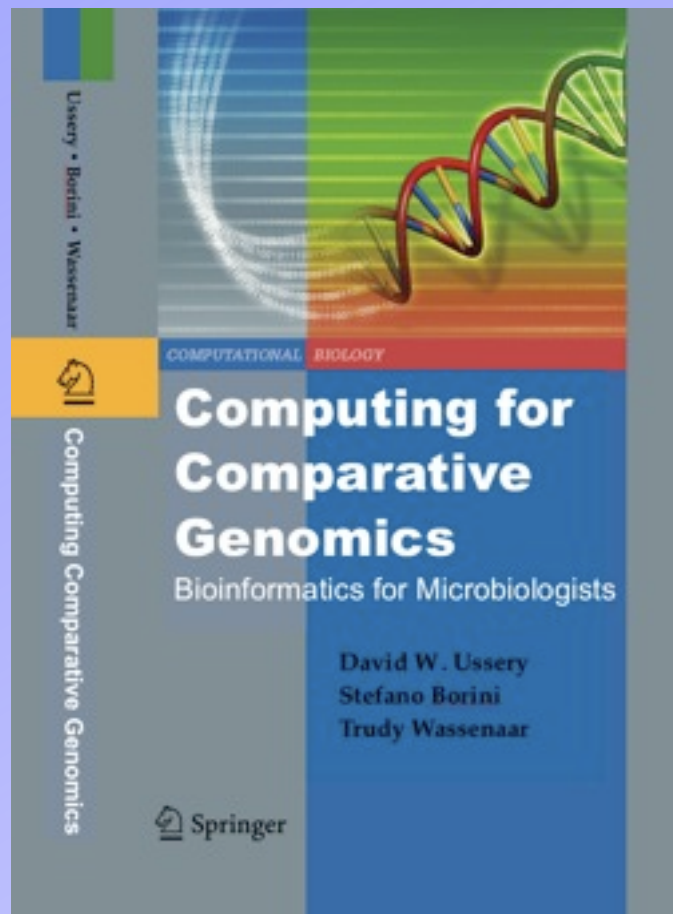


Pan-Genomics and Metagenomics



~~Dave Ussery
talk at EBI
Hinxton, England
Thursday, 27 August, 2009~~

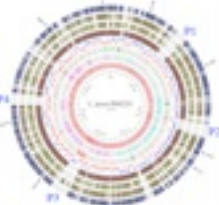
MBV-INF4410 – Bioinformatics for Molecular Biology
Second talk

Friday, 11 September, 2009



Outline

- **Genome Diversity is large!**
- **Using the Pan-genome to make trees.**
- **A few words about annotation....**



CENTER FOR BIOLOGICAL SEQUENCE ANALYSIS ■ TECHNICAL UNIVERSITY OF DENMARK DTU

EVENTS	RESEARCH GROUPS	CBS DATA SETS	PUBLICATIONS
NEWS	CBS PREDICTION SERVERS		BIOINFORMATICS EDUCATION PROGRAM
STAFF	ABOUT CBS	CBS BIOINFORMATICS TOOLS	OTHER BIOINFORMATICS LINKS
CONTACT	INTERNAL	CBS COURSES	

powered by sgi

www.cbs.dtu.dk

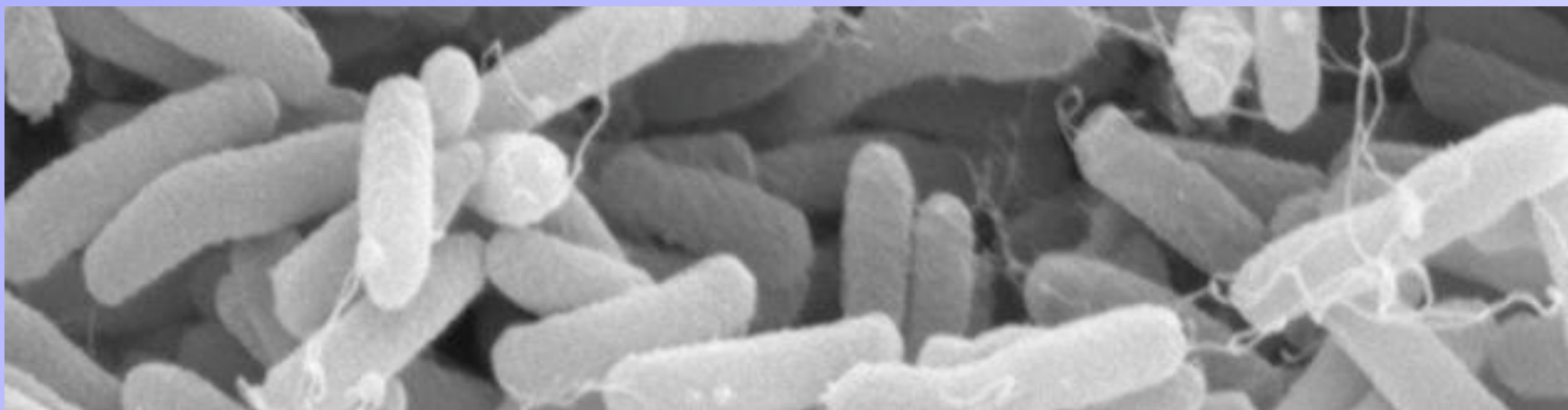
News and Views

[Nature](#) 412, 597-598 (9 August 2001) | doi:10.1038/35088167

Genome sequencing: The ABC of symbiosis

J. Allan Downie and J. Peter W. Young

It is a truth universally acknowledged, that there are only two kinds of bacteria. One is *Escherichia coli* and the other is not. Anything that *E. coli* does is a universal truth about bacteria; anything it does not do must be a specialization.



When are two proteins the same??

50% length of query
←→

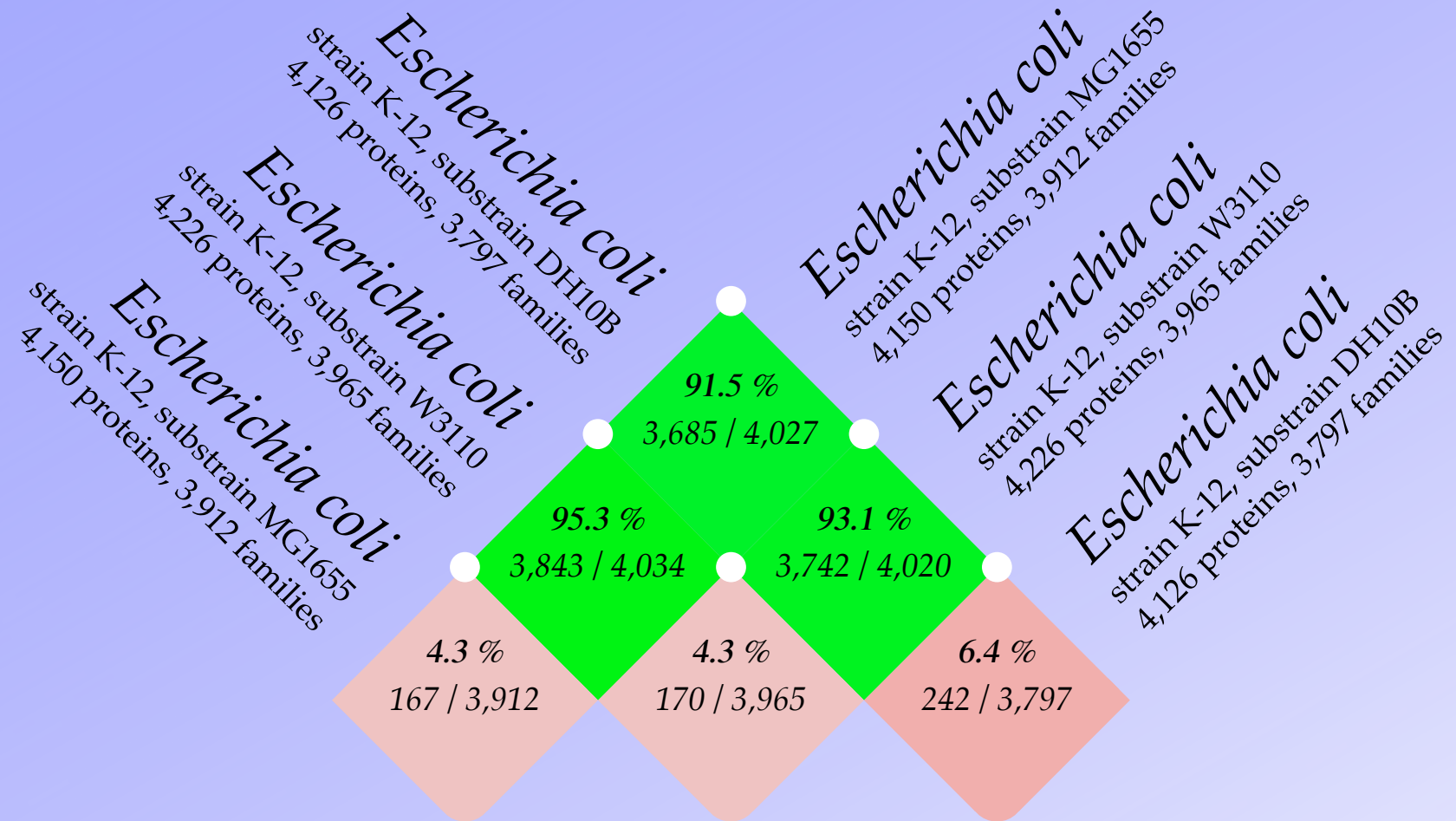
Query sequence (protein)

Subject sequence (protein)

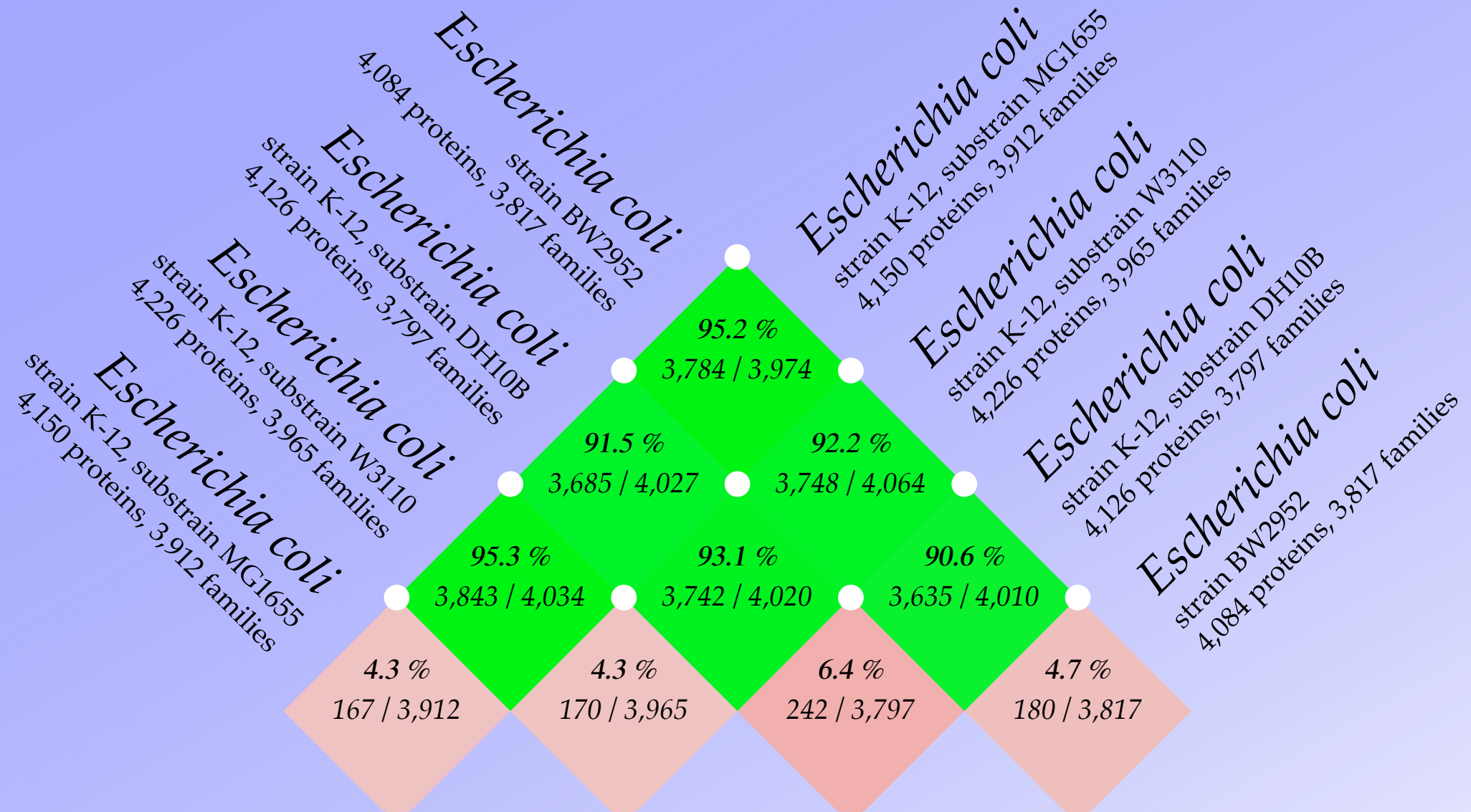
50% identity of match

Disclaimer - the "50-50 rule"

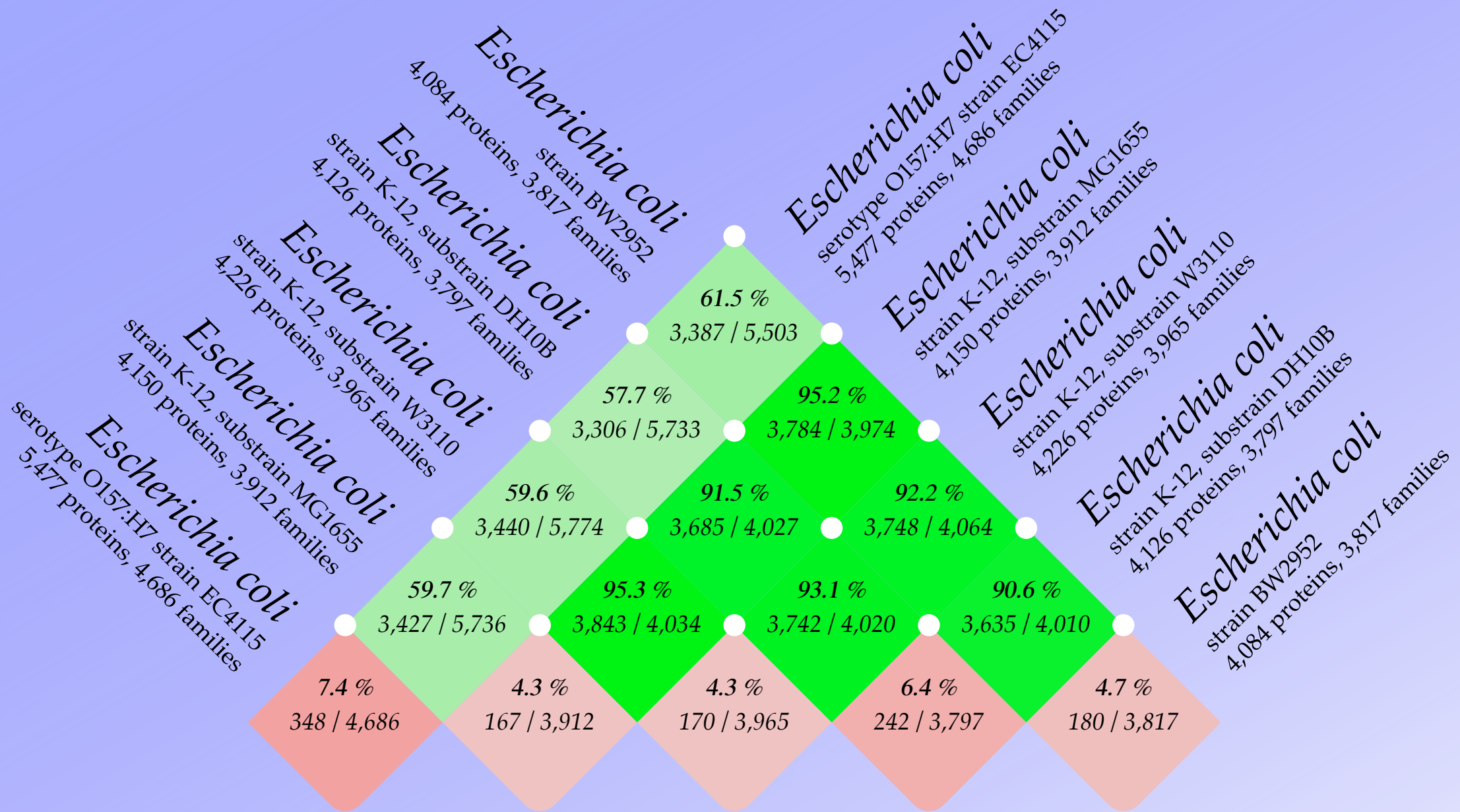
3 *E. coli* K-12 genomes



4 *E. coli* genomes



5 *E. coli* genomes

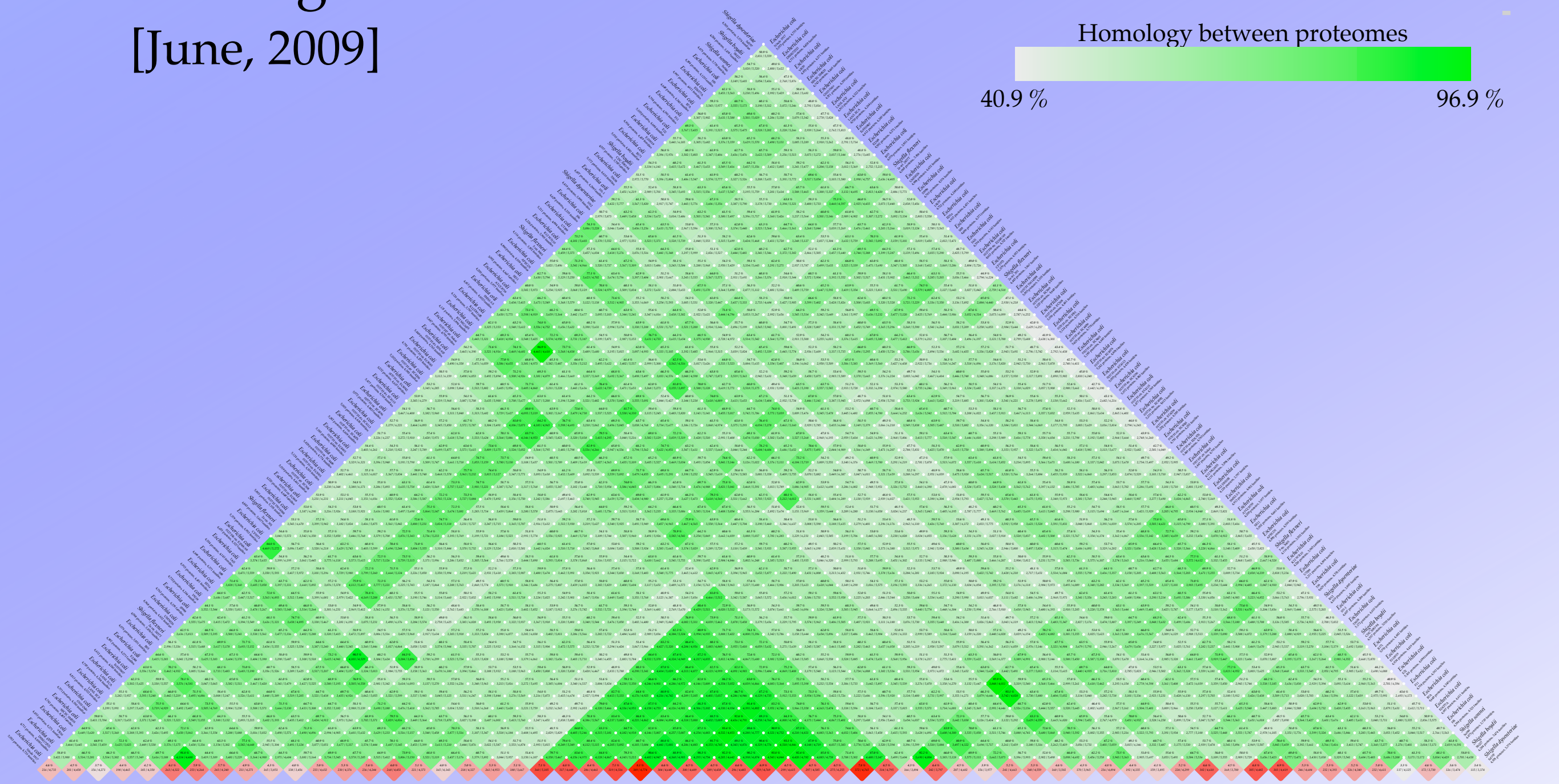


60 *E. coli* genomes [June, 2009]

Homology between proteomes

40.9%

96.9%



Homology within proteomes

3.3%

8.1%

87 projects found

as of 5 August, 2009



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	Photobacterium luminescens subsp. laumondii TTO1	BProt GE	9605	1	5,688,987	4,905	8	7	7	85	57.2
20	Escherichia coli IA139	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

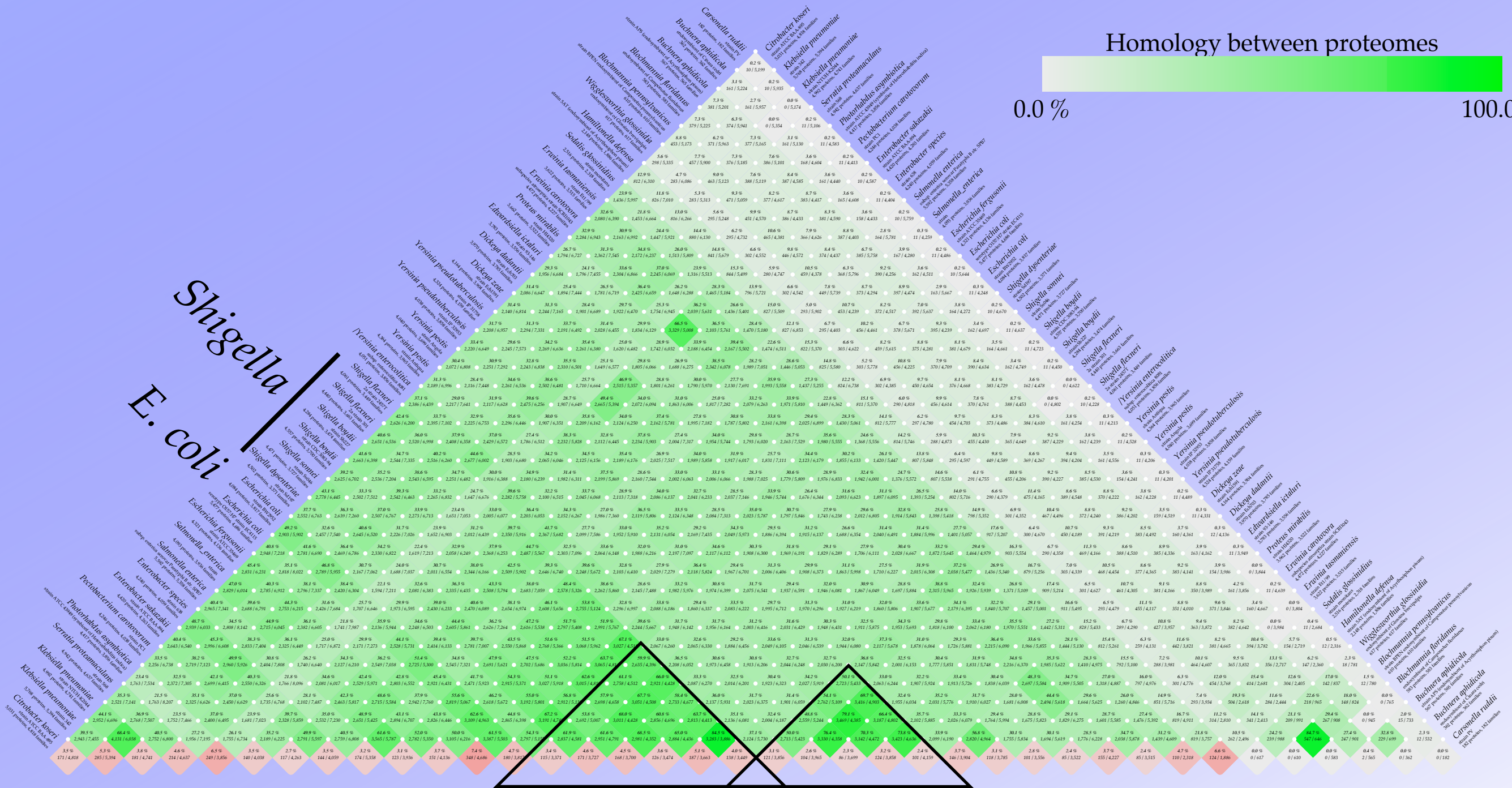
41	<u>Escherichia coli HS</u>	<u>BProt GE</u>	<u>13959</u>	1	4,643,538	4,384	8	7	7	88	49.2
42	<u>Escherichia fergusonii ATCC 35469</u>	<u>BProt GE</u>	<u>33369</u>	2	4,643,861	4,377	8	7	7	87	50.1
43	<u>Yersinia pestis Antiqua</u>	<u>BProt GE</u>	<u>16645</u>	4	4,879,836	4,364	8	7	7	68	52.3
44	<u>Escherichia coli str. K-12 substr. W3110</u>	<u>BProt GE</u>	<u>16351</u>	1	4,646,332	4,337	8	7	7	88	49.2
45	<u>Yersinia pseudotuberculosis IP 31758</u>	<u>BProt GE</u>	<u>16070</u>	3	4,935,125	4,324	8	7	7	86	52.8
46	<u>Salmonella enterica subsp. enterica serovar Typhi str. Ty2</u>	<u>BProt GE</u>	<u>371</u>	1	4,791,961	4,323	8	7	7	78	47.9
47	<u>Salmonella enterica subsp. enterica serovar Enteritidis str. P125109</u>	<u>BProt GE</u>	<u>30687</u>	1	4,685,848	4,318	8	7	7	83	47.8
48	<u>Shigella boydii Sb227</u>	<u>BProt GE</u>	<u>13146</u>	2	4,646,520	4,290	8	7	7	91	48.9
49	<u>Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU 12601</u>	<u>BProt GE</u>	<u>30943</u>	1	4,581,797	4,284	8	7	7	81	47.8
50	<u>Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91</u>	<u>BProt GE</u>	<u>30689</u>	1	4,658,697	4,274	8	7	7	75	47.8
51	<u>Escherichia coli str. K-12 substr. MG1655</u>	<u>BProt GE</u>	<u>225</u>	1	4,639,675	4,267	8	7	7	88	49.2
52	<u>Pectobacterium carotovorum subsp. carotovorum PC1</u>	<u>BProt GE</u>	<u>31289</u>	1	4,862,913	4,246	8	7	7	78	48.1
53	<u>Escherichia coli BL21</u>	<u>BProt GE</u>	<u>28965</u>	1	4,557,041	4,244	8	7	7	82	49.2
54	<u>Enterobacter sp. 638</u>	<u>BProt GE</u>	<u>17461</u>	2	4,676,461	4,240	8	7	7	84	47.1
55	<u>Yersinia pseudotuberculosis PB1/+</u>	<u>BProt GE</u>	<u>28745</u>	2	4,765,431	4,237	8	7	7	83	52.5
56	<u>Escherichia coli BL21(DE3)</u>	<u>BProt GE</u>	<u>30681</u>	1	4,570,938	4,228	8	7	7	86	49.2
57	<u>Yersinia pestis CO92</u>	<u>BProt GE</u>	<u>34</u>	4	4,829,855	4,217	7	6	6	70	52.4
58	<u>Yersinia pestis KIM</u>	<u>BProt GE</u>	<u>288</u>	2	4,701,745	4,205	8	7	7	73	52.3
59	<u>Escherichia coli ATCC 8739</u>	<u>BProt GE</u>	<u>18083</u>	1	4,746,218	4,200	8	7	7	87	49.1
60	<u>Escherichia coli str. K-12 substr. DH10B</u>	<u>BProt GE</u>	<u>20079</u>	1	4,686,137	4,200	8	7	7	87	49.2
61	<u>Yersinia pseudotuberculosis YPIII</u>	<u>BProt GE</u>	<u>28743</u>	1	4,689,441	4,192	8	7	7	84	52.5
62	<u>Dickeya zeae Ech1591</u>	<u>BProt GE</u>	<u>31295</u>	1	4,813,854	4,163	8	7	7	74	45.5
63	<u>Yersinia pestis biovar Microtus str. 91001</u>	<u>BProt GE</u>	<u>10638</u>	5	4,803,217	4,142	8	7	7	72	52.3
64	<u>Yersinia enterocolitica subsp. enterocolitica 8081</u>	<u>BProt GE</u>	<u>190</u>	2	4,683,620	4,137	8	7	7	81	52.8
65	<u>Yersinia pseudotuberculosis IP 32953</u>	<u>BProt GE</u>	<u>12950</u>	3	4,840,898	4,116	8	7	7	85	52.4
66	<u>Shigella flexneri 5 str. 8401</u>	<u>BProt GE</u>	<u>16375</u>	1	4,574,284	4,116	8	7	7	97	49.1
67	<u>Yersinia pestis Nepal516</u>	<u>BProt GE</u>	<u>16646</u>	3	4,646,286	4,094	8	7	7	72	52.4
68	<u>Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150</u>	<u>BProt GE</u>	<u>13086</u>	1	4,585,229	4,093	8	7	7	82	47.8
69	<u>Escherichia coli BW2952</u>	<u>BProt GE</u>	<u>33775</u>	1	4,578,159	4,084	8	7	7	87	49.2
70	<u>Shigella flexneri 2a str. 2457T</u>	<u>BProt GE</u>	<u>408</u>	1	4,599,354	4,073	8	7	7	100	49.1
71	<u>Yersinia pestis Pestoides F</u>	<u>BProt GE</u>	<u>16700</u>	3	4,725,862	4,069	8	7	7	72	52.3
72	<u>Yersinia pestis Angola</u>	<u>BProt GE</u>	<u>16067</u>	3	4,687,014	4,045	8	7	7	70	52.4
73	<u>Dickeya dadantii Ech703</u>	<u>BProt GE</u>	<u>33069</u>	1	4,679,450	3,970	8	7	7	74	45.0
74	<u>Edwardsiella ictaluri 93-146</u>	<u>BProt GE</u>	<u>34853</u>	1	3,812,315	3,784	9	8	8	94	42.6
75	<u>Proteus mirabilis HI4320</u>	<u>BProt GE</u>	<u>12624</u>	2	4,099,895	3,740	8	7	7	83	61.1
76	<u>Erwinia tasmaniensis Et1/99</u>	<u>BProt GE</u>	<u>20585</u>	6	4,067,864	3,696	8	7	7	81	46.6
77	<u>Sodalis glossinidius str. morsitans</u>	<u>BProt GE</u>	<u>16309</u>	4	4,292,502	2,516	8	7	7	71	45.5
78	<u>Candidatus Hamiltonella defensa 5AT (Acyrtosiphon pisum)</u>	<u>BProt GE</u>	<u>31259</u>	2	2,169,363	2,155	3	3	3	43	59.5
79	<u>Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis</u>	<u>BProt GE</u>	<u>274</u>	2	703,004	611	2	2	2	34	77.5
80	<u>Candidatus Blochmannia pennsylvanicus str. BPEN</u>	<u>BProt GE</u>	<u>13875</u>	1	791,654	610	1	1	1	40	70.4
81	<u>Candidatus Blochmannia floridanus</u>	<u>BProt GE</u>	<u>443</u>	1	705,557	589	1	1	1	37	72.6
82	<u>Buchnera aphidicola str. APS (Acyrtosiphon pisum)</u>	<u>BProt GE</u>	<u>245</u>	3	655,725	574	1	1	1	32	73.6
83	<u>Buchnera aphidicola str. 5A (Acyrtosiphon pisum)</u>	<u>BProt GE</u>	<u>31225</u>	1	642,122	555	1	1	1	32	73.7
84	<u>Buchnera aphidicola str. Tuc7 (Acyrtosiphon pisum)</u>	<u>BProt GE</u>	<u>31223</u>	1	641,895	553	1	1	1	32	73.7
85	<u>Buchnera aphidicola str. Sg (Schizaphis graminum)</u>	<u>BProt GE</u>	<u>312</u>	1	641,454	545	1	1	1	32	74.7
86	<u>Buchnera aphidicola str. Bp (Baizongia pistaciae)</u>	<u>BProt GE</u>	<u>256</u>	2	618,379	507	1	1	1	32	74.7
87	<u>Buchnera aphidicola str. Cc (Cinara cedri)</u>	<u>BProt GE</u>	<u>16372</u>	2	422,434	365	1	1	1	31	79.8

40 enteric genomes

Homology between proteomes

0.0 %

100.0 %



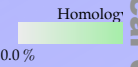
Shigella

Yersinia

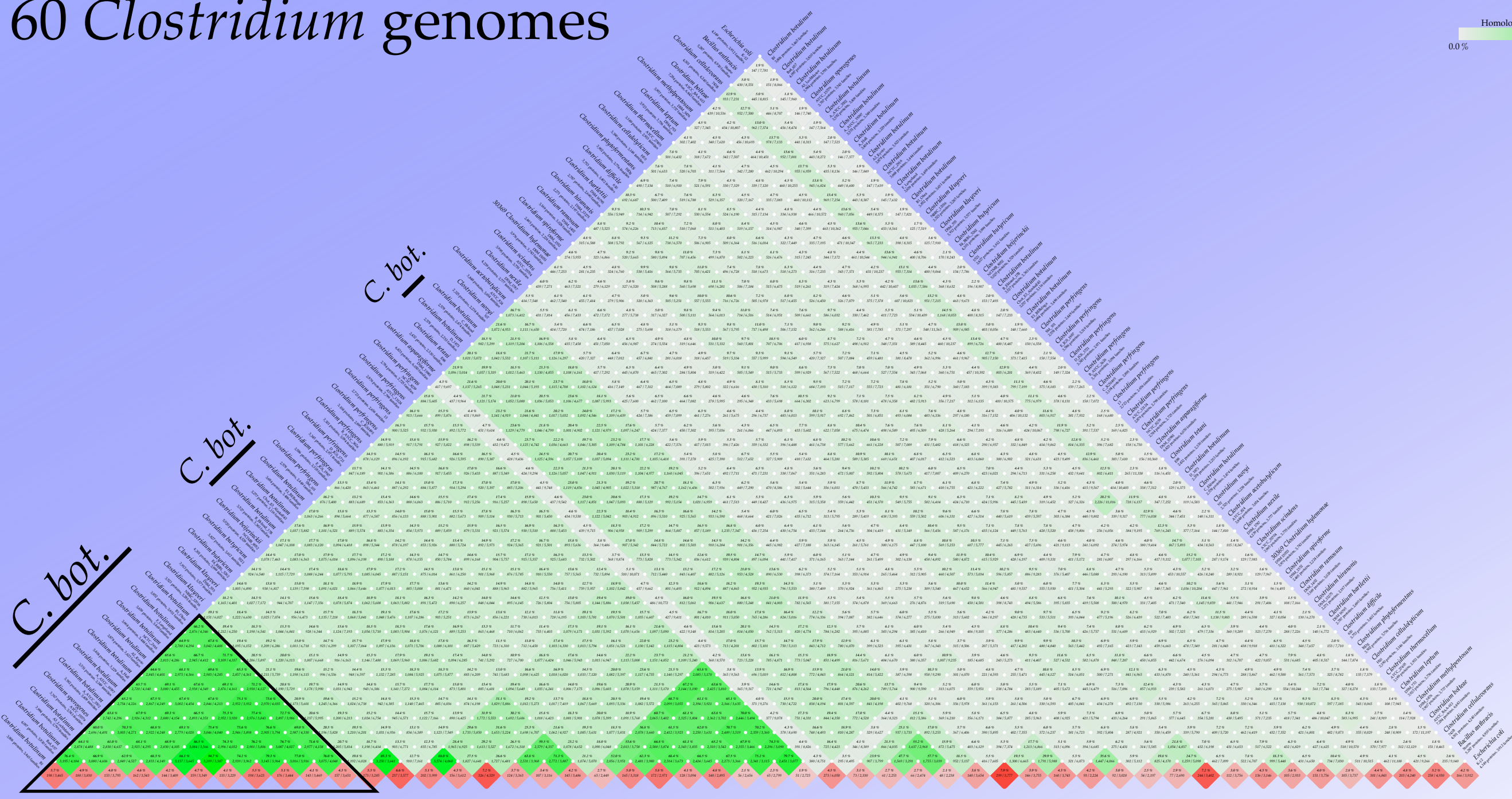
Homology within proteomes

0.0 %

20.0 %



60 Clostridium genomes



C. botulinium

Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"

Hervé Tettelin^{a,b}, Vega Masignani^{b,c}, Michael J. Cieslewicz^{b,d,e}, Claudio Donati^c, Duccio Medini^c, Naomi L. Ward^{a,f}, Samuel V. Angiuoli^a, Jonathan Crabtree^a, Amanda L. Jones^g, A. Scott Durkin^a, Robert T. DeBoy^a, Tanja M. Davidsen^a, Marirosa Mora^c, Maria Scarselli^c, Immaculada Margarit y Ros^c, Jeremy D. Peterson^a, Christopher R. Hauser^a, Jaideep P. Sundaram^a, William C. Nelson^a, Ramana Madupu^a, Lauren M. Brinkac^a, Robert J. Dodson^a, Mary J. Rosovitz^a, Steven A. Sullivan^a, Sean C. Daugherty^a, Daniel H. Haft^a, Jeremy Selengut^a, Michelle L. Gwinn^a, Liwei Zhou^a, Nikhat Zafar^a, Hoda Khouri^a, Diana Radune^a, George Dimitrov^a, Kisha Watkins^a, Kevin J. B. O'Connor^h, Shannon Smithⁱ, Teresa R. Utterbackⁱ, Owen White^a, Craig E. Rubens^g, Guido Grandi^c, Lawrence C. Madoff^{e,j}, Dennis L. Kasper^{e,j}, John L. Telford^c, Michael R. Wessels^{d,e}, Rino Rappuoli^{c,k,l}, and Claire M. Fraser^{a,b,k,m}

^aInstitute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850; ^cChiron Vaccines, Via Fiorentina 1, 53100 Siena, Italy; ^dDivision of Infectious Diseases, Children's Hospital, 300 Longwood Avenue, Boston, MA 02115; ^eHarvard Medical School, Boston, MA 02115; ^fCenter of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 East Pratt Street, Baltimore, MD 21202; ^gChildren's Hospital and Regional Medical Center, 307 Westlake Avenue N, Seattle, WA 98109; ^hThe Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218; ⁱCraig Venter Institute, 5 Research Place, Rockville, MD 20850; ^jChanning Laboratory, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115; and ^mGeorge Washington University Medical Center, 2300 Eye Street NW, Washington, DC 20037

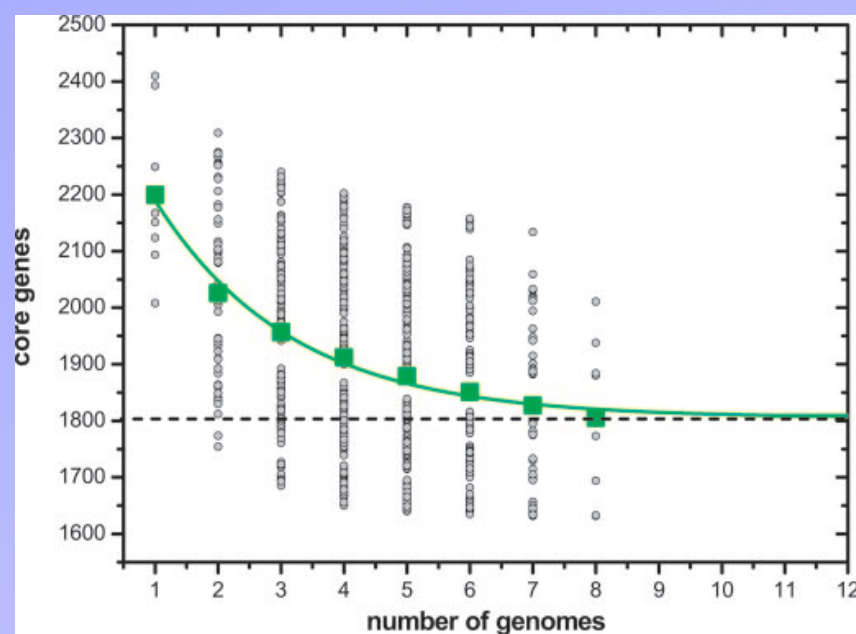


Fig. 2. GBS core genome. The number of shared genes is plotted as a function of the number n of strains sequentially added (see *Materials and Methods*). For each n , circles are the $8! / [(n-1)! \cdot (8-n)!]$ values obtained for the different strain combinations. Squares are the average of each value. The

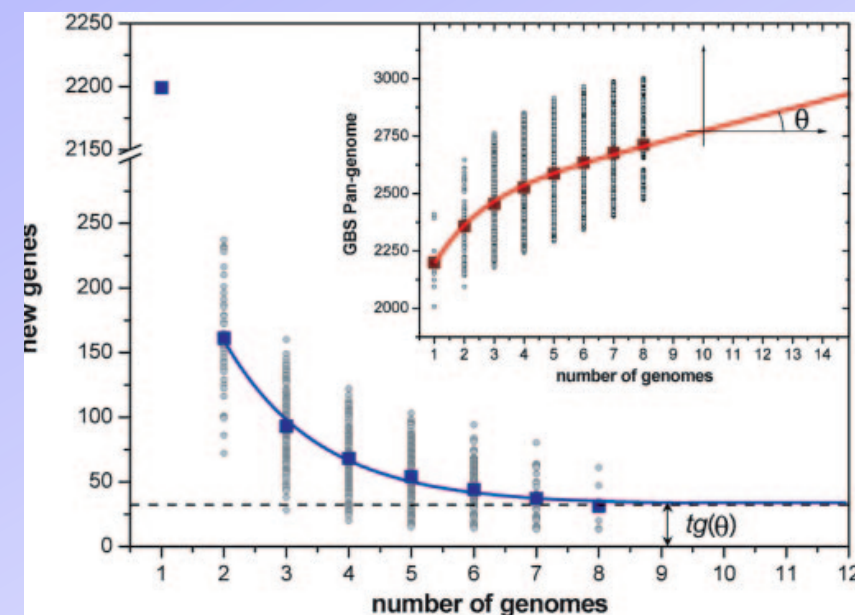
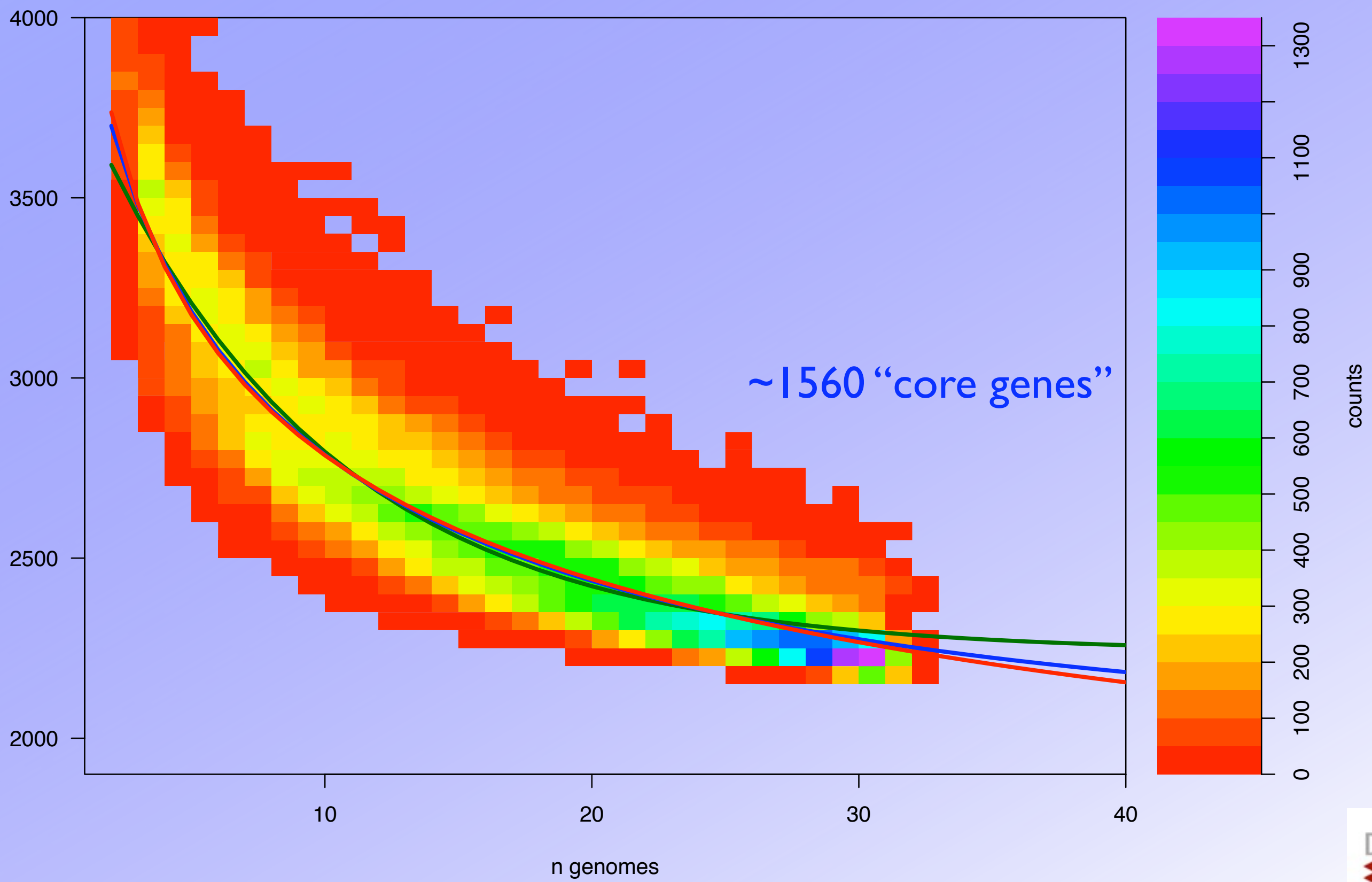
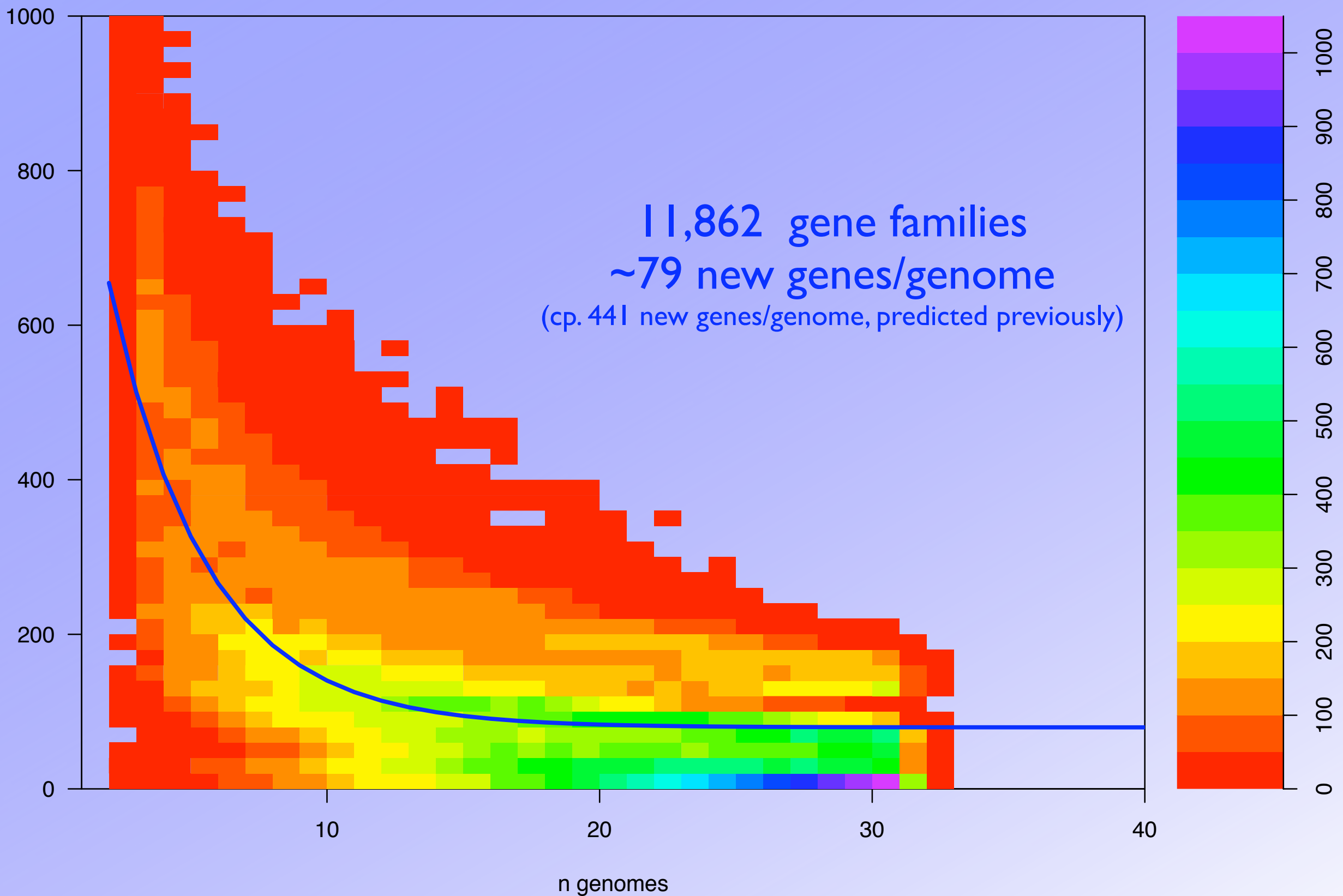


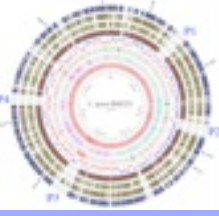
Fig. 3. GBS pan-genome. The number of specific genes is plotted as a function of the number n of strains sequentially added (see *Materials and Methods*). For each n , circles are the $8! / [(n-1)! \cdot (8-n)!]$ values obtained for the different strain combinations. Squares are the average of each value. The

E. coli "core genes" in 32 genomes

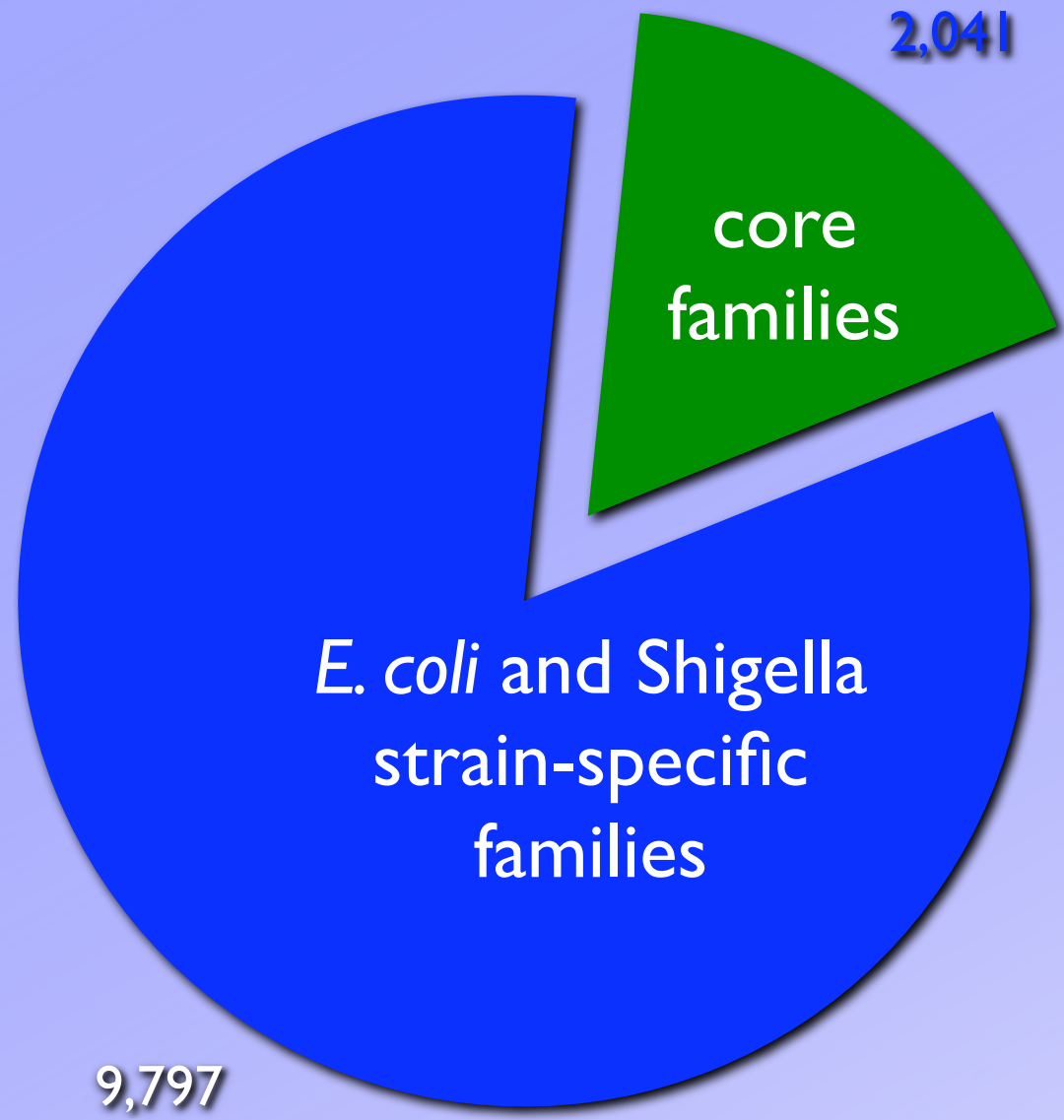


E. coli "pan-genome" based on 32 genomes





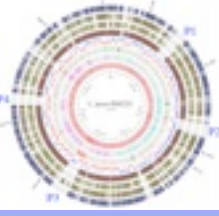
9,797 unique to *E. coli* and *Shigella*



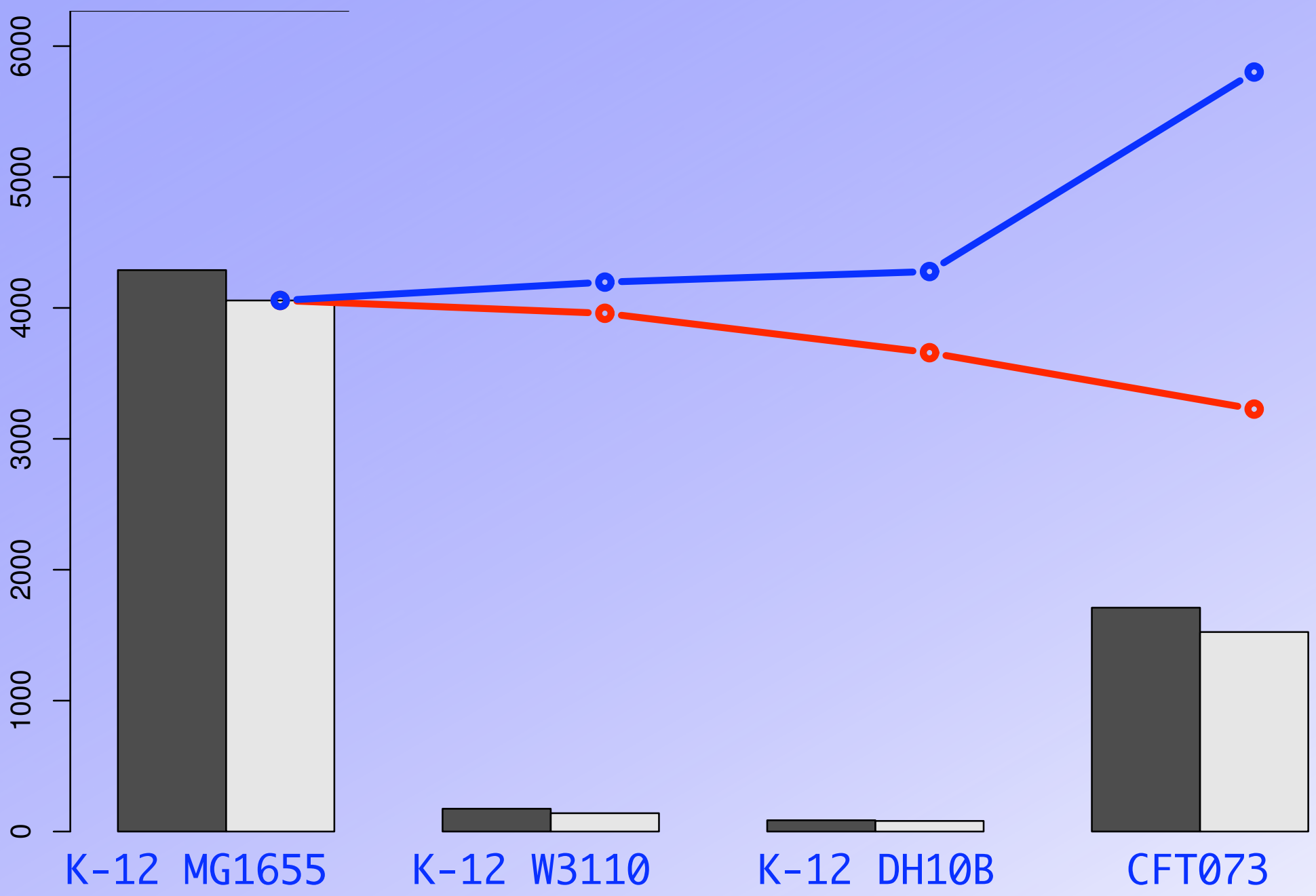
2,041 *E. coli* and *Shigella* core families

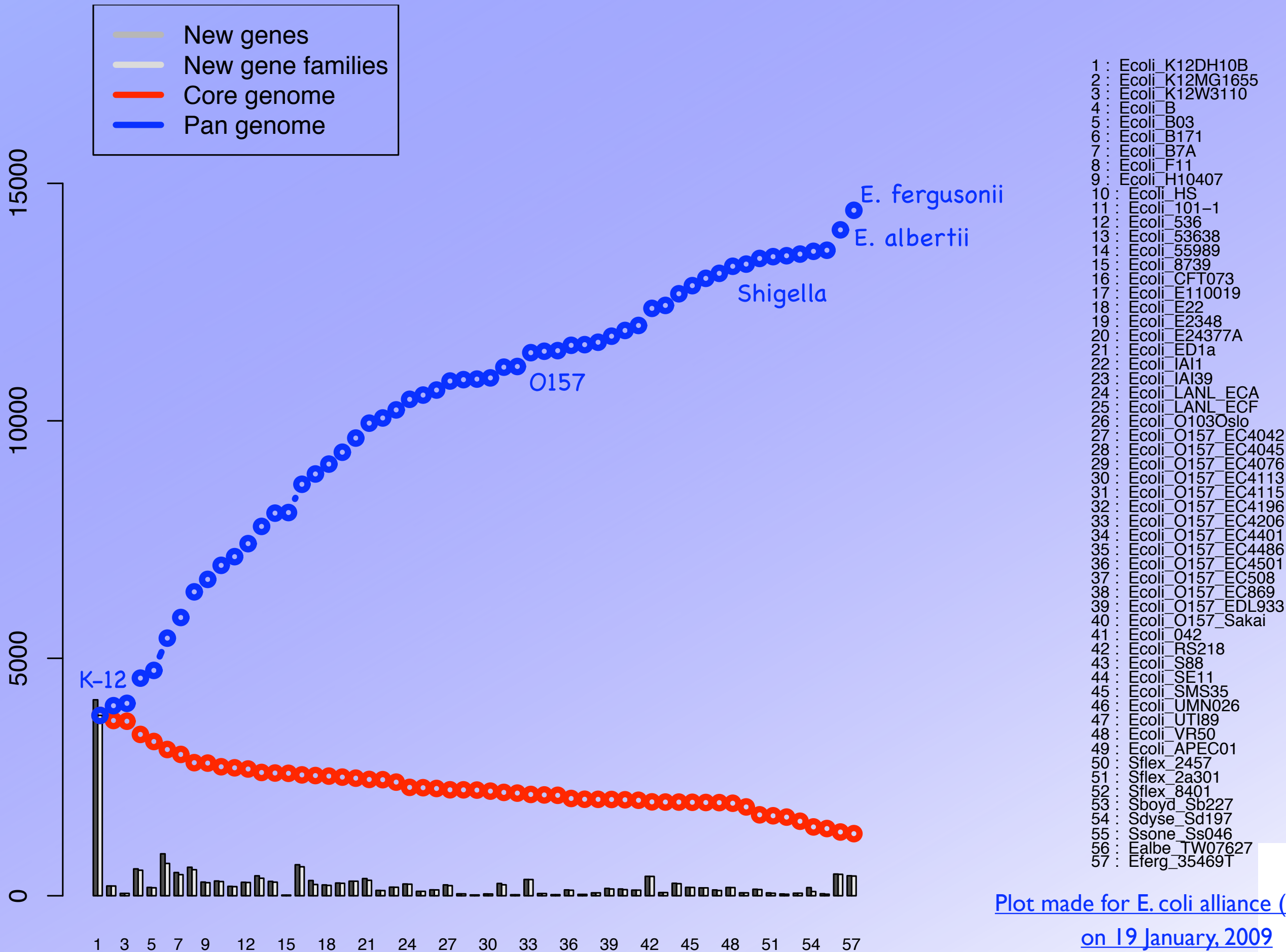
Total of 11,838 *E. coli* and *Shigella* gene families (32 genome sequences)

Comparative Microbial Genomics group



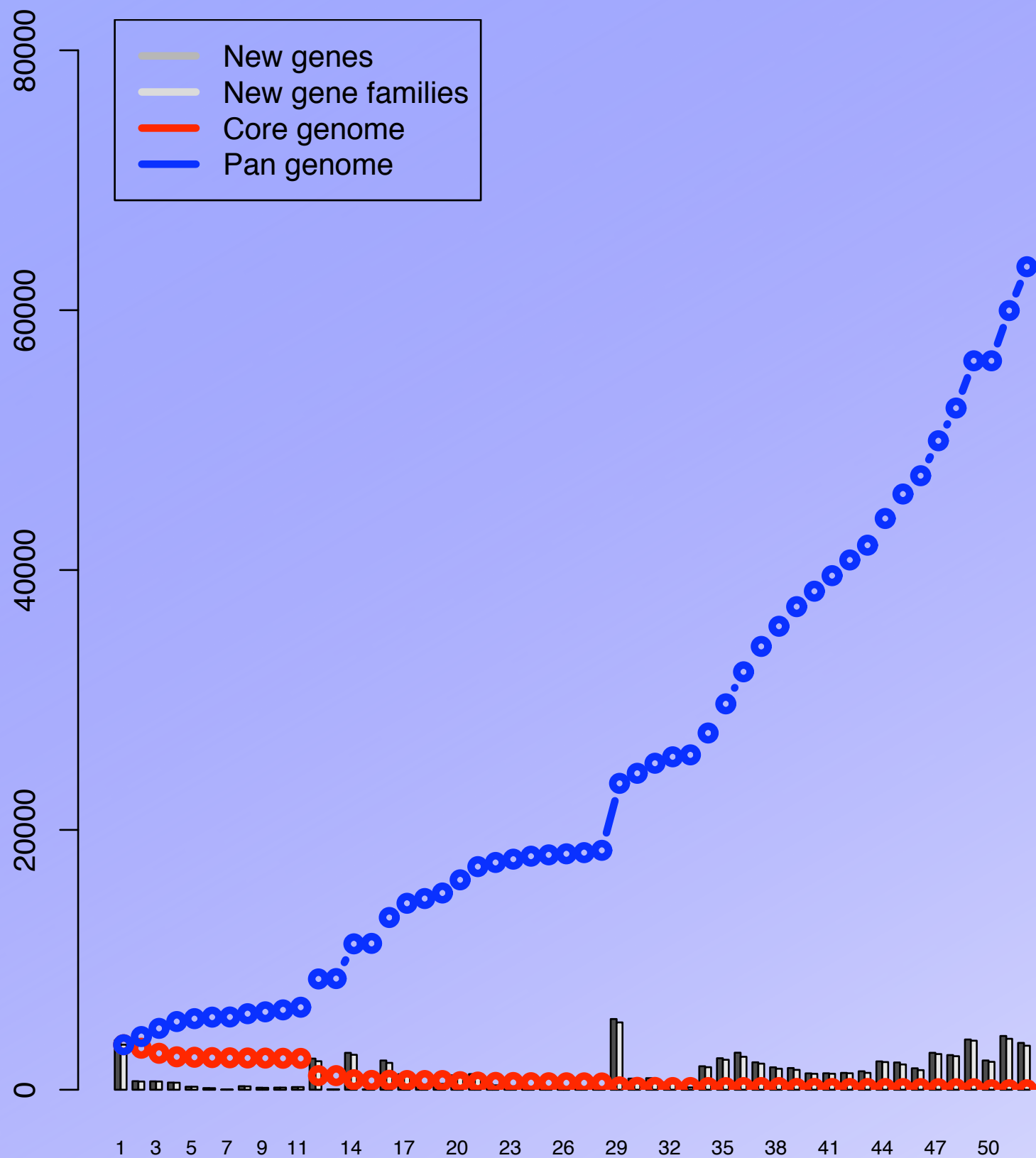
			#genes	# new genes	# new gene families	pan-	core-
1	E. coli K-12 MG1655		4289	4289	4057	4057	4057
2	E. coli K-12 W3110		4387	174	140	4197	3959
3	E. coli K-12 DH10B		4126	86	81	4278	3658
4	E. coli CFT073		5379	1710	1524	5802	3227





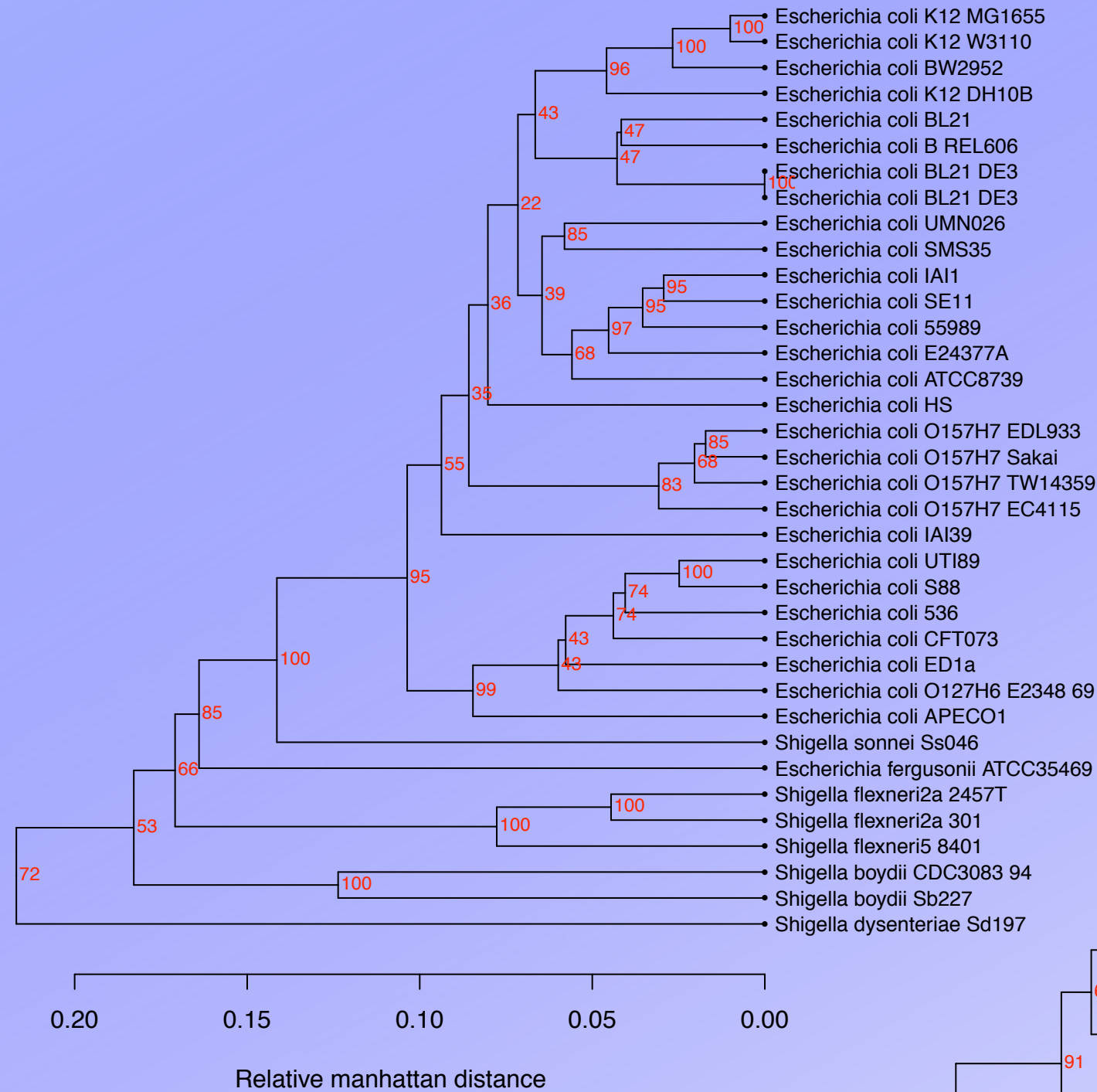
[Plot made for E. coli alliance \(by DU\)](#)

[on 19 January, 2009](#)

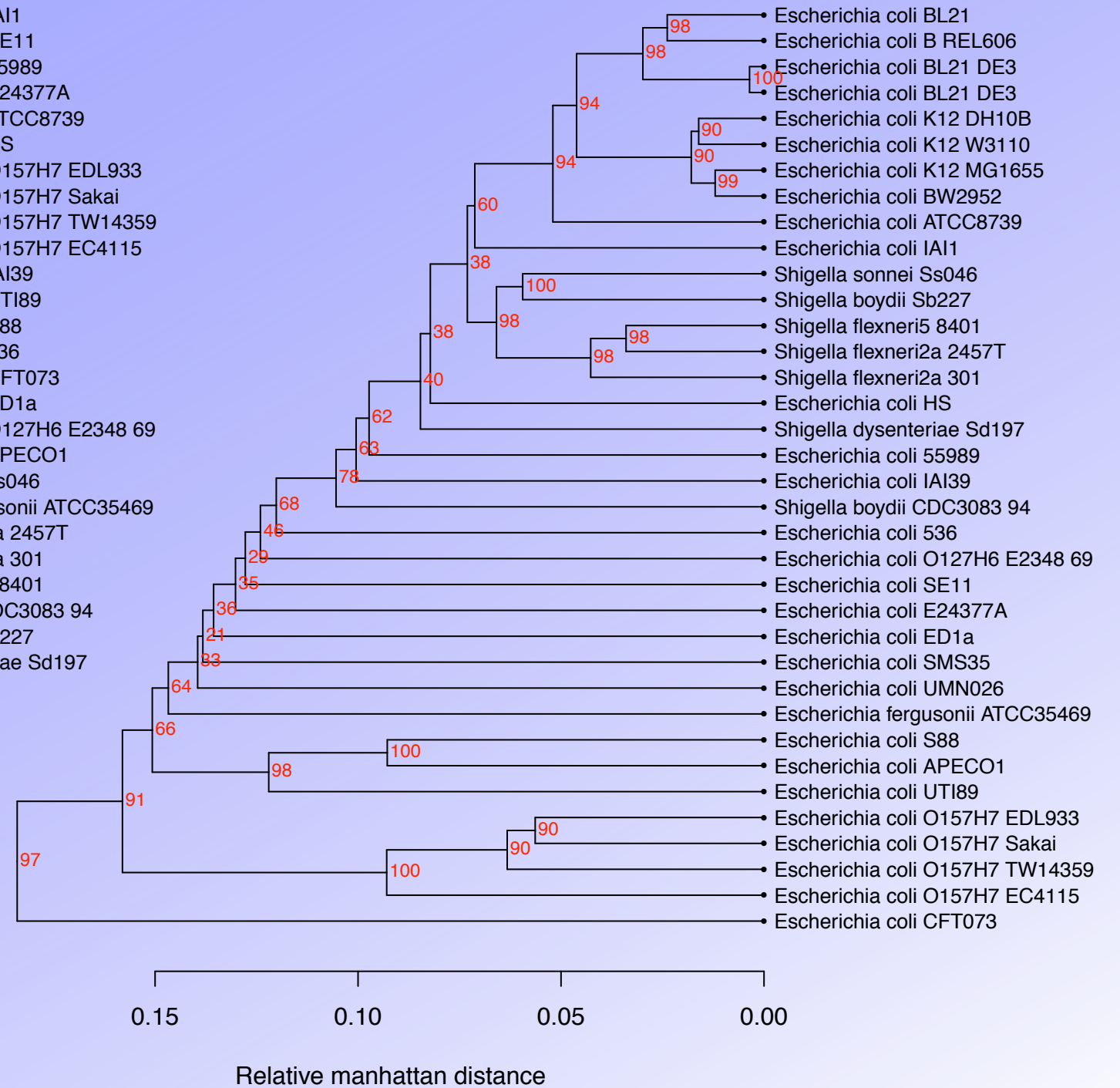


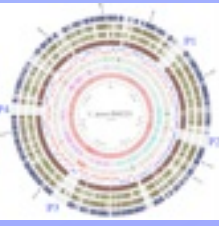
- 1 : *Clostridium botulinum* Bf
- 2 : *Clostridium botulinum* Ba4 str. 657
- 3 : *Clostridium botulinum* A3 str. Loch Maree
- 4 : *Clostridium sporogenes* ATCC 15579
- 5 : *Clostridium botulinum* A str. ATCC 3502
- 6 : *Clostridium botulinum* A str. ATCC 19397
- 7 : *Clostridium botulinum* A str. Hall
- 8 : *Clostridium botulinum* A2 str. Kyoto
- 9 : *Clostridium botulinum* NCTC 2916
- 10 : *Clostridium botulinum* F str. Langeland
- 11 : *Clostridium botulinum* B1 str. Okra
- 12 : *Clostridium kluyveri* DSM 555
- 13 : *Clostridium kluyveri* NBRC 12016
- 14 : *Clostridium butyricum* E4 str. BoNT E BL5262
- 15 : *Clostridium butyricum* 5521
- 16 : *Clostridium beijerinckii* NCIMB 8052
- 17 : *Clostridium botulinum* B str. Eklund 17B
- 18 : *Clostridium botulinum* E3 str. Alaska E43
- 19 : *Clostridium botulinum* E1 str. BoNT E Beluga
- 20 : *Clostridium perfringens* SM101
- 21 : *Clostridium perfringens* E str. JGS1987
- 22 : *Clostridium perfringens* D str. JGS1721
- 23 : *Clostridium perfringens* B str. ATCC 3626
- 24 : *Clostridium perfringens* C str. JGS1495
- 25 : *Clostridium perfringens* str. 13
- 26 : *Clostridium perfringens* ATCC 13124
- 27 : *Clostridium perfringens* NCTC 8239
- 28 : *Clostridium perfringens* CPE str. F4969
- 29 : *Clostridium asparagiforme* DSM 15981
- 30 : *Clostridium tetani* E88
- 31 : *Clostridium botulinum* D str. 1873
- 32 : *Clostridium botulinum* C str. Eklund
- 33 : *Clostridium novyi* NT
- 34 : *Clostridium acetobutylicum* ATCC 824
- 35 : *Clostridium cellulovorans* 743B
- 36 : *Clostridium nexile* DSM 1787
- 37 : *Clostridium scindens* ATCC 35704
- 38 : *Clostridium hylemonae* DSM 15053
- 39 : *Clostridium spiroforme* DSM 1552
- 40 : *Clostridium ramosum* DSM 1402
- 41 : *Clostridium hiranonis* DSM 13275
- 42 : *Clostridium bartlettii* DSM 16795
- 43 : *Clostridium difficile* 630
- 44 : *Clostridium phytofermentans* ISDg
- 45 : *Clostridium cellulolyticum* H10
- 46 : *Clostridium thermocellum* ATCC 27405
- 47 : *Clostridium leptum* DSM 753
- 48 : *Clostridium methylpentosum* DSM 5476
- 49 : *Clostridium bolteae* ATCC BAA-613
- 50 : *Clostridium cellulovorans* 743B
- 51 : *Bacillus anthracis* str. Sterne
- 52 : *Escherichia Coli* str. K12 MG1655

E. coli and Shigella, stabilome

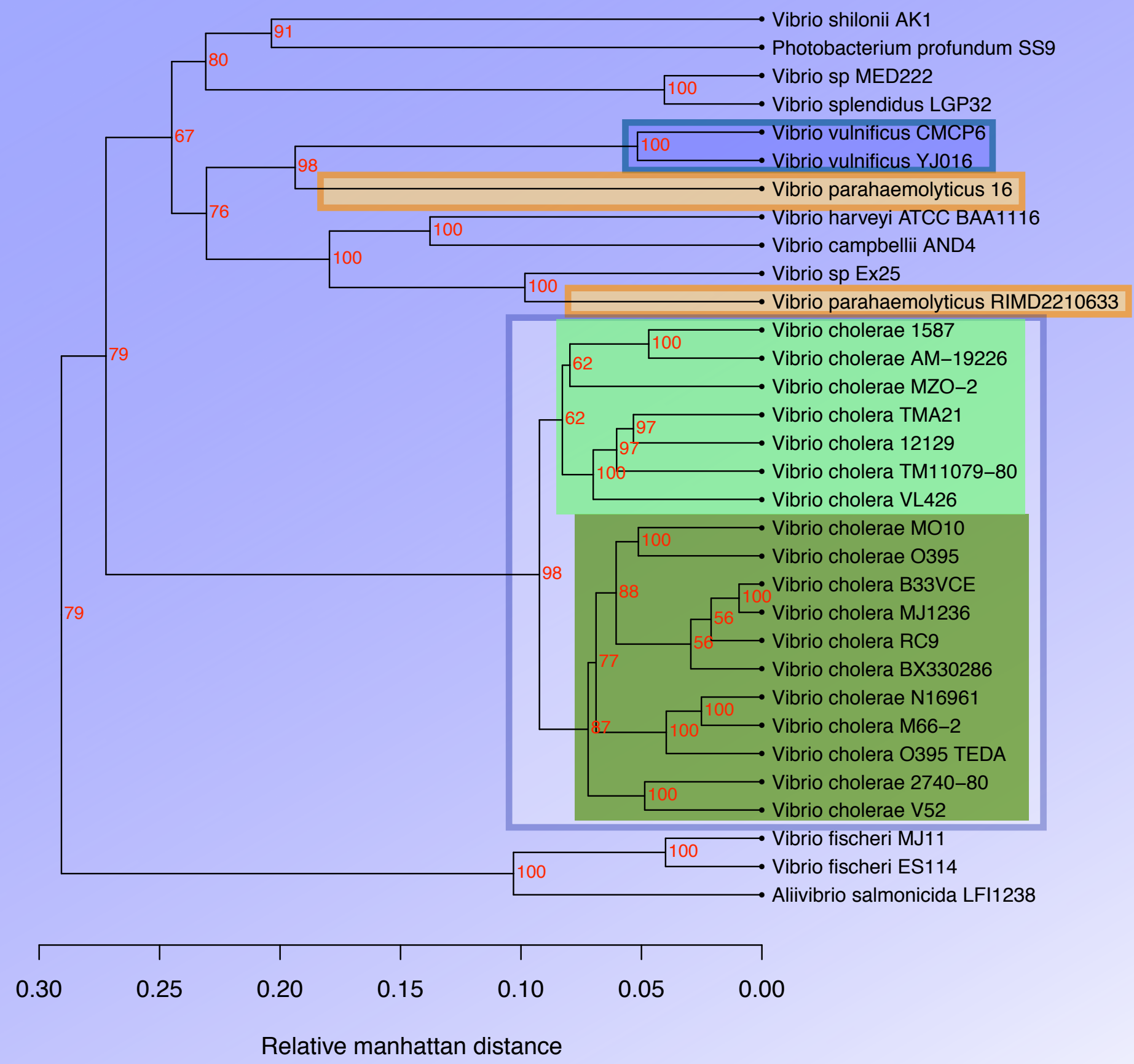


E. coli and Shigella, mobilome

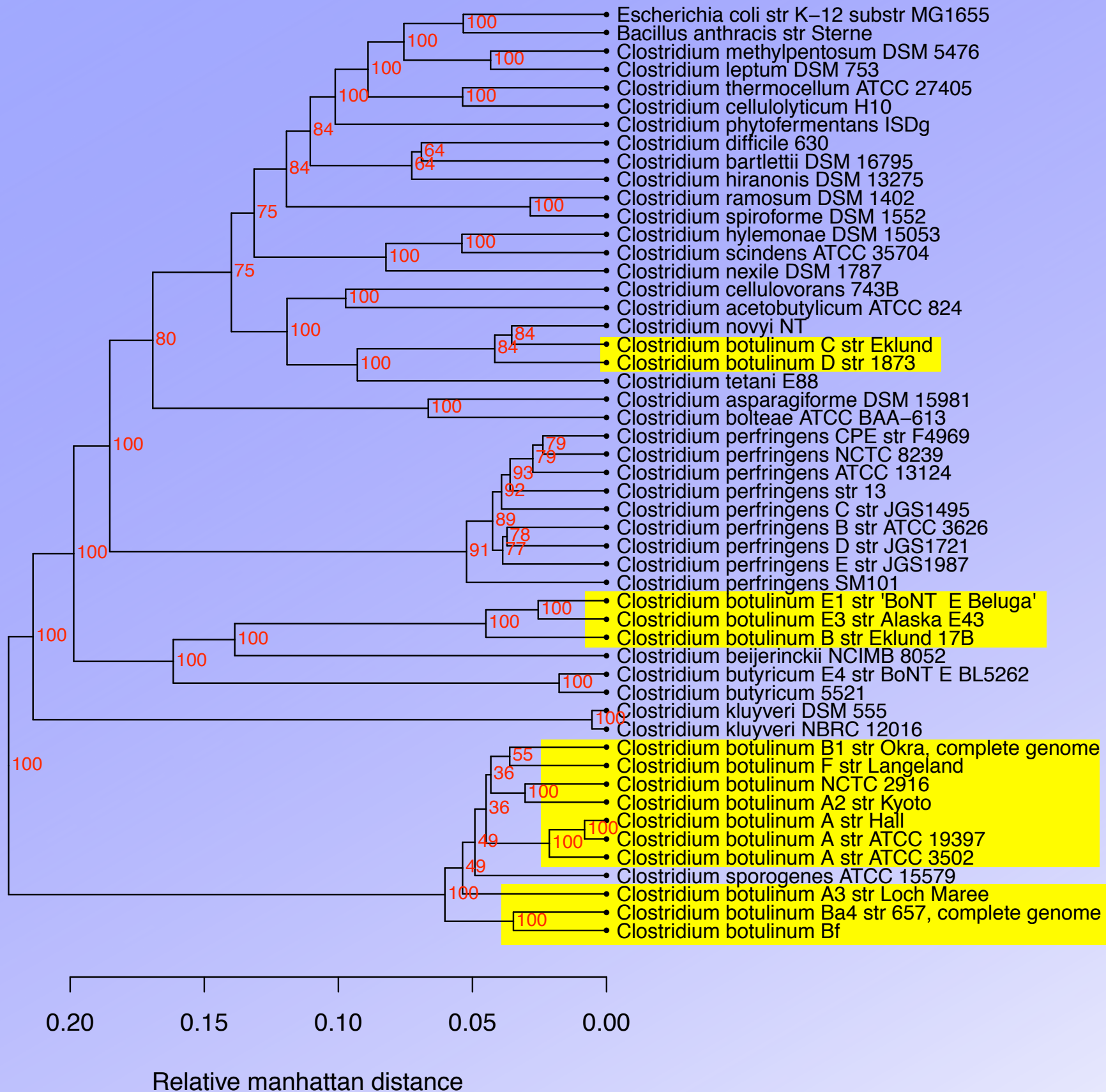




Vibrio



Clostridium and friends



Gene acquisition {
by insertion (mobile elements including phages)
by duplication (recombinations)
by plasmid uptake

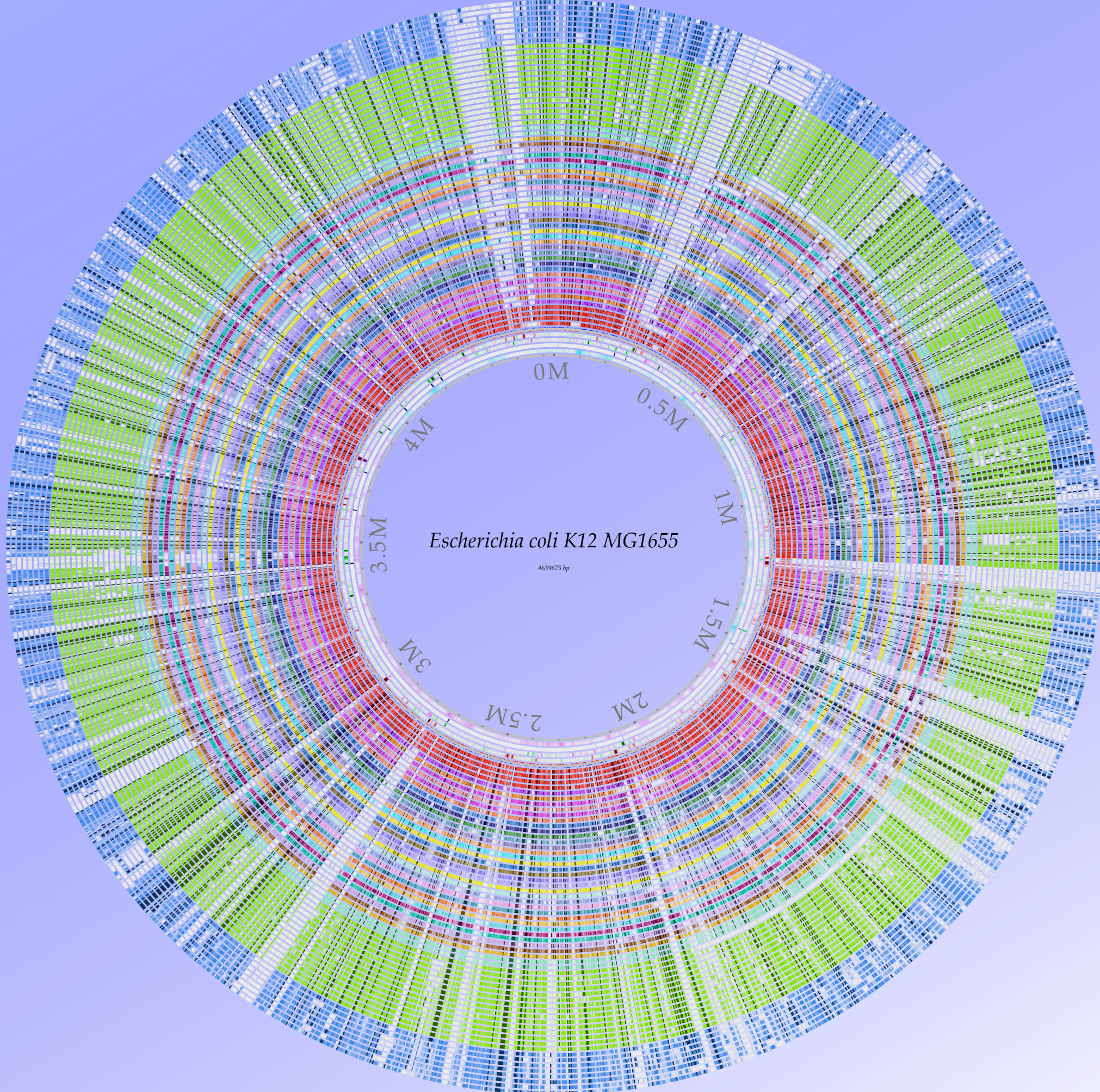
Genome enrichment

Expanding genome

Shrinking genome

Genome reduction

Gene loss {
by excision (mobile elements including phages)
by deletion (recombinations)
by plasmid loss



Escherichia coli K12 MG1655

4639675 bp

0M

0.5M

1M

1.5M

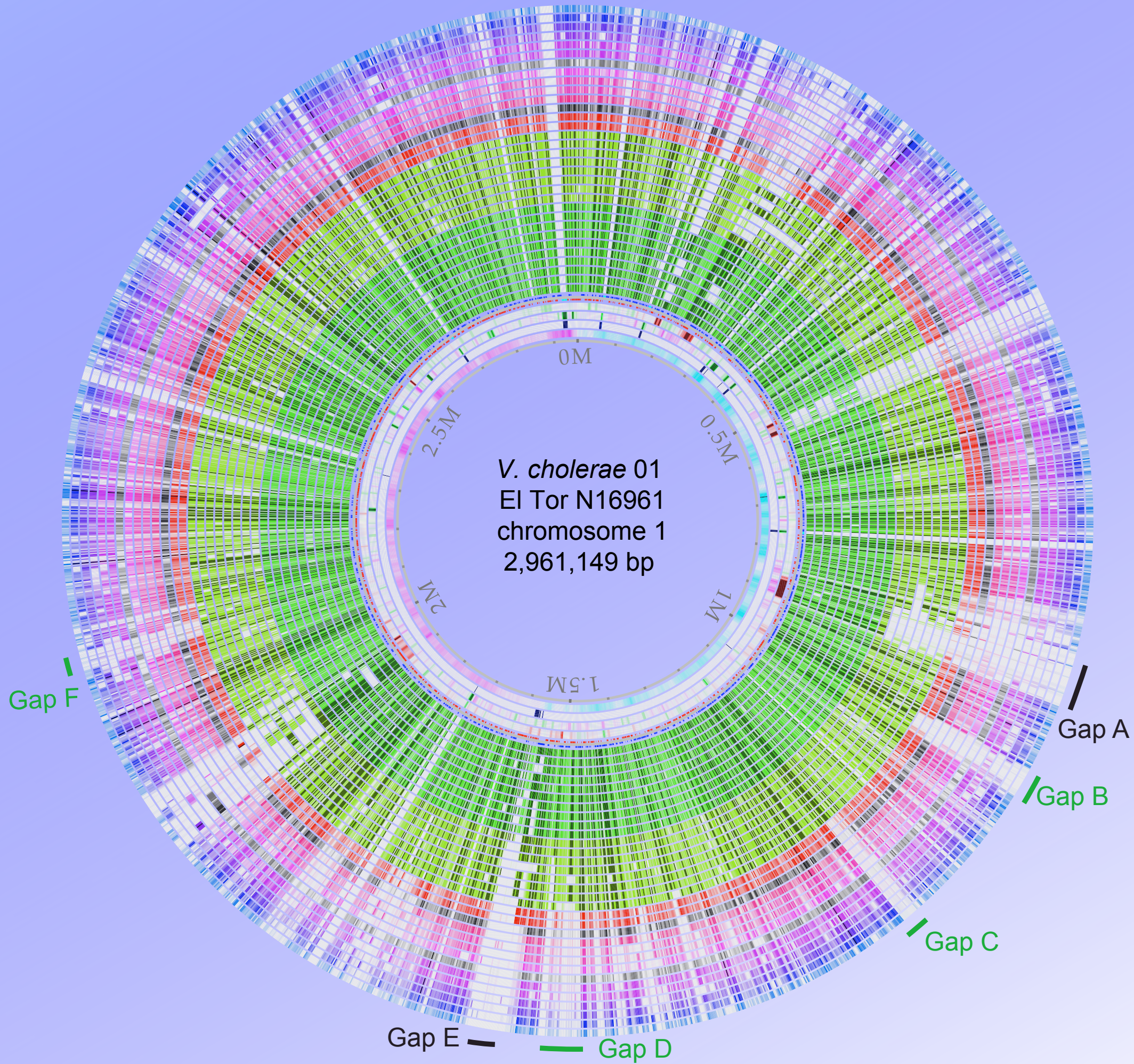
2M

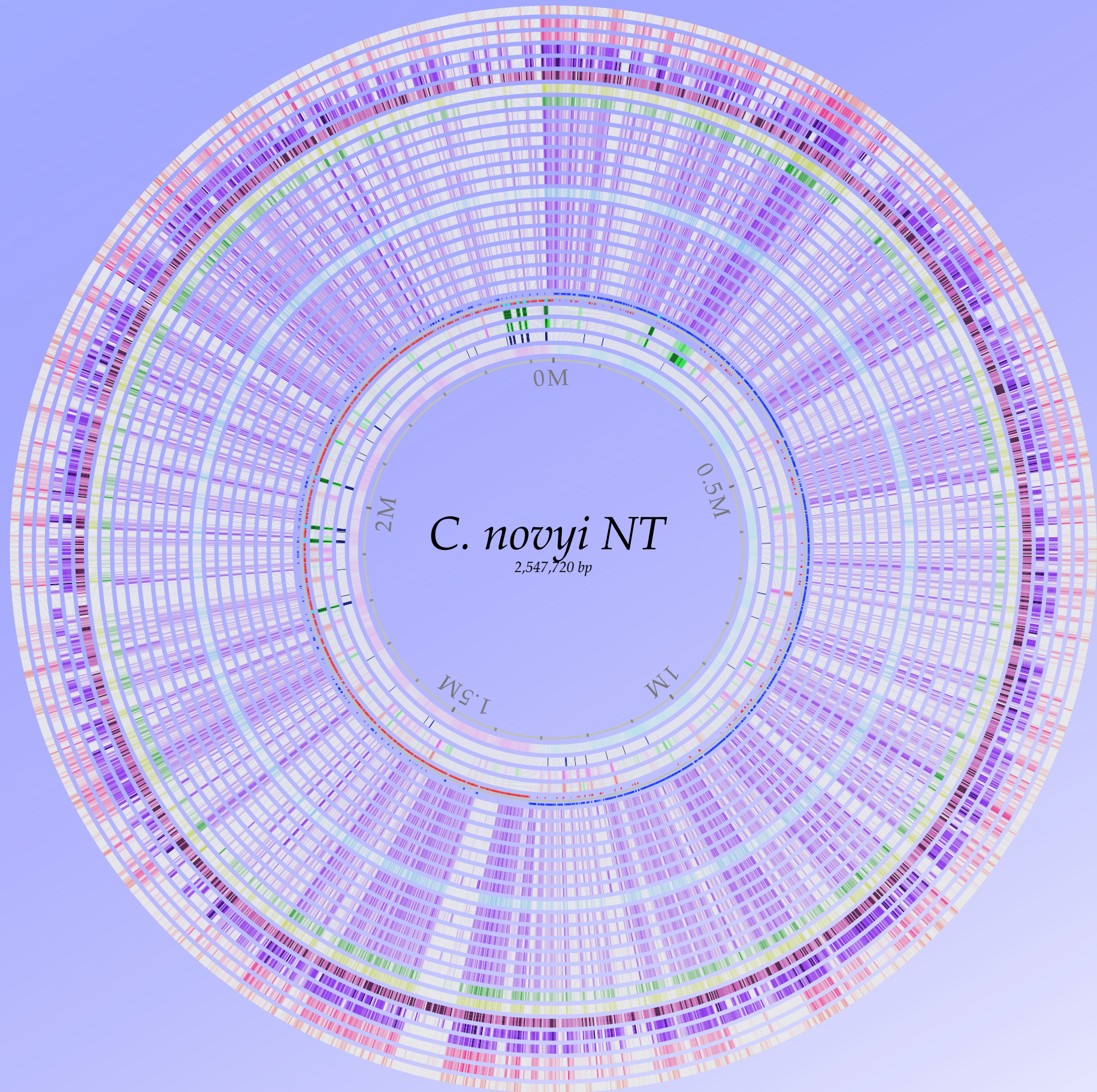
2.5M

3M

3.5M

4M





C. novyi NT
2,547,720 bp

0M

0.5M

1M

1.5M

2M

Whole Genome Amplification and *De novo* Assembly of Single Bacterial Cells

Sébastien Rodrigue¹✉, Rex R. Malmstrom¹✉, Aaron M. Berlin², Bruce W. Birren², Matthew R. Henn², Sallie W. Chisholm^{1*}

1 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **2** The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

Abstract

Background: Single-cell genome sequencing has the potential to allow the in-depth exploration of the vast genetic diversity found in uncultured microbes. We used the marine cyanobacterium *Prochlorococcus* as a model system for addressing important challenges facing high-throughput whole genome amplification (WGA) and complete genome sequencing of individual cells.

Methodology/Principal Findings: We describe a pipeline that enables single-cell WGA on hundreds of cells at a time while virtually eliminating non-target DNA from the reactions. We further developed a post-amplification normalization procedure that mitigates extreme variations in sequencing coverage associated with multiple displacement amplification (MDA), and demonstrated that the procedure increased sequencing efficiency and facilitated genome assembly. We report genome recovery as high as 99.6% with reference-guided assembly, and 95% with *de novo* assembly starting from a single cell. We also analyzed the impact of chimera formation during MDA on *de novo* assembly, and discuss strategies to minimize the presence of incorrectly joined regions in contigs.

Conclusions/Significance: The methods describe in this paper will be useful for sequencing genomes of individual cells from a variety of samples.

Citation: Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, et al. (2009) Whole Genome Amplification and *De novo* Assembly of Single Bacterial Cells. PLoS ONE 4(9): e6864. doi:10.1371/journal.pone.0006864

Editor: Niyaz Ahmed, University of Hyderabad, India

