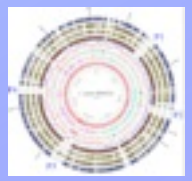
GC Skew
dev avg
–0.10 0.14

Percent AT
dev avg
0.47 0.63

Resolution: 3

BASE ATLAS

Center for Biological Sequence analysis
Department of Systems Biology, Technical University of Denmark



Direct repeats

5‘-AGAGAGAGATCTAGACTAGAGAGAGAGAGATACCCTAGAGAGAGAGAGAGAAAG-3’
3‘-TCTCTCTCTAGATCTGATCTCCCTCTCTCTATGGGATCTCTCTCTCTCTCTTTC-5’

Local direct simple repeats

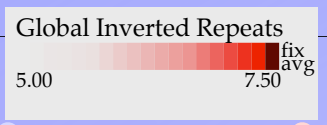
5‘-AGGCAAAGTCGCCATGCTGCCATCTGGAGGCAAAGTCGCCATGCTGCCATCTGG-3’
3‘-TCCGTTTCAGCGGTACGACGGTAGACGTCCGTTTCAGCGGTACGACGGTAGACC-5’

Local direct repeats

5‘-AGGCAAAGTCGCCATGCTGCCATCTGGTGCCCAGGCAAAGTCGCCATGCTGCCATCTGG-3’
3‘-TCCGTTTCAGCGGTACGACGGTAGACGACGGGTCCGTTTCAGCGGTACGACGGTAGACC-5’

Local direct repeats with spacer

Inverted repeats



5‘-AGGCAAAGTCGCCATGCTGCCATCTGGCCAGATGGCAGCATGGCGACTTTGCCT-3’
3‘-TCCGTTTCAGCGGTACGACGGTAGACCGGTCTACCGTCGTACCGCTGAAACGGA-5’

Local inverted repeat

5‘-AGGCAAAGTCGCCATGCTGCCATCTGGTGCCCCCAGATGGCAGCATGGCGACTTTGCCT-3’
3‘-TCCGTTTCAGCGGTACGACGGTAGACCACGCCGGTCTACCGTCGTACCGCTGAAACGGA-5’

Inverted repeat with spacer

5‘-AGGCACGTGAGAATGAATTCCTGCTTTAGATCTAAAAGGCAAAGCTTTGCCT-3’
3‘-TCCGTGCACTCTTACTTAAGGACGAAATCTAGATTTTCCGTTTCGAAACGGT-5’

Palindromic repeats

Two global direct repeats

and one global inverted repeat

(not on scale)

Mirror and Everted repeats

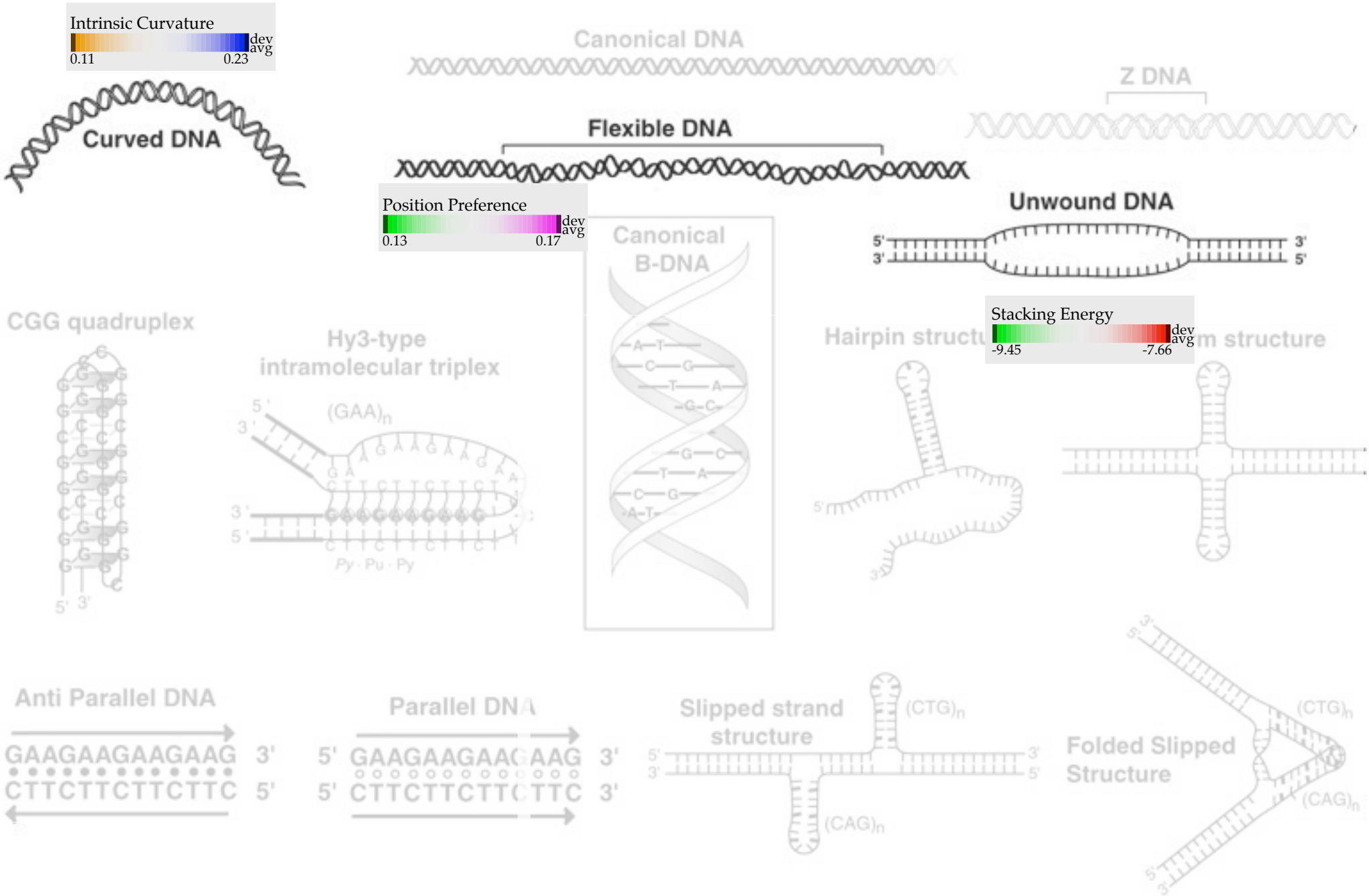
5‘-AGGCAAAGTCGCCATGCTGCCATCTGGGGTCTACCGTCGTACCGCTGAAACGGA-3’
3‘-TCCGTTTCAGCGGTACGACGGTAGACCCCAGATGGCAGCATGGCGACTTTGCCT-5’

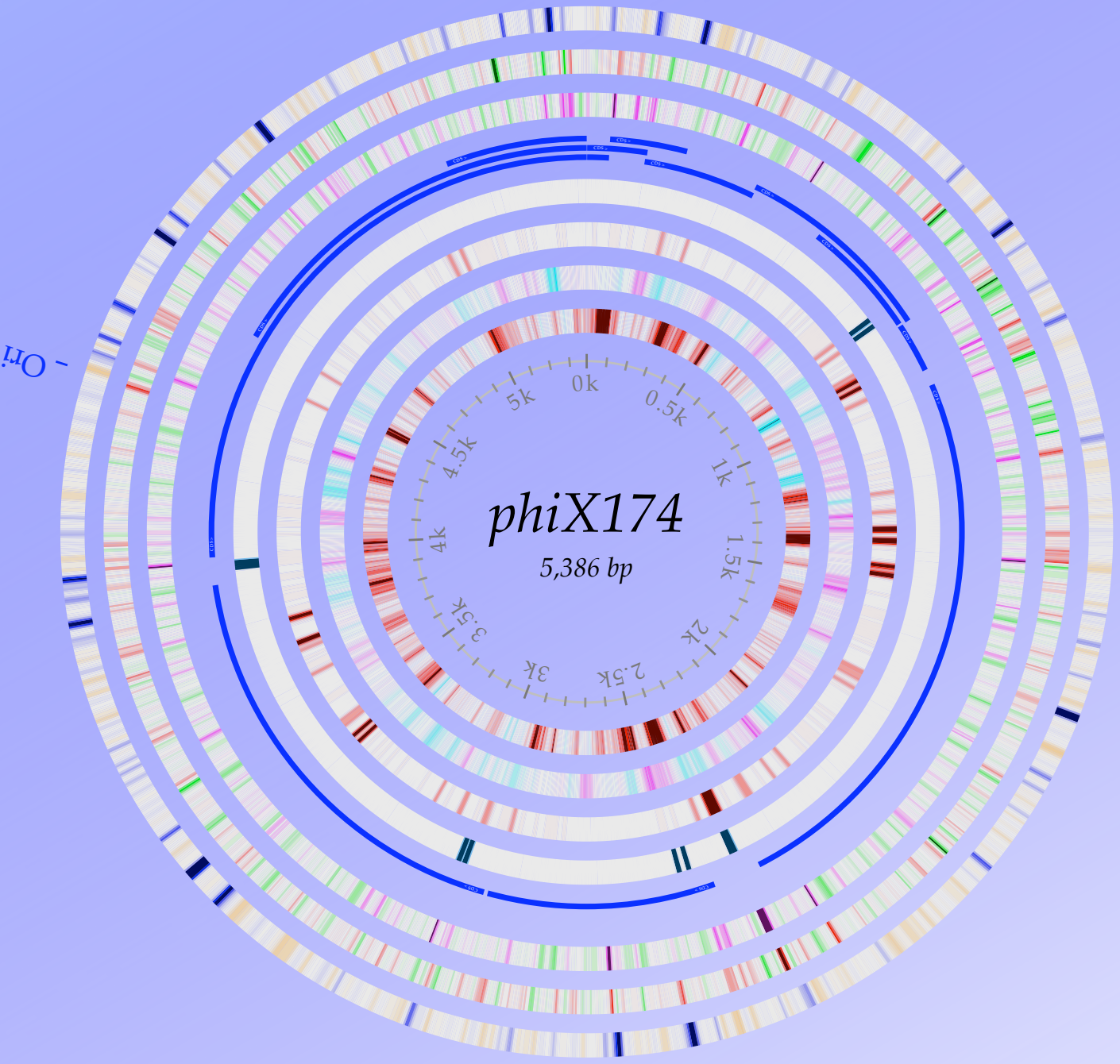
Mirror repeat

5‘-AGGCAAAGTCGCCATGCTGCCATCTGGTCCGTTTCAGCGGTACGACGGTAGACC-3’
3‘-TCCGTTTCAGCGGTACGACGGTAGACCAGGCAAAGTCGCCATGCTGCCATCTGG-5’

Everted repeat

Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark


Intrinsic Curvature
dev
avg
0.11
0.23
Curved DNA
Canonical DNA
Z DNA
Flexible DNA
Position Preference
dev
avg
0.13
0.17
Unwound DNA
5'
3'
3'
5'
Canonical
B-DNA
Stacking Energy
dev
avg
-9.45
-7.66
CGG quadruplex
Hy3-type
intramolecular triplex
(GAA)n
Py · Pu · Py
Hairpin structu
m structure
Anti Parallel DNA
5' GAAGAAGAAGAAG 3'
3' CTTCTTCTTCTTC 5'
Parallel DNA
5' GAAGAAGAAGAAG 3'
5' CTTCTTCTTCTTC 3'
Slipped strand
structure
(CTG)n
(CAG)n
Folded Slipped
Structure
(CTG)n
(CAG)n



Intrinsic Curvature

-0.09 0.51

Stacking Energy

-10.31 -5.40

Position Preference

0.03 0.25

Annotations:

CDS +

CDS -

Perfect-palindromes

-0.25 0.28

Inverted Repeats

3.54 7.83

GC Skew

-0.25 0.29

Percent AT

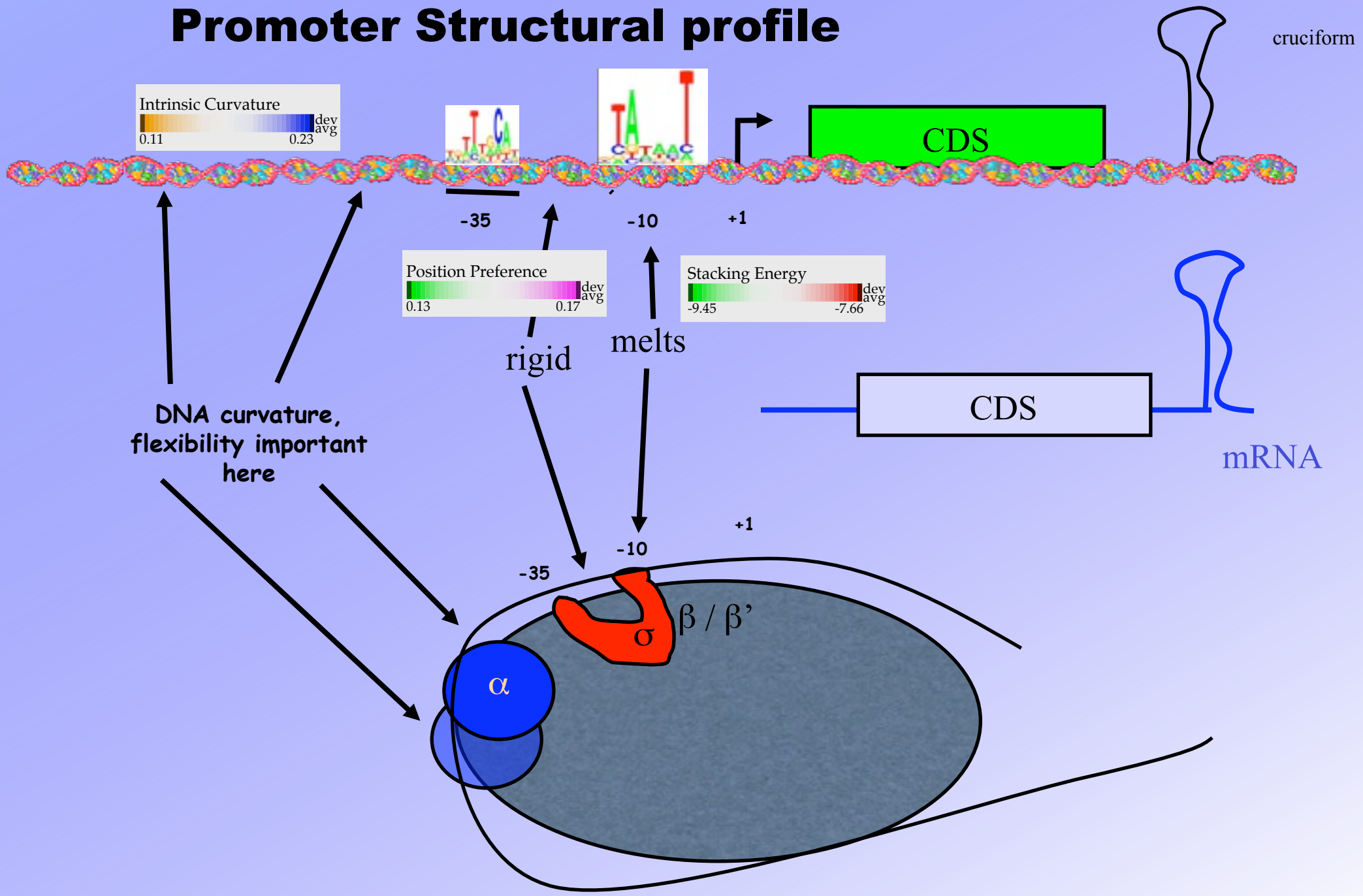
0.35 0.65

Resolution: 3

http://www.cbs.dtu.dk/
Center for Biological Sequence An

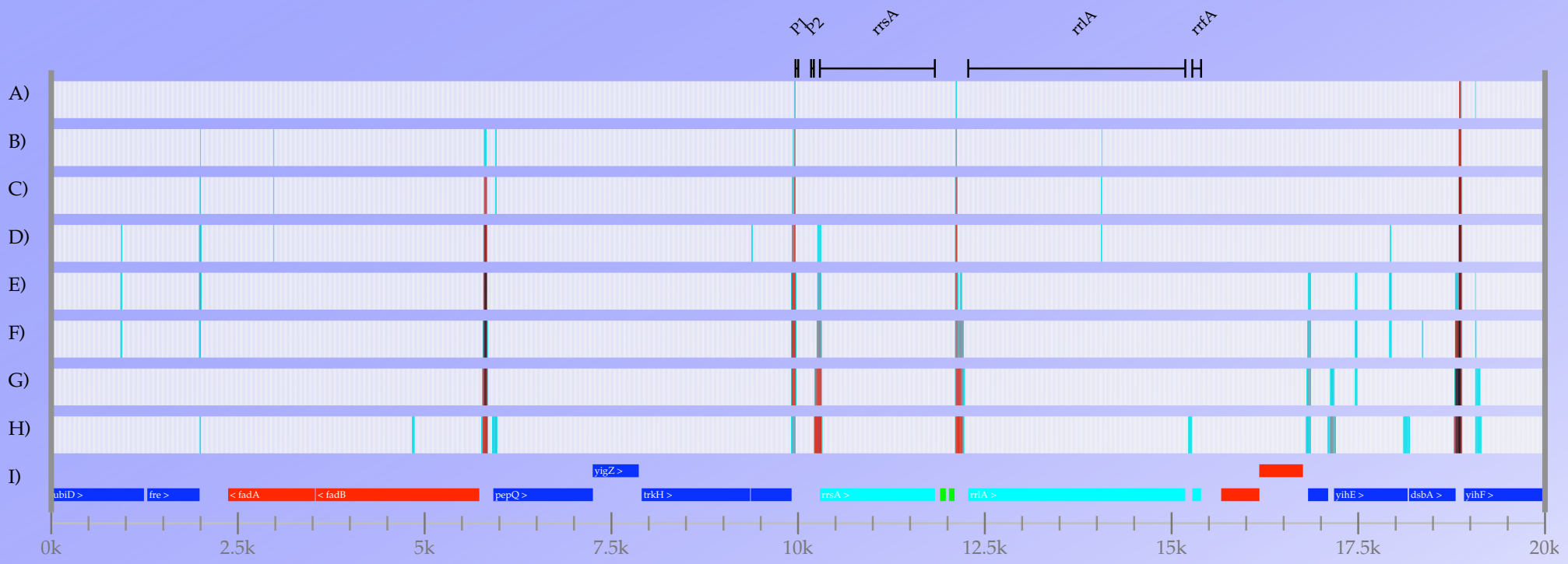
GENOME ATLAS

Promoter Structural profile

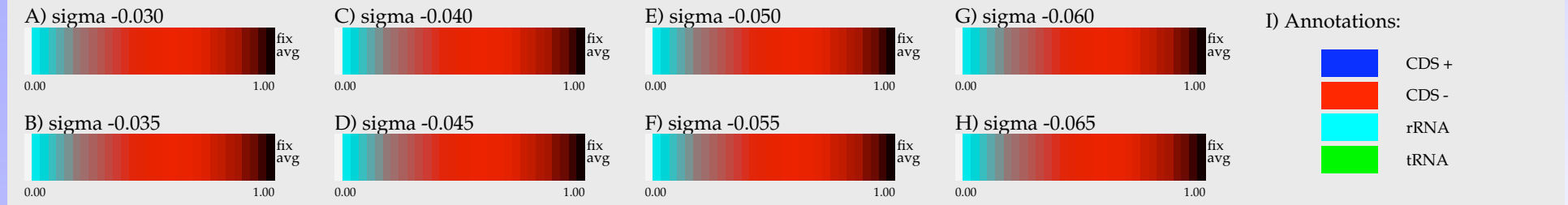


Ecoli_K12_MG1655 rrsA

20,000 bp



Resolution: 9



Center for Biological Sequence Analysis
http://www.cbs.dtu.dk/

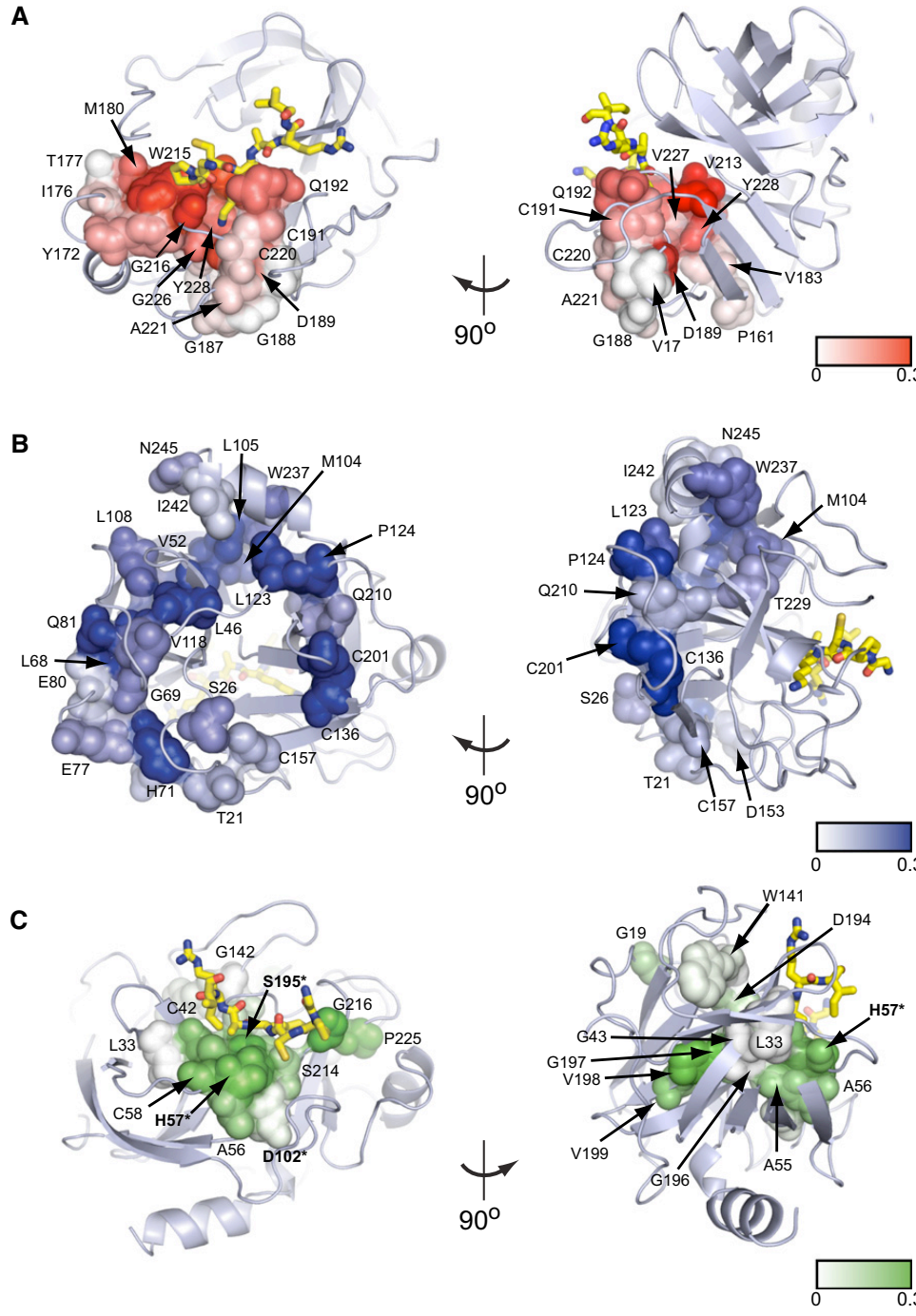
SIDD ATLAS

Summary (so far!)

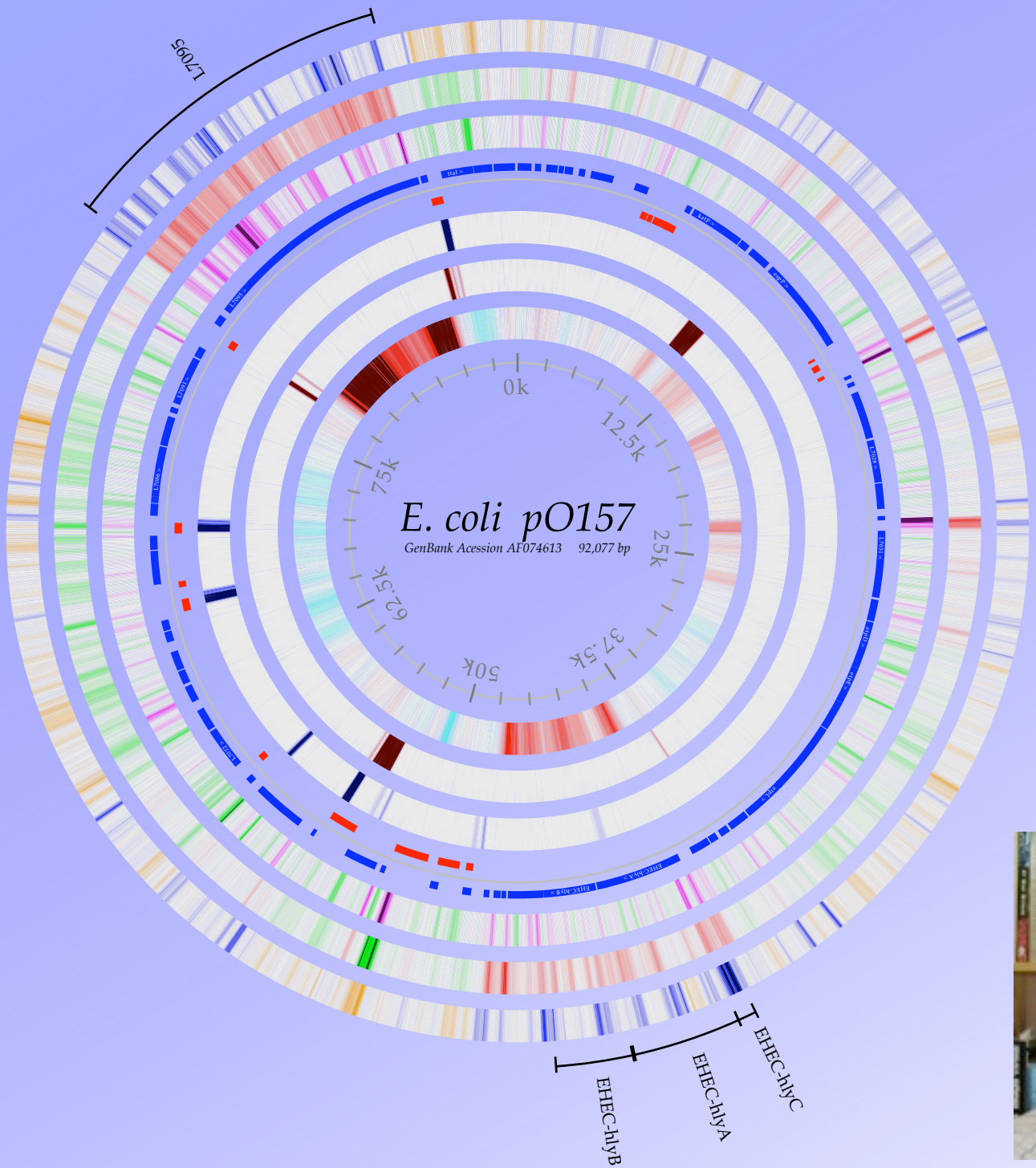
DNA
RNA
Protein

Sequences → Structure → Function

Center for Biological Sequence analysis
Department of Systems Biology, Technical University of Denmark


A
M180
T177
W215
I176
Q192
C191
Y172
C220
G216
Y228
G226
D189
A221
G187
G188
90°
V227
V213
Y228
Q192
C191
C220
V183
A221
G188
V17
D189
P161
0
0.3
B
N245
L105
W237
M104
I242
L108
P124
V52
L123
Q210
Q81
L46
V118
C201
L68
E80
G69
S26
C136
C157
E77
H71
T21
90°
N245
I242
W237
M104
L123
P124
Q210
T229
C201
C136
S26
T21
C157
D153
0
0.3
C
G142
S195*
C42
G216
P225
L33
S214
C58
H57*
A56
D102*
90°
W141
D194
G19
H57*
G43
L33
G197
V198
A56
V199
A55
G196
0
0.3

%AT content varies locally, along chromosomes



Intrinsic Curvature

0.08 0.30

Stacking Energy

-9.52 -6.41

Position Preference

0.11 0.17

Annotations:

CDS +

CDS -

Watson Repeats

5.00 7.50

Crick Repeats

5.00 7.50

Percent AT

0.30 0.70

Resolution: 37

Friis, Jensen, and Ussery

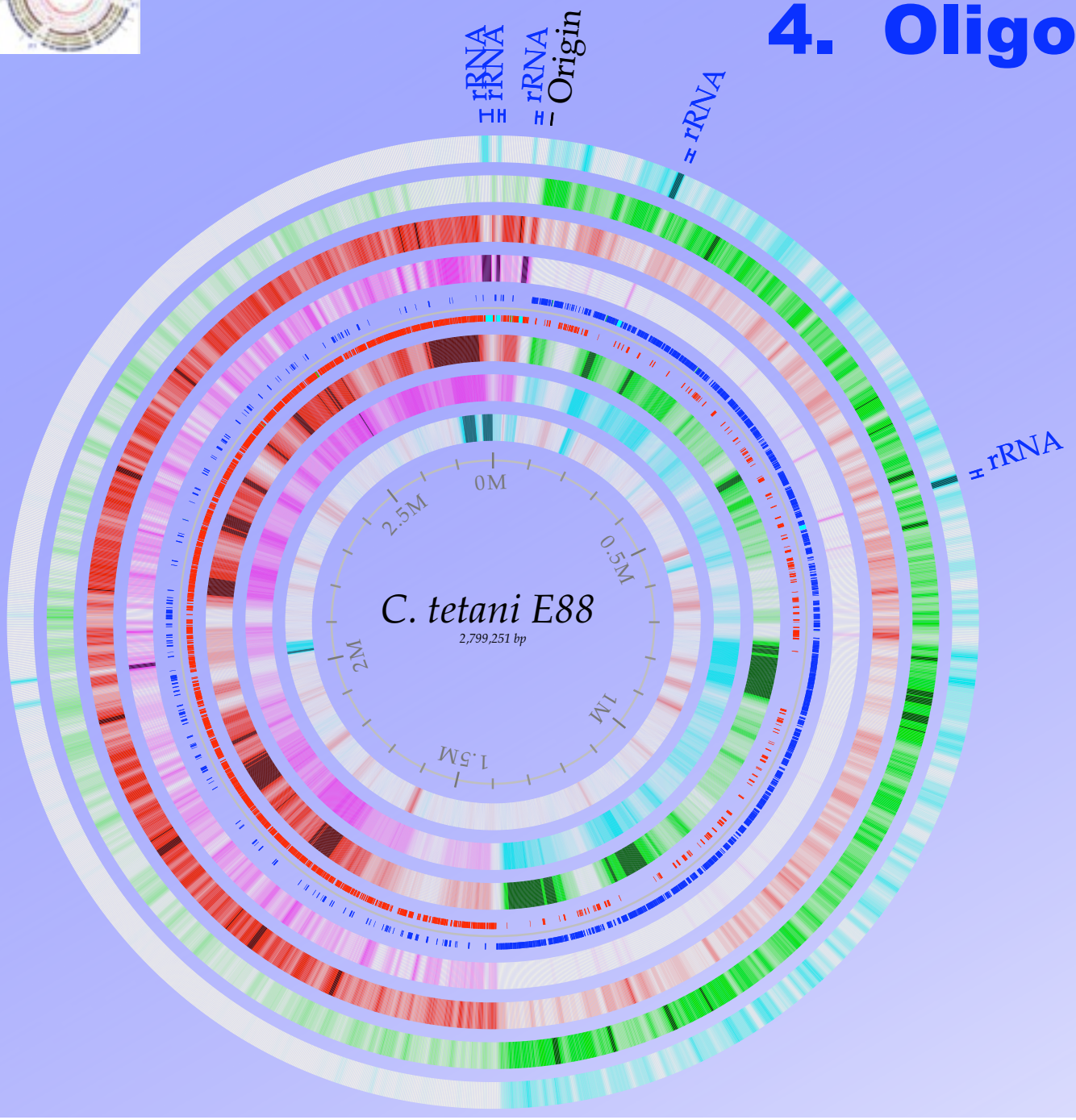
Genetica, **108**:47-51, (2000).

http://www.cbs.dtu.dk/

Center for Biological Sequence Ar

GENOME ATLAS

4. Oligomer skews



G Content
fix avg
0.00 0.23

A Content
fix avg
0.00 0.45

T Content
fix avg
0.00 0.45

C Content
fix avg
0.00 0.23

Annotations:
CDS +
CDS -
rRNA
tRNA

AT Skew
fix avg
-0.12 0.12

GC Skew
fix avg
-0.10 0.10

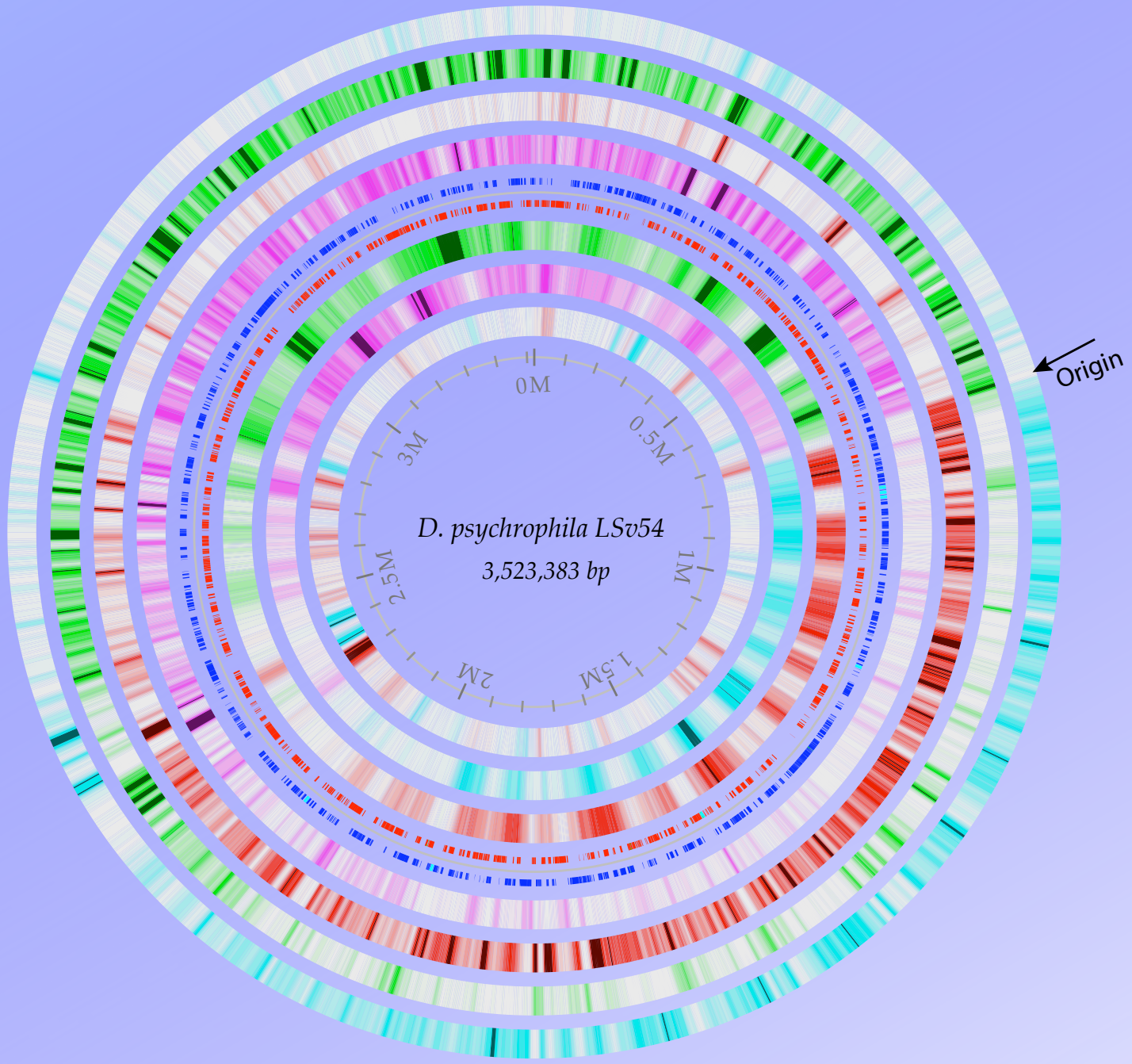
Percent AT
dev avg
0.67 0.75

Resolution: 1120

http://www.cbs.dtu.dk/
Center for Biological Sequence Analysis

Center for Biological Sequence analysis
Department of Systems Biology, Technical University of Denmark

Diagnosis, epidemiology and antibiotic resistance of the genus Clostridium, (Duchesnes C., Mainil J, Pelkonen S. and Menozzi MG, editors), Proceedings of the meeting, Presses de la Faculté de Médecine Vétérinaire de l'Université de Liège, pages 60-66, (2004).



G Content
0.10 0.29 fix avg

A Content
0.15 0.31 fix avg

T Content
0.15 0.31 fix avg

C Content
0.10 0.29 fix avg

Annotations:
CDS +
CDS -
rRNA
tRNA

AT Skew
-0.07 0.07 fix avg

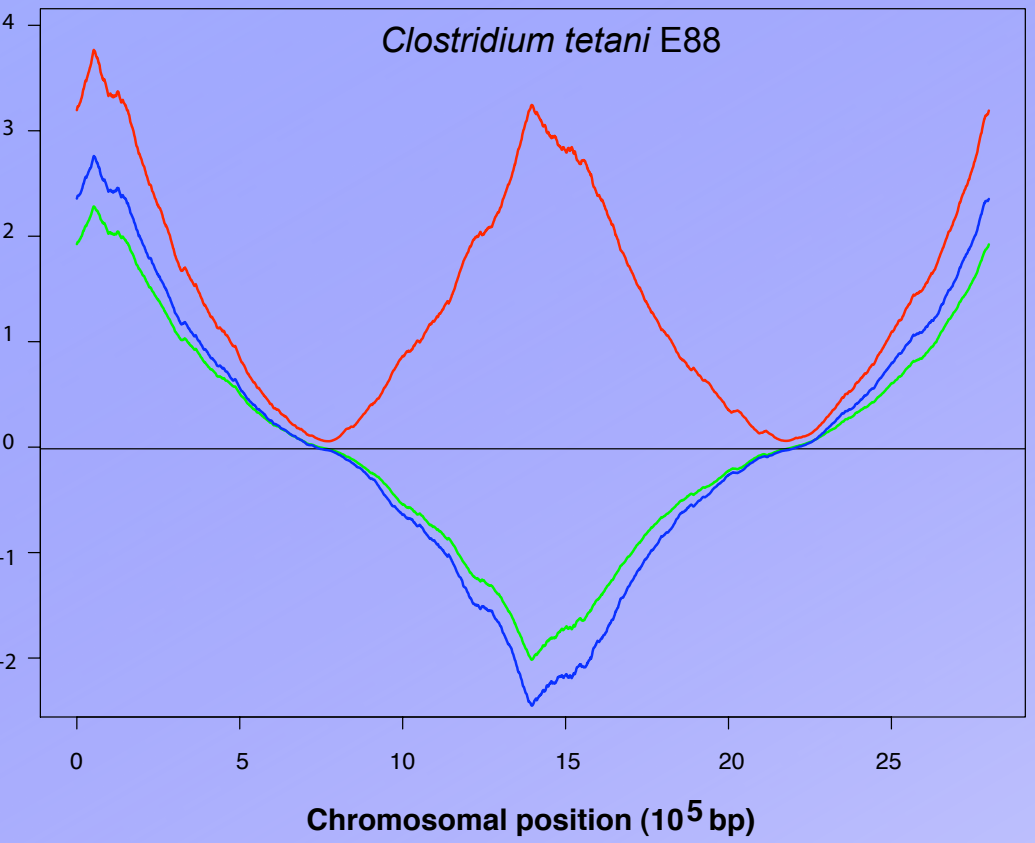
GC Skew
-0.07 0.07 fix avg

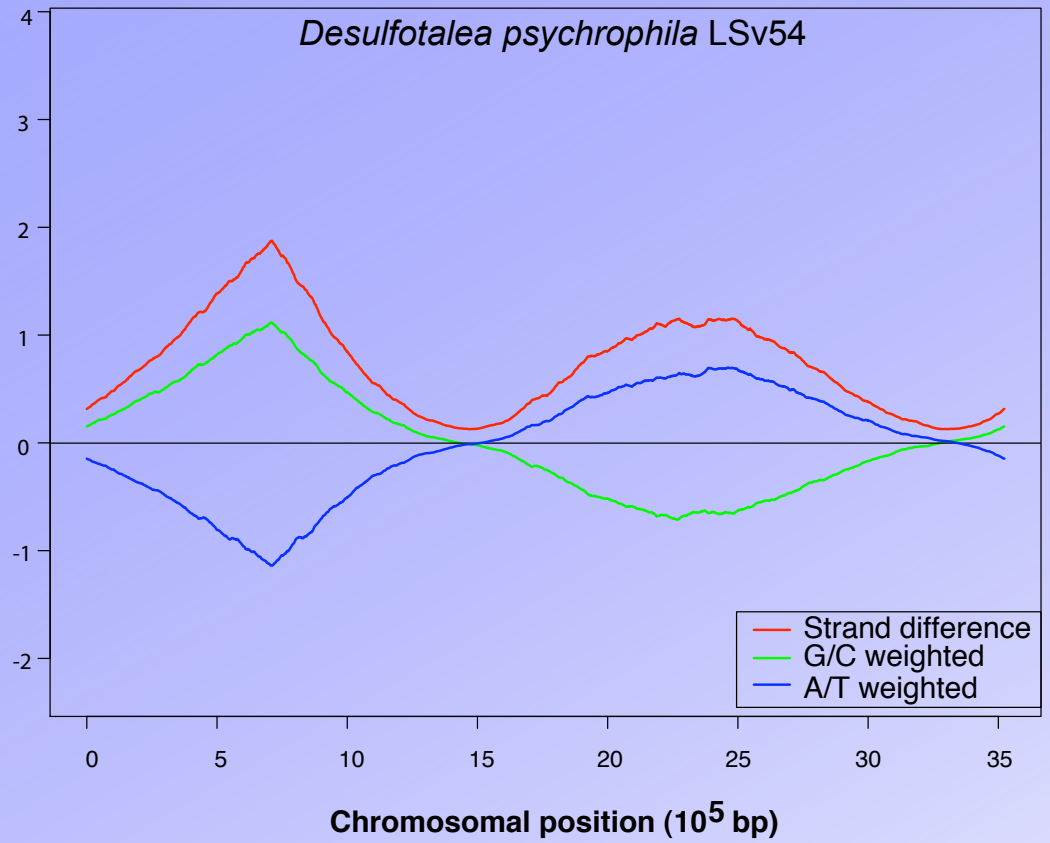
Percent AT
0.48 0.58 dev avg

Resolution: 1410

BASE ATLAS

D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009



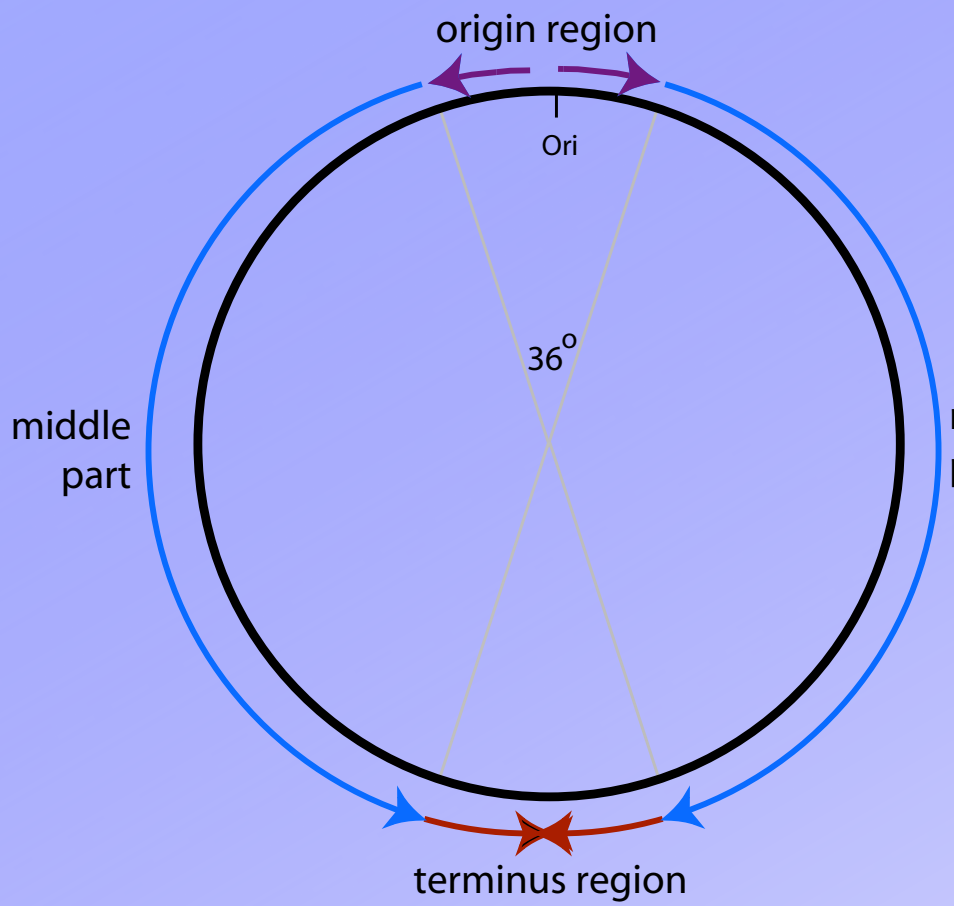


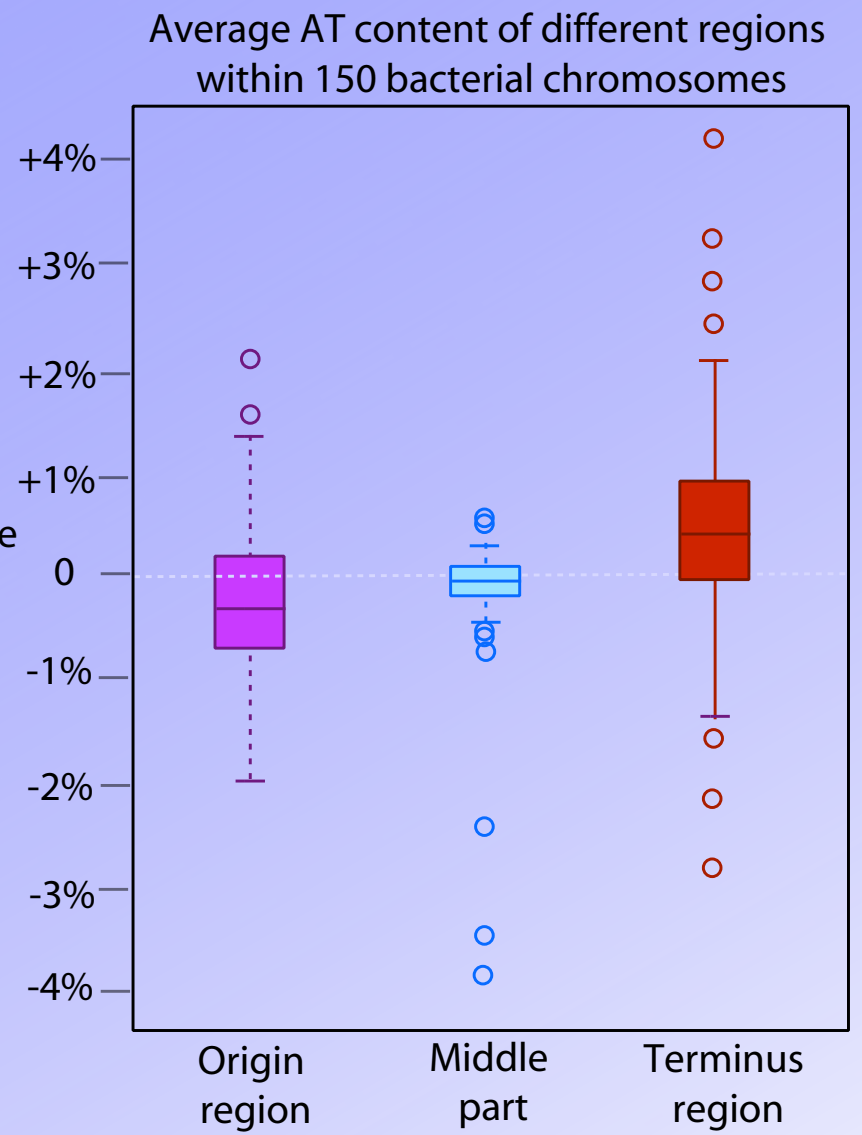
Lars

Peder

D.W. Ussery et al., *Computing for Comparative Microbial Genomi*
Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag

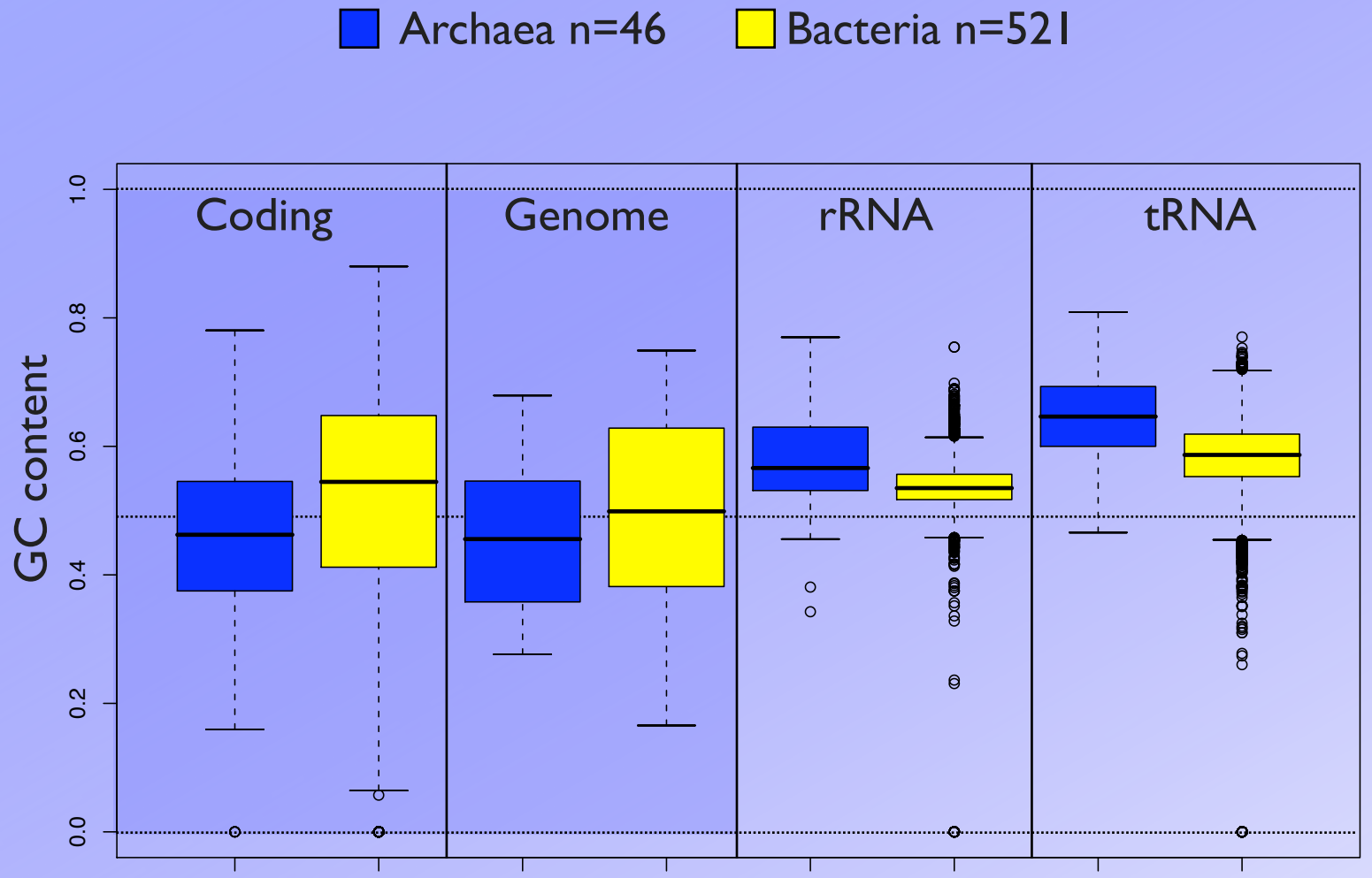
Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark



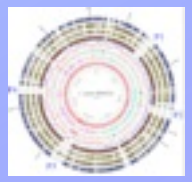


D.W. Ussery et al., *Computing for Comparative Microbial Genomics*, Computational Biology 8, DOI 10.1007/978-1-84800-255-5_14, © Springer-Verlag London Limited 2009

Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark

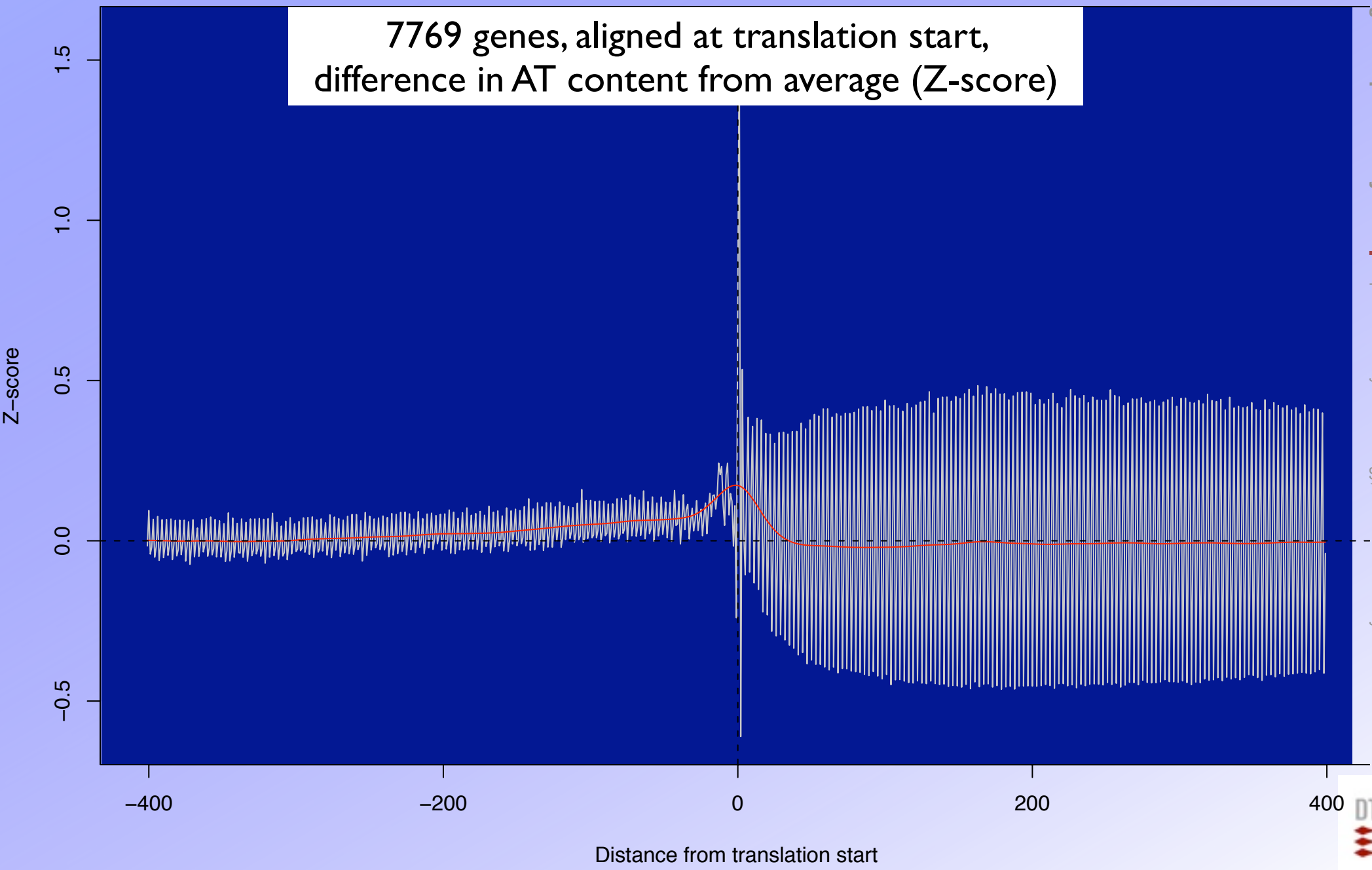


Center for Biological Sequence analysis
Department of Systems Biology, Technical University of Denmark

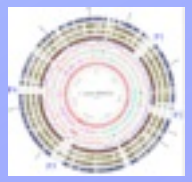


Streptomyces coelicolor, strain A3(2)

7769 genes, aligned at translation start,
difference in AT content from average (Z-score)



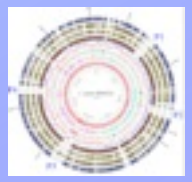
Center for Biological Sequence analysis
Department of Systems Biology, Technical University of Denmark



Codon Usage in *Streptomyces coelicolor* strain A3

2,577,562 codons in 7,825 orfs examined

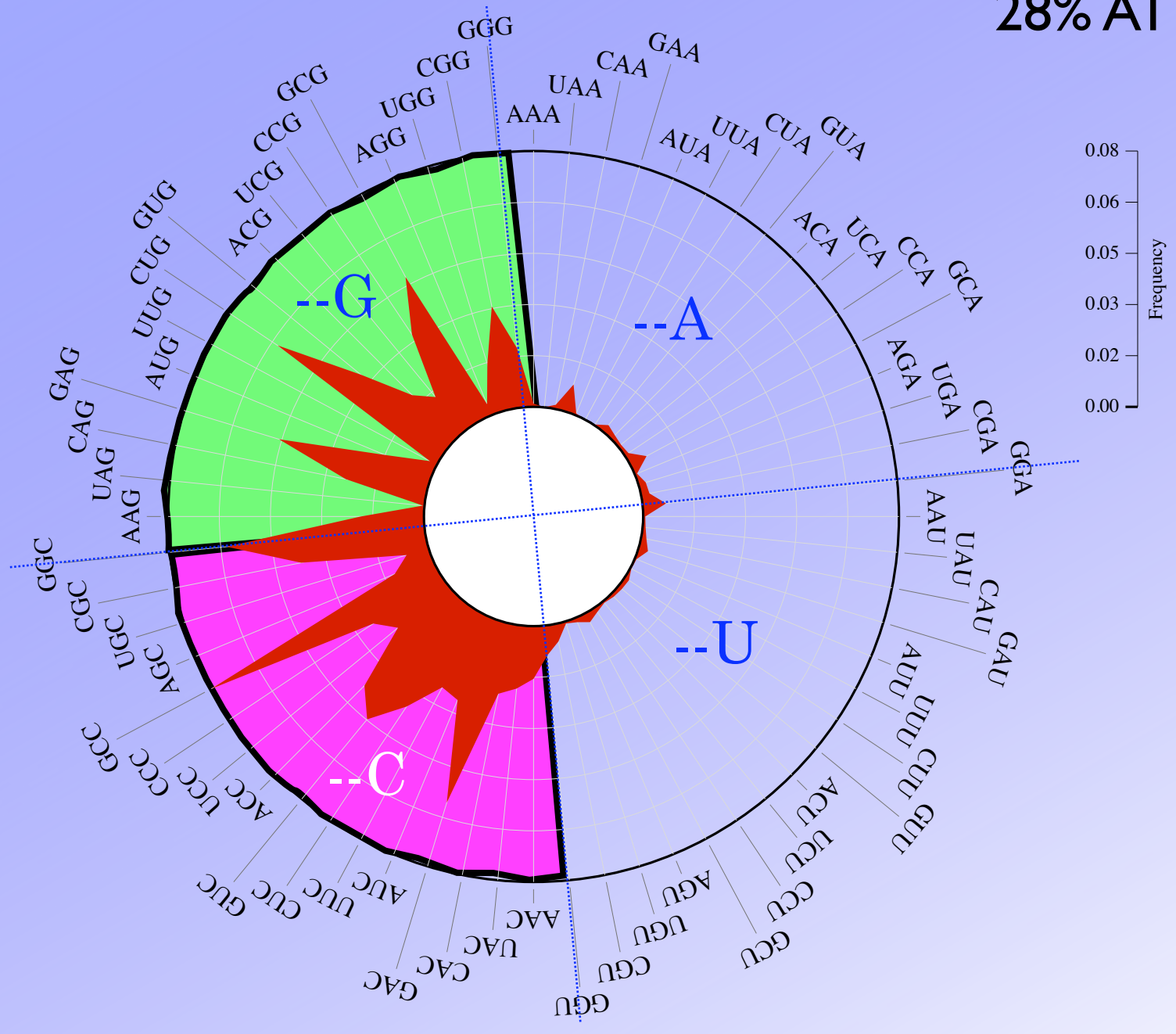
1st position	2nd position U			C			A			G			3rd position
U	0.04%	1,099	Phe	0.06%	1,609	Ser	0.10%	2,474	Tyr	0.07%	1,830	Cys	U
	2.60%	67,046	Phe	2.03%	52,291	Ser	1.95%	50,282	Tyr	0.70%	18,179	Cys	C
	0.01%	156	Leu	0.11%	2,745	Ser	0.01%	347	STOP	0.24%	6,117	STOP	A
	0.24%	6,141	Leu	1.39%	35,899	Ser	0.05%	1,380	STOP	1.51%	39,048	Trp	G
C	0.15%	3,869	Leu	0.15%	3,878	Pro	0.17%	4,250	His	0.54%	14,003	Arg	U
	3.65%	94,248	Leu	2.55%	65,721	Pro	2.17%	56,041	His	3.90%	100,568	Arg	C
	0.03%	831	Leu	0.13%	3,367	Pro	0.13%	3,294	Gln	0.25%	6,528	Arg	A
	6.12%	158,040	Leu	3.37%	87,004	Pro	2.49%	64,313	Gln	3.23%	83,228	Arg	G
A	0.06%	1,533	Ile	0.11%	2,940	Thr	0.07%	1,765	Asn	0.15%	3,857	Ser	U
	2.73%	70,370	Ile	3.96%	102,329	Thr	1.62%	41,802	Asn	1.23%	31,833	Ser	C
	0.07%	1,683	Ile	0.16%	4,025	Thr	0.10%	2,515	Lys	0.08%	1,991	Arg	A
	1.57%	40,571	Met	1.91%	49,224	Thr	1.94%	50,088	Lys	0.36%	9,398	Arg	G
G	0.14%	3,554	Val	0.29%	7,513	Ala	0.29%	7,545	Asp	0.93%	23,935	Gly	U
	4.71%	121,541	Val	7.85%	202,479	Ala	5.81%	149,973	Asp	6.15%	158,630	Gly	C
	0.26%	6,814	Val	0.54%	13,930	Ala	0.84%	21,727	Glu	0.72%	18,489	Gly	A
	3.53%	91,176	Val	5.01%	129,403	Ala	4.83%	124,591	Glu	1.85%	47,689	Gly	G

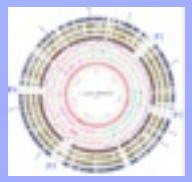


Codon Usage

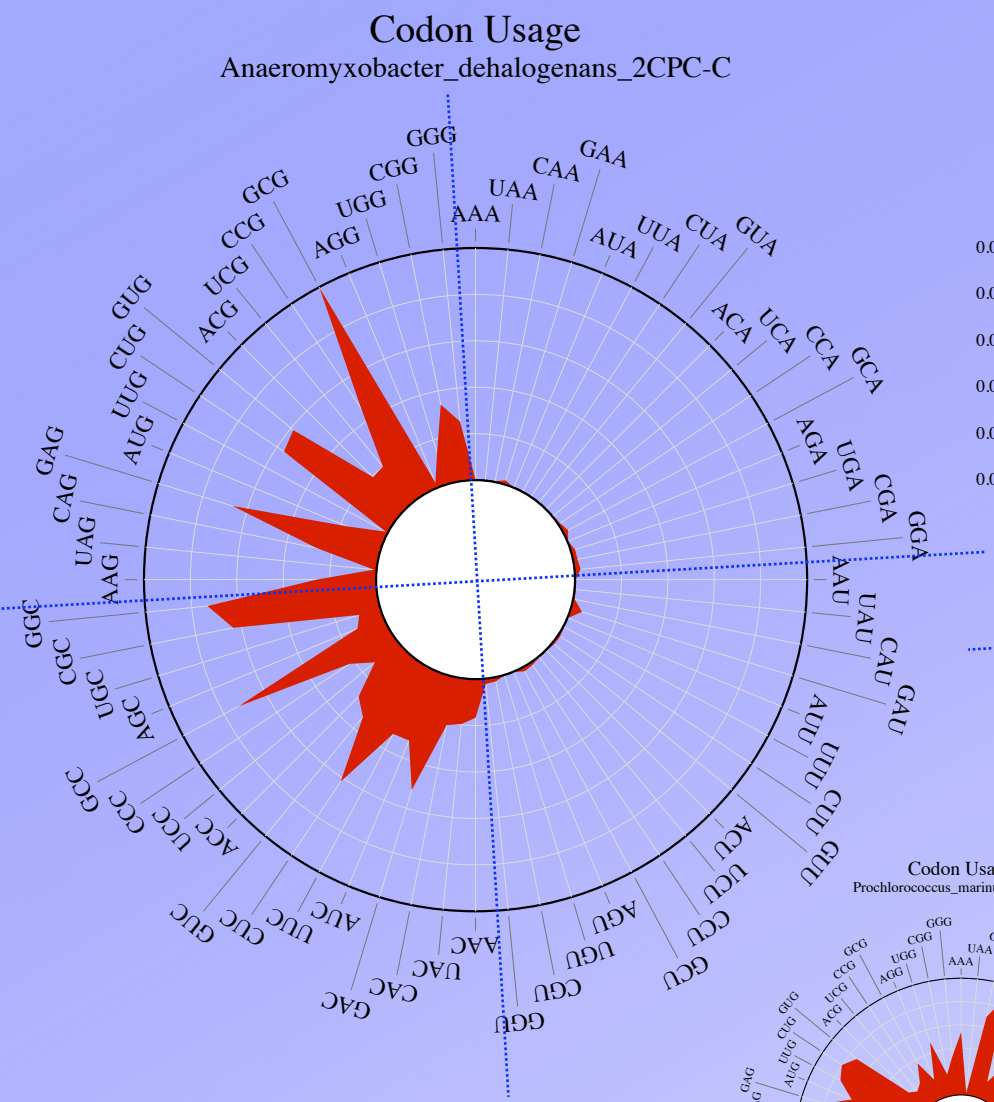
Streptomyces_coelicolor_A3

28% AT

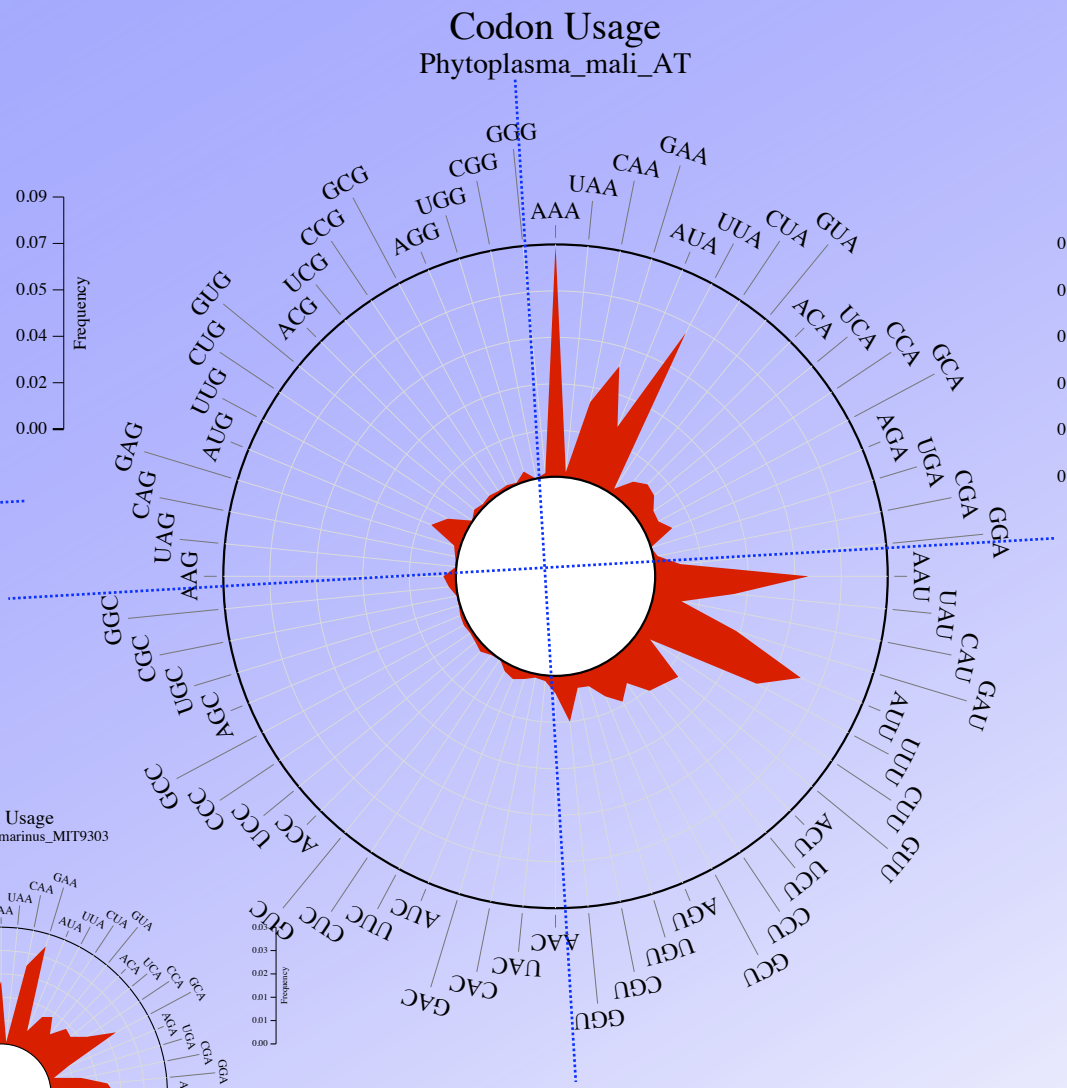


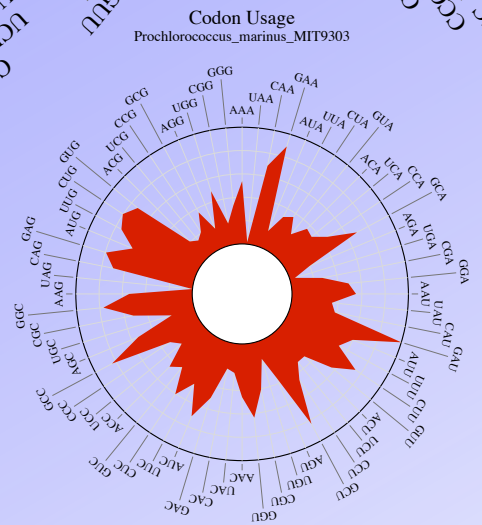


25% AT

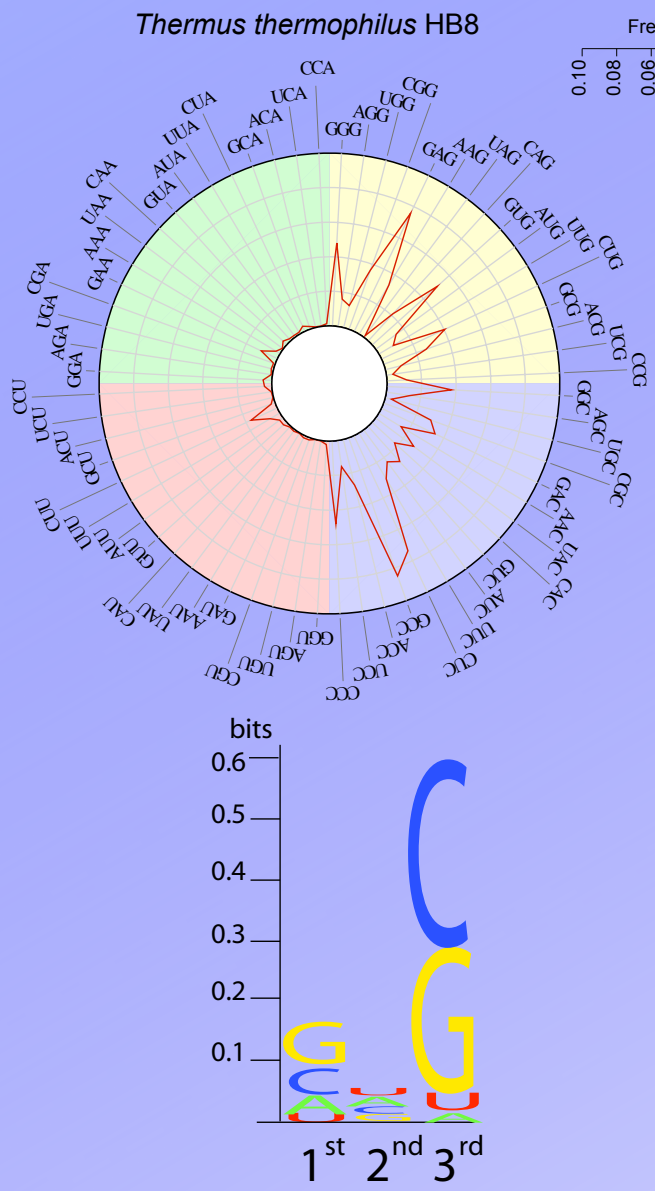


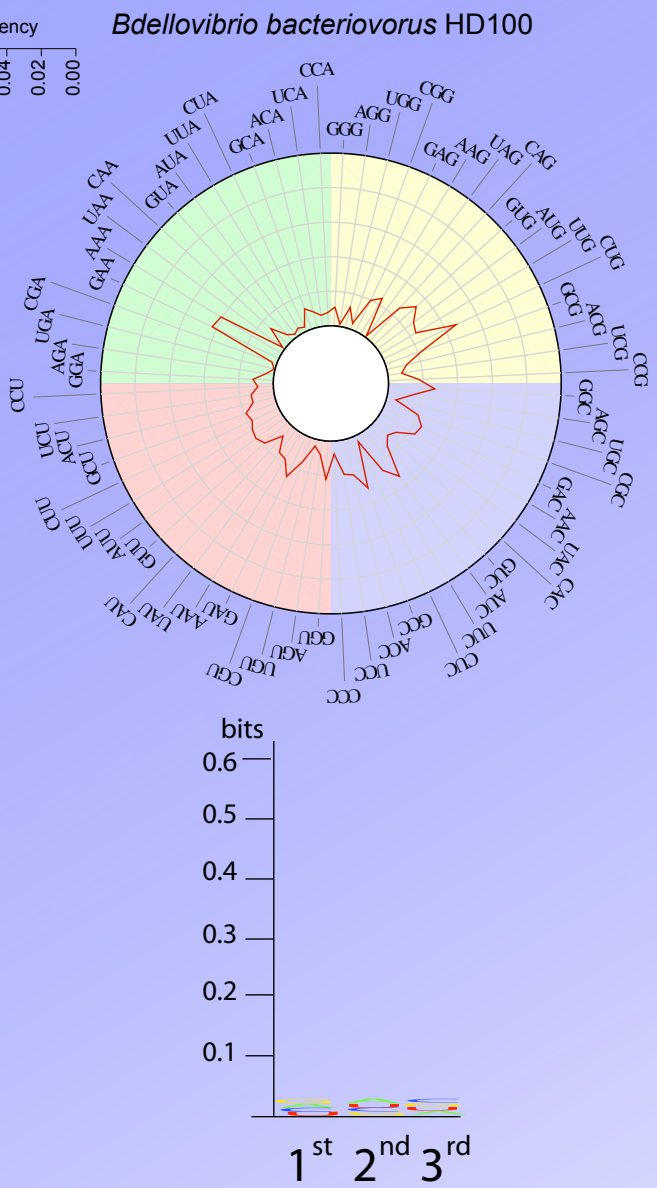
79% AT

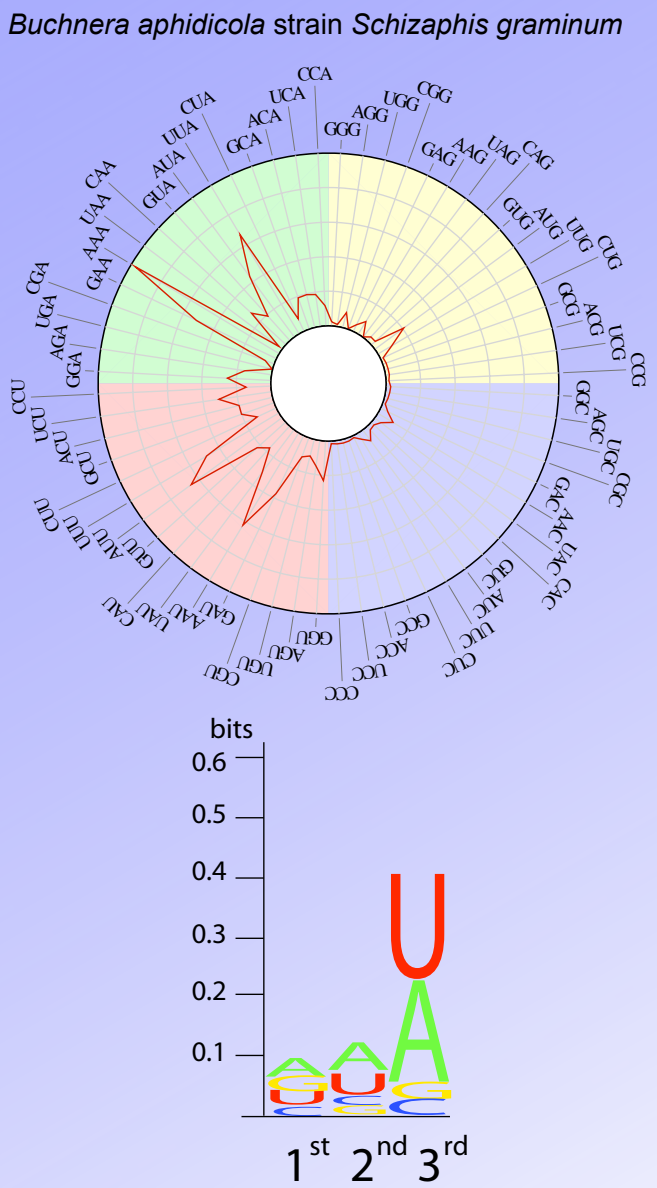




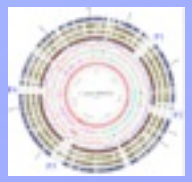
50% AT







Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark



proteins	Size (bp)	Organism	%AT	tRNA	rRNA	Accession
5,379	5,231,428 bp	Escherichia coli CFT073	49.5	89	7	AE014075
5,361	5,498,450 bp	Escherichia coli O157 RIMD	49.5	105	7	BA000007
5,349	5,528,445 bp	Escherichia coli O157 EDL	49.5	98	7	AE005174
5,066	5,065,741 bp	Escherichia coli UTI89	49.4	88	7	CP000243
4,905	5,688,987 bp	Photoghabdus luminescens	57.2	85	7	AP009048
4,685	4,938,920 bp	Escherichia coli strain 536	49.5	81	7	CP000247
4,600	4,809,037 bp	Salmonella entericia CT18	47.9	79	7	AL513382
4,492	5,064,019 bp	Erwinia carotovora	49.0	76	7	BX950851
4,468	5,082,025 bp	Escherichia coli APEC 01	49.4	103	7	CP000468
4,452	4,857,432 bp	Salmonella typhimurium LT2	47.8	85	7	AE006468
4,445	4,755,700 bp	Salmonella entericia SCB67	47.8	85	7	AE017220
4,436	4,607,203 bp	Shigella flexneri 2a301	49.1	97	7	AE005674
4,337	4,646,332 bp	Escherichia coli K-12 W3110	49.2	86	7	U00096
4,331	4,639,675 bp	Escherichia coli K-12 MG1655	49.2	86	7	AP009048
4,323	4,791,961 bp	Salmonella enterica Ty2	47.2	78	7	AE014613
4,277	4,369,232 bp	Shigella dysenteriae Sd197	48.8	85	7	CP000034
4,224	4,825,265 bp	Shigella sonnei Ss046	49.0	97	7	CP000038
4,167	4,702,289 bp	Yersinia pestis Antiqua	52.3	68	7	CP000308
4,142	4,519,823 bp	Shigella boydii Sb227	48.8	91	7	CP000036
4,116	4,574,284 bp	Shigella flexneri 5str8401	49.1	91	7	CP000266
4,093	4,585,229 bp	Salmonella entericia ATCC9150	47.8	82	7	CP000026
4,090	4,600,755 bp	Yersinia pestis KIM	52.4	73	7	AE009952
4,073	4,599,354 bp	Shigella flexneri 2457T	49.1	98	7	AE014073
4,008	4,653,728 bp	Yersinia pestis CO-92	52.4	70	6	AL590842
3,981	4,534,590 bp	Yersinia pestis Nepal516	52.4	72	7	CP000305
3,974	4,744,671 bp	Yersinia pseudotuber. IP32953	52.4	85	7	BX936398
3,895	4,595,065 bp	Yersinia pestis Mediaevails	52.3	72	7	AE017042
2,432	4,171,146 bp	Sodalis glossinidius	45.3	69	7	AP008232
611	697,724 bp	Wiggelsworthia glossinidia	77.5	34	2	BA000021
610	791,654 bp	Blochmannia pennsylvanicus	70.4	39	1	CP000016
595	686,194 bp	Baumannia cicadellinicola	61.2	39	2	CP000238
589	705,557 bp	Blochmannia floridanus	72.6	37	1	BX248583
564	640,681 bp	Buchnera aphidicola APS	73.7	32	1	BA000003
545	641,454 bp	Buchnera aphidicola Sg	74.7	32	1	AE013218
504	615,980 bp	Buchnera aphidicola BBp	74.7	32	1	AE016826
182	159,662 bp	Carsonella ruddii Pv	85.4	28	1	AP009180

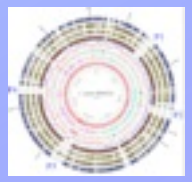
87 projects found

as of 5 August, 2009

www.cbs.dtu.dk/services/GenomeAtlas/

Row	Organism	Tax Group	NCBI Project ID	Replicons	Total Size (bp)	Number of genes	5S rRNA count	16S rRNA count	23S rRNA count	tRNA count	% AT
1	Klebsiella pneumoniae 342	BProt GE	28471	3	5,920,257	5,768	9	8	8	88	43.1
2	Salmonella enterica subsp. enterica serovar Paratyphi B str. SPB7	BProt GE	27803	1	4,858,887	5,601	8	7	7	85	47.9
3	Escherichia coli O157:H7 str. EC4115	BProt GE	27739	3	5,704,171	5,477	8	7	7	109	49.6
4	Escherichia coli O157:H7 EDL933	BProt GE	259	2	5,620,522	5,449	8	7	7	100	49.5
5	Escherichia coli O157:H7 str. Sakai	BProt GE	226	3	5,594,477	5,447	8	7	7	103	49.5
6	Escherichia coli CFT073	BProt GE	313	1	5,231,428	5,379	7	7	7	89	49.5
7	Escherichia coli UTI89	BProt GE	16259	2	5,179,971	5,211	8	7	7	89	49.4
8	Klebsiella pneumoniae subsp. pneumoniae MGH 78578	BProt GE	31	6	5,694,894	5,185	9	8	8	86	42.9
9	Escherichia coli ED1a	BProt GE	33409	1	5,209,548	5,123	8	7	7	91	49.3
10	Escherichia coli UMN026	BProt GE	33415	3	5,324,391	5,058	8	7	7	88	49.3
11	Citrobacter koseri ATCC BAA-895	BProt GE	12716	3	4,735,357	5,031	8	7	7	83	46.2
12	Klebsiella pneumoniae NTUH-K2044	BProt GE	21069	1	5,248,520	5,006	9	8	8	86	42.3
13	Escherichia coli SE11	BProt GE	18057	7	5,155,626	5,002	8	7	7	90	49.3
14	Escherichia coli E24377A	BProt GE	13960	7	5,249,288	4,997	8	7	7	91	49.4
15	Salmonella enterica subsp. enterica serovar Typhi str. CT18	BProt GE	236	3	5,133,713	4,980	8	7	7	79	48.1
16	Serratia proteamaculans 568	BProt GE	17459	2	5,495,657	4,942	8	7	7	85	45.0
17	Escherichia coli 55989	BProt GE	33413	1	5,154,862	4,919	8	7	7	94	49.3
18	Escherichia coli SMS-3-5	BProt GE	19469	5	5,215,377	4,913	8	7	7	90	49.5
19	Photorhabdus luminescens subsp. laumondii TTO1	BProt GE	9605	1	5,688,987	4,905	8	7	7	85	57.2
20	Escherichia coli IAI39	BProt GE	33411	1	5,132,068	4,892	8	7	7	88	49.4
21	Escherichia coli APEC O1	BProt GE	16718	3	5,497,653	4,890	8	7	7	96	49.7
22	Escherichia coli O127:H6 str. E2348/69	BProt GE	32571	3	5,069,678	4,824	8	7	7	92	49.5
23	Salmonella enterica subsp. enterica serovar Newport str. SL254	BProt GE	18747	3	5,007,719	4,805	8	7	7	85	47.8
24	Salmonella enterica subsp. enterica serovar Heidelberg str. SL476	BProt GE	20045	3	4,983,515	4,779	8	7	7	85	47.9
25	Shigella flexneri 2a str. 301	BProt GE	310	2	4,828,821	4,703	8	7	7	97	49.3
26	Escherichia coli 536	BProt GE	16235	1	4,938,920	4,685	8	7	7	81	49.5
27	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	BProt GE	9618	3	4,944,000	4,666	8	7	7	85	47.9
28	Salmonella enterica subsp. enterica serovar Paratyphi C strain RKS4594	BProt GE	20993	2	4,888,494	4,640	8	7	7	83	47.8
29	Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633	BProt GE	19459	3	4,823,887	4,627	8	7	7	83	47.8
30	Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853	BProt GE	19467	2	4,917,459	4,617	8	7	7	82	47.9
31	Salmonella enterica subsp. enterica serovar Agona str. SL483	BProt GE	20063	2	4,836,638	4,614	8	7	7	84	48.0
32	Shigella boydii CDC 3083-94	BProt GE	15637	6	4,874,659	4,557	8	7	7	99	49.0
33	Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	BProt GE	241	2	4,951,371	4,554	8	7	7	86	47.8
34	Salmonella enterica subsp. arizonae serovar 62:z4,z23:--	BProt GE	13030	1	4,600,800	4,510	8	7	7	85	48.6
35	Shigella dysenteriae Sd197	BProt GE	13145	3	4,560,911	4,508	8	7	7	85	49.0

41	Escherichia coli HS	BProt GE	13959	1	4,643,538	4,384	8	7	7	88	49.2
42	Escherichia fergusonii ATCC 35469	BProt GE	33369	2	4,643,861	4,377	8	7	7	87	50.1
43	Yersinia pestis Antiqua	BProt GE	16645	4	4,879,836	4,364	8	7	7	68	52.3
44	Escherichia coli str. K-12 substr. W3110	BProt GE	16351	1	4,646,332	4,337	8	7	7	88	49.2
45	Yersinia pseudotuberculosis IP 31758	BProt GE	16070	3	4,935,125	4,324	8	7	7	86	52.8
46	Salmonella enterica subsp. enterica serovar Typhi str. Ty2	BProt GE	371	1	4,791,961	4,323	8	7	7	78	47.9
47	Salmonella enterica subsp. enterica serovar Enteritidis str. P125109	BProt GE	30687	1	4,685,848	4,318	8	7	7	83	47.8
48	Shigella boydii Sb227	BProt GE	13146	2	4,646,520	4,290	8	7	7	91	48.9
49	Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601	BProt GE	30943	1	4,581,797	4,284	8	7	7	81	47.8
50	Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91	BProt GE	30689	1	4,658,697	4,274	8	7	7	75	47.8
51	Escherichia coli str. K-12 substr. MG1655	BProt GE	225	1	4,639,675	4,267	8	7	7	88	49.2
52	Pectobacterium carotovorum subsp. carotovorum PC1	BProt GE	31289	1	4,862,913	4,246	8	7	7	78	48.1
53	Escherichia coli BL21	BProt GE	28965	1	4,557,041	4,244	8	7	7	82	49.2
54	Enterobacter sp. 638	BProt GE	17461	2	4,676,461	4,240	8	7	7	84	47.1
55	Yersinia pseudotuberculosis PB1/+	BProt GE	28745	2	4,765,431	4,237	8	7	7	83	52.5
56	Escherichia coli BL21(DE3)	BProt GE	30681	1	4,570,938	4,228	8	7	7	86	49.2
57	Yersinia pestis CO92	BProt GE	34	4	4,829,855	4,217	7	6	6	70	52.4
58	Yersinia pestis KIM	BProt GE	288	2	4,701,745	4,205	8	7	7	73	52.3
59	Escherichia coli ATCC 8739	BProt GE	18083	1	4,746,218	4,200	8	7	7	87	49.1
60	Escherichia coli str. K-12 substr. DH10B	BProt GE	20079	1	4,686,137	4,200	8	7	7	87	49.2
61	Yersinia pseudotuberculosis YPIII	BProt GE	28743	1	4,689,441	4,192	8	7	7	84	52.5
62	Dickeya zeae Ech1591	BProt GE	31295	1	4,813,854	4,163	8	7	7	74	45.5
63	Yersinia pestis biovar Microtus str. 91001	BProt GE	10638	5	4,803,217	4,142	8	7	7	72	52.3
64	Yersinia enterocolitica subsp. enterocolitica 8081	BProt GE	190	2	4,683,620	4,137	8	7	7	81	52.8
65	Yersinia pseudotuberculosis IP 32953	BProt GE	12950	3	4,840,898	4,116	8	7	7	85	52.4
66	Shigella flexneri 5 str. 8401	BProt GE	16375	1	4,574,284	4,116	8	7	7	97	49.1
67	Yersinia pestis Nepal516	BProt GE	16646	3	4,646,286	4,094	8	7	7	72	52.4
68	Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150	BProt GE	13086	1	4,585,229	4,093	8	7	7	82	47.8
69	Escherichia coli BW2952	BProt GE	33775	1	4,578,159	4,084	8	7	7	87	49.2
70	Shigella flexneri 2a str. 2457T	BProt GE	408	1	4,599,354	4,073	8	7	7	100	49.1
71	Yersinia pestis Pestoides F	BProt GE	16700	3	4,725,862	4,069	8	7	7	72	52.3
72	Yersinia pestis Angola	BProt GE	16067	3	4,687,014	4,045	8	7	7	70	52.4
73	Dickeya dadantii Ech703	BProt GE	33069	1	4,679,450	3,970	8	7	7	74	45.0
74	Edwardsiella ictaluri 93-146	BProt GE	34853	1	3,812,315	3,784	9	8	8	94	42.6
75	Proteus mirabilis HI4320	BProt GE	12624	2	4,099,895	3,740	8	7	7	83	61.1
76	Erwinia tasmaniensis Et1/99	BProt GE	20585	6	4,067,864	3,696	8	7	7	81	46.6
77	Sodalis glossinidius str. morsitans	BProt GE	16309	4	4,292,502	2,516	8	7	7	71	45.5
78	Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)	BProt GE	31259	2	2,169,363	2,155	3	3	3	43	59.5
79	Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	BProt GE	274	2	703,004	611	2	2	2	34	77.5
80	Candidatus Blochmannia pennsylvanicus str. BPEN	BProt GE	13875	1	791,654	610	1	1	1	40	70.4
81	Candidatus Blochmannia floridanus	BProt GE	443	1	705,557	589	1	1	1	37	72.6
82	Buchnera aphidicola str. APS (Acyrthosiphon pisum)	BProt GE	245	3	655,725	574	1	1	1	32	73.6
83	Buchnera aphidicola str. 5A (Acyrthosiphon pisum)	BProt GE	31225	1	642,122	555	1	1	1	32	73.7
84	Buchnera aphidicola str. Tuc7 (Acyrthosiphon pisum)	BProt GE	31223	1	641,895	553	1	1	1	32	73.7
85	Buchnera aphidicola str. Sg (Schizaphis graminum)	BProt GE	312	1	641,454	545	1	1	1	32	74.7
86	Buchnera aphidicola str. Bp (Baizongia pistaciae)	BProt GE	256	2	618,379	507	1	1	1	32	74.7
87	Buchnera aphidicola str. Cc (Cinara cedri)	BProt GE	16372	2	422,434	365	1	1	1	31	79.8

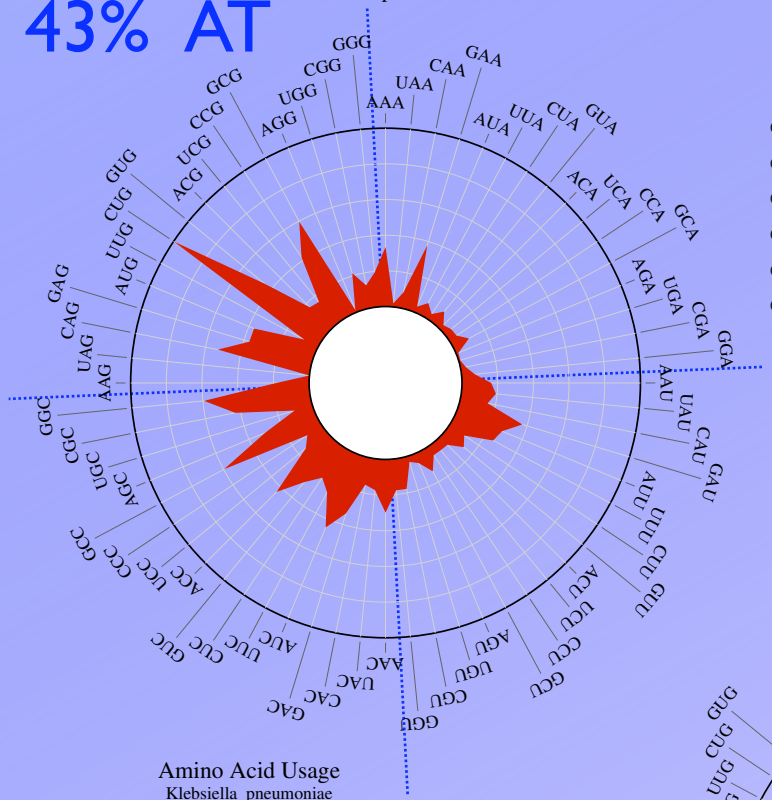


CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

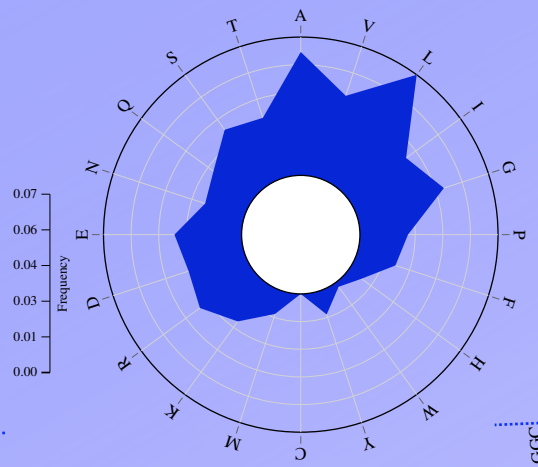
Center for Biological Sequence analysis | Department of Systems Biology, Technical University of Denmark

Codon Usage
Klebsiella_pneumoniae

43% AT

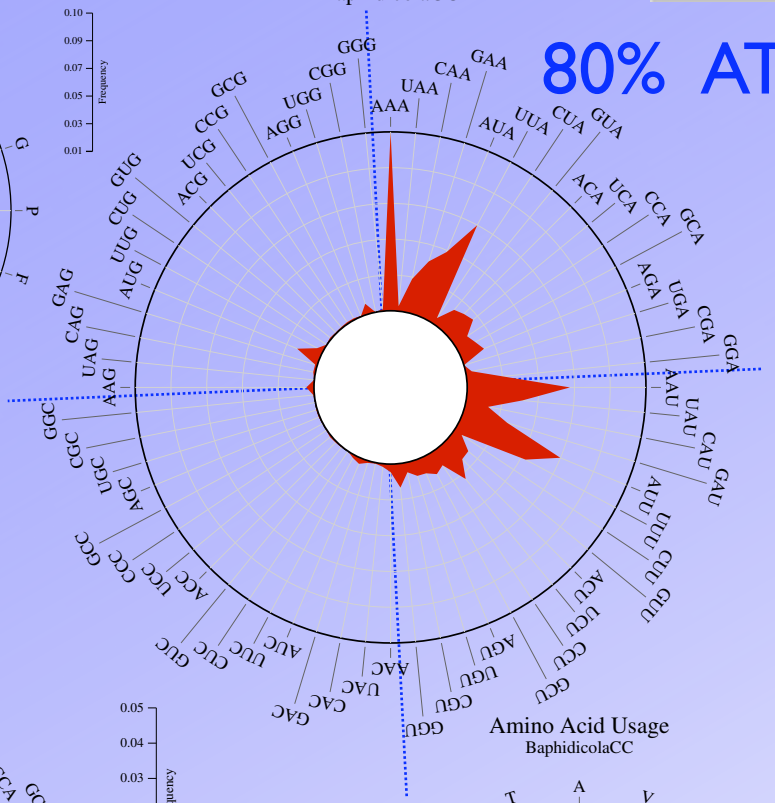


Amino Acid Usage
ecoli_APEC01

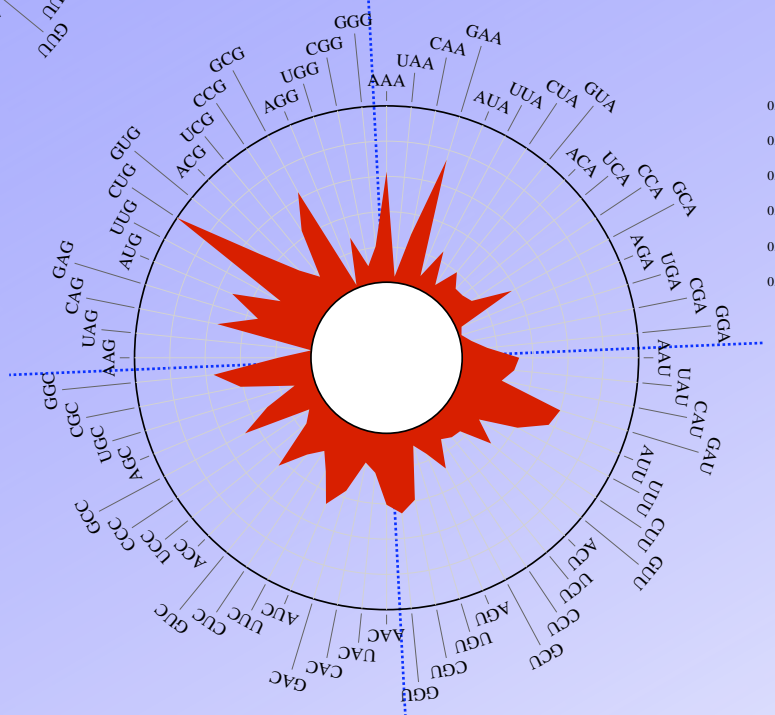


Codon Usage
BaphidicolaCC

80% AT

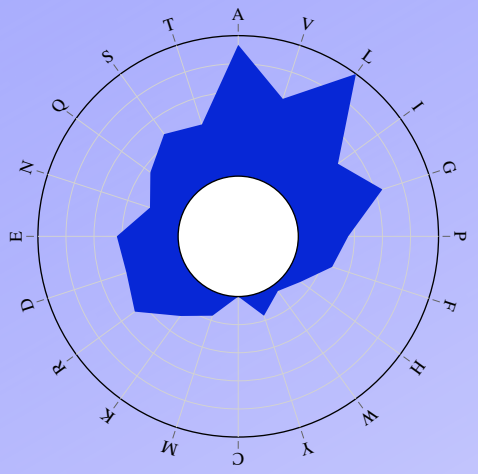


Codon Usage
ecoli_ATCC8739

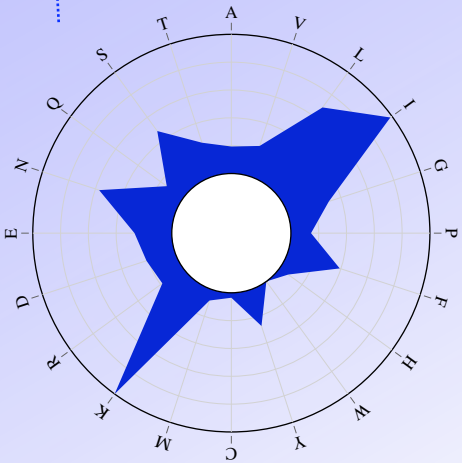


49% AT

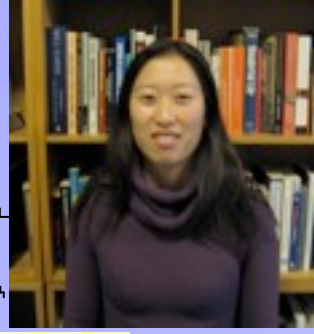
Amino Acid Usage
Klebsiella_pneumoniae



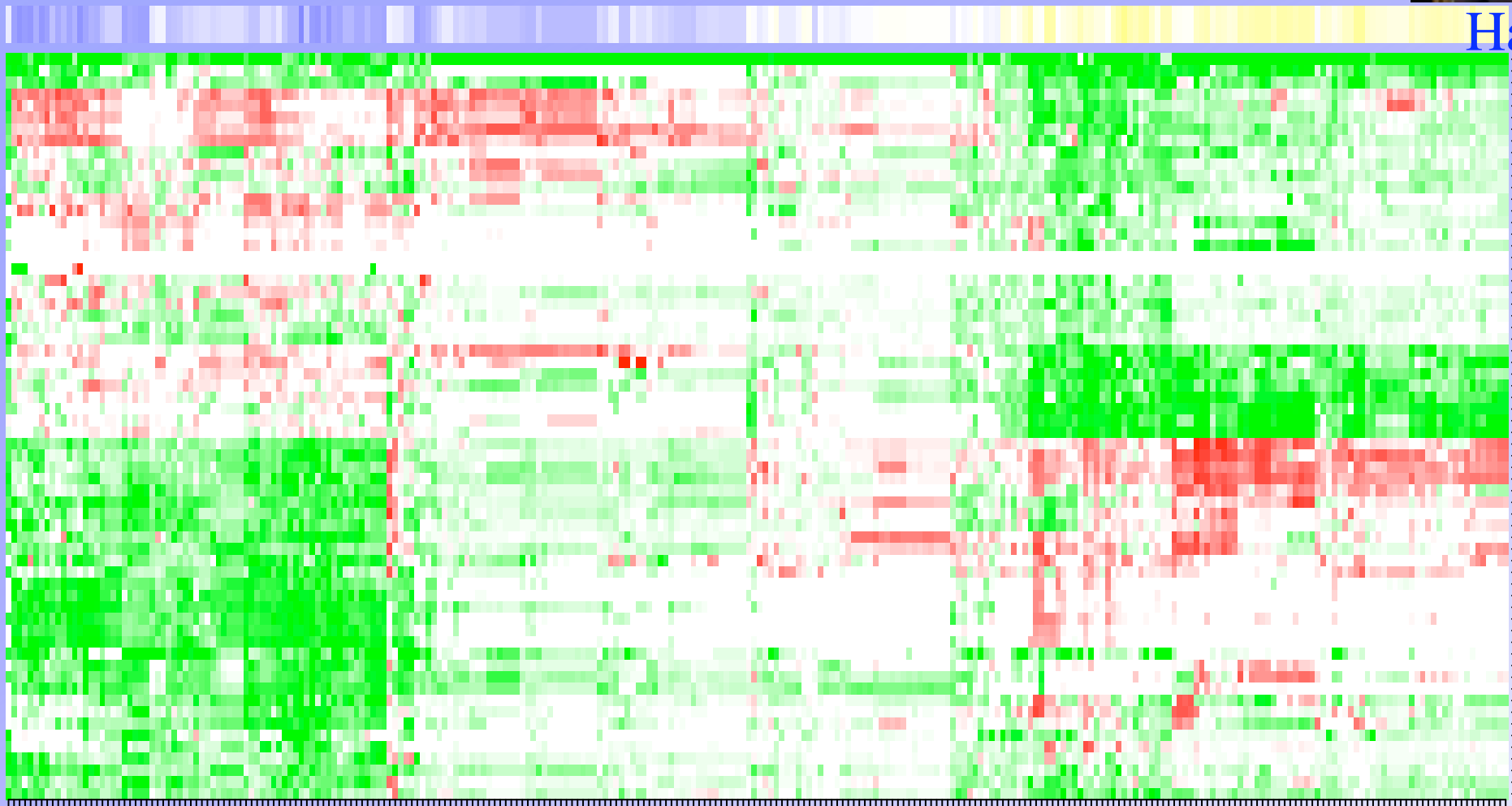
Amino Acid Usage
BaphidicolaCC



Microbiology, **150**:1603-1606, (2004).



Hanni



GenomeBiology, 7:R114, (2006). Highly accessed Open Access