

Supplementary information

Features of smaller ribosomes in Candidate Phyla Radiation (CPR) bacteria revealed with a molecular evolutionary analysis

Megumi Tsurumaki^{1,2}, Motofumi Saito^{1,2}, Masaru Tomita^{1,2,3}, and Akio Kanai^{1,2,3*}

¹Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan

²Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa 252-0882, Japan

³Faculty of Environment and Information Studies, Keio University, Fujisawa 252-0882, Japan

*Corresponding Author

Akio Kanai, PhD

Institute for Advanced Biosciences, Keio University

Tsuruoka, Yamagata 997-0017, Japan

Tel: +81-235-29-0524

Fax: +81-235-29-0525

E-mail: akio@sfc.keio.ac.jp

Running title: Features of smaller ribosomes in CPR bacteria

Keywords: Candidate phyla radiation, ribosome, rRNA, ribosomal protein, bioinformatic

Supplementary Table S1. Summary of the number of bacterial genome sequences used in this study

A. CPR bacteria

Subgroup	Phyla	Complete genomes	Draft genomes
Microgenomates (MG)	14	9	222
Parcubacteria 1 (PB1)	9	2	117
Parcubacteria 2 (PB2)	5	0	59
Parcubacteria 3 (PB3)	5	1	44
Parcubacteria 4 (PB4)	9	6	160
Other Parcubacteria (Other PB)	9	5	119
Katanobacteria (WWE3)	1	1	25
Saccharibacteria (Sa)	1	32	11
Peregrinibacteria (Pe)	1	7	22
Others	11	6	49
Total	65	69	828

B. non-CPR bacteria

Lifestyle	Phyla	Complete genomes	Draft genomes
Symbiotic	10	167	0
Free-living	31	1,494	0
Total	32*	1,661	0

* Nine phyla were common between Symbiotic and Free-living.

Supplementary Table S2. List of complete genomes of CPR bacteria used in this study

Accession #	Subgroup	Phylum	Organism name	Genome size (base)
GCA_001029675.1	MG	Beckwithbacteria	Candidatus Beckwithbacteria bacterium GW2011_GWC1_49_16	1,049,888
GCA_016699265.1	MG	Roizmanbacteria	Candidatus Roizmanbacteria bacterium	928,688
GCA_016700135.1	MG	Roizmanbacteria	Candidatus Roizmanbacteria bacterium	1,448,136
GCA_013374835.1	MG	Woesebacteria	Candidatus Woesebacteria bacterium	901,682
GCA_013374855.1	MG	Woesebacteria	Candidatus Woesebacteria bacterium	802,554
GCA_013426185.1	MG	Woesebacteria	Candidatus Woesebacteria bacterium	818,817
GCA_016700095.1	MG	Woesebacteria	Candidatus Woesebacteria bacterium	1,274,238
GCA_001029755.1	MG	Woesebacteria	Candidatus Woesebacteria bacterium GW2011_GWF1_31_35	819,458
GCA_009936135.1	MG	unclassified *	Candidatus <i>Chazhembacterium aquaticus</i>	801,504
GCA_016699775.1	PB1	Falkowbacteria	Candidatus Falkowbacteria bacterium	938,851
GCA_016699205.1	PB1	Uhrbacteria	Candidatus Uhrbacteria bacterium	1,082,254
GCA_001029635.1	PB3	Wolfbacteria	Candidatus Wolfbacteria bacterium GW2011_GWB1_47_1	984,447
GCA_016699465.1	PB4	Campbellbacteria	Candidatus Campbellbacteria bacterium	811,376
GCA_001029775.1	PB4	Campbellbacteria	Candidatus Campbellbacteria bacterium GW2011_OD1_34_28	752,630
GCA_016699245.1	PB4	Kaiserbacteria	Candidatus Kaiserbacteria bacterium	1,054,716
GCA_017134395.1	PB4	Kaiserbacteria	Candidatus Kaiserbacteria bacterium	687,746
GCA_016699085.1	PB4	Nomurabacteria	Candidatus Nomurabacteria bacterium	808,221
GCA_016699365.1	PB4	Nomurabacteria	Candidatus Nomurabacteria bacterium	705,143
GCA_016699385.1	Other PB	Moranbacteria	Candidatus Moranbacteria bacterium	1,111,322
GCA_016699425.1	Other PB	Moranbacteria	Candidatus Moranbacteria bacterium	1,217,328
GCA_016699795.1	Other PB	Moranbacteria	Candidatus Moranbacteria bacterium	1,186,598
GCA_016699875.1	Other PB	Moranbacteria	Candidatus Moranbacteria bacterium	1,088,644
GCA_016699975.1	Other PB	Moranbacteria	Candidatus Moranbacteria bacterium	1,048,819
GCA_001007975.1	Others	Absconditabacteria	candidate division SR1 bacterium Aalborg_AAW-1	1,044,756
GCA_000503875.1	Others	Absconditabacteria	candidate division SR1 bacterium RAAC1_SR1_1	1,177,760
GCA_001029735.1	Others	Berkelbacteria	Berkelbacteria bacterium GW2011_GWE1_39_12	915,059
GCA_004563595.2	Others	Gracilibacteria	Candidatus Gracilibacteria bacterium 28_42_T64	1,343,103
GCA_002761215.1	Others	Gracilibacteria	Candidatus Gracilibacteria bacterium HOT-871	1,184,529
GCA_001029795.1	Others	Kazan	candidate division Kazan bacterium GW2011_GWA1_50_15	602,646
GCA_016699145.1	Pe	Peregrinibacteria	Candidatus Peregrinibacteria bacterium	1,338,973
GCA_016699515.1	Pe	Peregrinibacteria	Candidatus Peregrinibacteria bacterium	1,338,661
GCA_016699755.1	Pe	Peregrinibacteria	Candidatus Peregrinibacteria bacterium	1,560,578
GCA_001430755.1	Pe	Peregrinibacteria	Candidatus <i>Peribacter riflensis</i>	1,248,026
GCA_001430845.1	Pe	Peregrinibacteria	Candidatus <i>Peribacter riflensis</i>	1,248,112

GCA_001430885.1	Pe	Peregrinibacteria	Candidatus <i>Peribacter riflensis</i>	1,247,854
GCA_001430905.1	Pe	Peregrinibacteria	Candidatus <i>Peribacter riflensis</i>	1,248,180
GCA_003516025.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	1,005,782
GCA_016699065.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	979,137
GCA_016699895.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	749,073
GCA_016699935.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	1,226,308
GCA_016699955.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	1,211,496
GCA_016700015.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	1,011,181
GCA_016700315.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	988,272
GCA_016700355.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	1,006,672
GCA_016700375.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	1,256,494
GCA_016700395.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium	936,132
GCA_001029695.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium GW2011_GWC2_44_17	1,038,683
GCA_005697215.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	889,964
GCA_010202115.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	931,438
GCA_010202465.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	831,234
GCA_010202645.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	848,784
GCA_010202845.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	853,179
GCA_013098515.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	839,600
GCA_013098855.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	835,009
GCA_013099015.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	877,855
GCA_013099195.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	856,991
GCA_013100805.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	858,111
GCA_013100825.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	893,660
GCA_013394755.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 488	847,999
GCA_005697055.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 955	842,202
GCA_010201925.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 955	885,960
GCA_010202265.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 955	873,056
GCA_013098655.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 955	841,891
GCA_005697565.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium oral taxon 957	782,344
GCA_000503915.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium RAAC3_TM7_1	845,464
GCA_002952755.1	Sa	Saccharibacteria	Candidatus Saccharibacteria bacterium YM_S32_TM7_50_20	1,450,269
GCA_000392435.1	Sa	Saccharibacteria	Candidatus <i>Saccharimonas aalborgensis</i>	1,013,781
GCA_005697395.1	Sa	Saccharibacteria	TM7 phylum sp. oral taxon 952	705,178
GCA_000503835.1	WWE3	WWE3	candidate division WWE3 bacterium RAAC2_WWE3_1	878,109

* Candidatus Chazhembacterium aquaticus is not a member of any candidate phyla described in NCBI taxonomy database.

Supplementary Table S2. Continued

Supplementary Table S3. List of representative genomes of non-CPR free-living bacteria

Accession #	Phylum	Organism name	Genome size (base)
NC_014318	Actinobacteria	<i>Amycolatopsis mediterranei</i> U32	10,236,715
NC_004307	Actinobacteria	<i>Bifidobacterium longum</i> NCC2705	2,256,640
NC_013172	Actinobacteria	<i>Brachybacterium faecium</i> DSM 4810	3,614,992
NC_003450	Actinobacteria	<i>Corynebacterium glutamicum</i> ATCC 13032	3,309,401
NC_008596	Actinobacteria	<i>Mycobacterium smegmatis</i> str. MC2 155	6,988,209
NC_010397	Actinobacteria	<i>Mycobacteroides abscessus</i>	5,067,172
NC_003888	Actinobacteria	<i>Streptomyces coelicolor</i> A3(2)	8,667,507
NC_000918	Aquificae	<i>Aquifex aeolicus</i> VF5	1,551,335
NC_006347	Bacteroidetes	<i>Bacteroides fragilis</i> YCH46	5,277,274
NC_009613	Bacteroidetes	<i>Flavobacterium psychrophilum</i> JIP02/86	2,860,382
NC_007677	Bacteroidetes	<i>Salinibacter ruber</i> DSM 13855	3,551,823
NC_002932	Chlorobi	<i>Chlorobium tepidum</i> TLS	2,154,946
NC_010175	Chloroflexi	<i>Chloroflexus aurantiacus</i> J-10-fl	5,258,541
NC_005125	Cyanobacteria	<i>Gloeobacter violaceus</i> PCC 7421	4,659,019
NC_005042	Cyanobacteria	<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. CCMP1375	1,751,080
NC_004113	Cyanobacteria	<i>Thermosynechococcus elongatus</i> BP-1	2,593,857
NC_001263	Deinococcus-Thermus	<i>Deinococcus radiodurans</i> R1	3,060,986
NC_006461	Deinococcus-Thermus	<i>Thermus thermophilus</i> HB8	1,849,742
NC_011661	Dictyoglomi	<i>Dictyoglomus turgidum</i> DSM 6724	1,855,560
NC_000964	Firmicutes	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	4,215,606
NC_004722	Firmicutes	<i>Bacillus cereus</i> ATCC 14579	5,411,809
NC_003030	Firmicutes	<i>Clostridium acetobutylicum</i> ATCC 824	3,940,880
NC_009495	Firmicutes	<i>Clostridium botulinum</i> A str. ATCC 3502	3,886,916
NC_004668	Firmicutes	<i>Enterococcus faecalis</i> V583	3,218,031
NC_017960	Firmicutes	<i>Enterococcus faecium</i> DO	2,698,137
NC_004567	Firmicutes	<i>Lactobacillus plantarum</i> WCFS1	3,308,273
NC_008526	Firmicutes	<i>Lactobacillus paracasei</i> ATCC 334	2,895,264
NC_002662	Firmicutes	<i>Lactococcus lactis</i> subsp. <i>lactis</i> II1403	2,365,589
NC_003210	Firmicutes	<i>Listeria monocytogenes</i> EGD-e	2,944,528
NC_004461	Firmicutes	<i>Staphylococcus epidermidis</i> ATCC 12228	2,499,279
NC_007795	Firmicutes	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> NCTC 8325	2,821,361
NC_003098	Firmicutes	<i>Streptococcus pneumoniae</i> R6	2,038,615
NC_004350	Firmicutes	<i>Streptococcus mutans</i> UA159	2,032,925
NC_012926	Firmicutes	<i>Streptococcus suis</i> BM407	2,146,229
NC_013853	Firmicutes	<i>Streptococcus mitis</i> B6	2,146,611
NC_005957	Firmicutes	[<i>Bacillus thuringiensis</i>] serovar <i>konkukian</i> str. 97-27	5,237,682
NC_003454	Fusobacteria	<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i> ATCC 25586	2,174,500
NC_011296	Nitrospirae	<i>Thermodesulfovibrio yellowstonii</i> DSM 11347	2,003,803
NC_005027	Planctomycetes	<i>Rhodopirellula baltica</i> SH 1	7,145,576

NC_016603	Proteobacteria	<i>Acinetobacter pittii</i> PHEA-2	3,862,530
NC_003062	Proteobacteria	<i>Agrobacterium fabrum</i> str. C58	4,917,157
NC_002929	Proteobacteria	<i>Bordetella pertussis</i> Tohama I	4,086,189
NC_019382	Proteobacteria	<i>Bordetella bronchiseptica</i> 253	5,264,383
NC_006348	Proteobacteria	<i>Burkholderia mallei</i> ATCC 23344	5,835,527
NC_006350	Proteobacteria	<i>Burkholderia pseudomallei</i> K96243	7,247,547
NC_002696	Proteobacteria	<i>Caulobacter crescentus</i> CB15	4,016,947
NC_002937	Proteobacteria	<i>Desulfovibrio vulgaris</i> str. Hildenborough	3,570,858
NC_014121	Proteobacteria	<i>Enterobacter cloacae</i> subsp. <i>cloacae</i> ATCC 13047	5,314,581
NC_000913	Proteobacteria	<i>Escherichia coli</i> str. K-12 substr. MG1655	4,641,652
NC_006570	Proteobacteria	<i>Francisella tularensis</i> subsp. <i>tularensis</i> SCHU S4	1,892,775
NC_002939	Proteobacteria	<i>Geobacter sulfurreducens</i> PCA	3,814,128
NC_000915	Proteobacteria	<i>Helicobacter pylori</i> 26695	1,667,867
NC_017384	Proteobacteria	<i>Ketogulonicigenium vulgare</i> WSH-001	2,766,400
NC_016845	Proteobacteria	<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> HS11286	5,333,942
NC_002942	Proteobacteria	<i>Legionella pneumophila</i> subsp. <i>pneumophila</i> str. Philadelphia 1	3,397,754
NC_002946	Proteobacteria	<i>Neisseria gonorrhoeae</i> FA 1090	2,153,922
NC_002516	Proteobacteria	<i>Pseudomonas aeruginosa</i> PAO1	6,264,404
NC_002947	Proteobacteria	<i>Pseudomonas putida</i> KT2440	6,181,873
NC_004578	Proteobacteria	<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	6,397,126
NC_007643	Proteobacteria	<i>Rhodospirillum rubrum</i> ATCC 11170	4,352,825
NC_003197	Proteobacteria	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i> str. LT2	4,857,450
NC_004347	Proteobacteria	<i>Shewanella oneidensis</i> MR-1	4,969,811
NC_004337	Proteobacteria	<i>Shigella flexneri</i> 2a str. 301	4,607,202
NC_003047	Proteobacteria	<i>Sinorhizobium meliloti</i> 1021	3,654,135
NC_012587	Proteobacteria	<i>Sinorhizobium fredii</i> NGR234	3,925,702
NC_003902	Proteobacteria	<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	5,076,188
NC_003143	Proteobacteria	<i>Yersinia pestis</i> CO92	4,653,728
NC_004342	Spirochaetes	<i>Leptospira interrogans</i> serovar <i>Lai</i> str. 56601	4,698,134
NC_002967	Spirochaetes	<i>Treponema denticola</i> ATCC 35405	2,843,201
NC_000853	Thermotogae	<i>Thermotoga maritima</i> MSB8	1,860,725

Supplementary Table S3. Continued

Supplementary Table S4. List of representative genomes of non-CPR symbiotic bacteria

Accession #	Phylum	Organism name	Genome size (base)
NC_004572	Actinobacteria	<i>Tropheryma whipplei</i> str. Twist	927,303
NC_016621	Bacteroidetes	<i>Blattabacterium</i> sp. (<i>Cryptocercus punctulatus</i>) str. Cpu	605,745
NC_020510	Bacteroidetes	<i>Blattabacterium cuenoti</i> BPAA	632,490
NC_018605	Bacteroidetes	<i>Cardinium</i> endosymbiont cEper1 of <i>Encarsia pergandiella</i>	887,130
NC_000117	Chlamydiae	<i>Chlamydia trachomatis</i> D/UW-3/CX	1,042,519
NC_002620	Chlamydiae	<i>Chlamydia muridarum</i> str. Nigg	1,072,950
NC_007899	Chlamydiae	<i>Chlamydia felis</i> Fe/C-56	1,166,239
NC_010287	Chlamydiae	<i>Chlamydia trachomatis</i> 434/Bu	1,038,842
NC_017287	Chlamydiae	<i>Chlamydia psittaci</i> 6BC	1,171,667
NZ_CP006571	Chlamydiae	<i>Chlamydia avium</i> 10DC88	1,041,170
NZ_CP015840	Chlamydiae	<i>Chlamydia gallinacea</i> 08-1274/3	1,059,583
NZ_LT821323	Chlamydiae	<i>Chlamydia suis</i>	1,075,789
NC_000922	Chlamydiae	<i>Chlamydophila pneumoniae</i> CWL029	1,230,230
NC_003361	Chlamydiae	<i>Chlamydophila caviae</i> GPIC	1,173,390
NC_015408	Chlamydiae	<i>Chlamydophila pecorum</i> E58	1,106,197
NC_020419	Elusimicrobia	uncultured Termite group 1 bacterium phylotype Rs-D17	1,125,857
NZ_CP017037	Firmicutes	<i>Dialister pneumosintes</i>	1,274,073
NZ_CP009224	Firmicutes	<i>Weissella ceti</i>	1,390,397
NZ_CP014606	Firmicutes	endosymbiont 'TC1' of <i>Trimyema compressum</i>	1,586,453
NZ_CP011280	Fusobacteria	<i>Sneathia amnii</i>	1,330,224
NC_007797	Proteobacteria	<i>Anaplasma phagocytophilum</i> str. HZ	1,471,282
NC_012026	Proteobacteria	<i>Anaplasma marginale</i> str. Florida	1,202,435
NC_013532	Proteobacteria	<i>Anaplasma centrale</i> str. Israel	1,206,806
NZ_CP013920	Proteobacteria	<i>Arsenophonus</i> symbiont of <i>Lipoptena fortisetosa</i>	836,724
NC_008783	Proteobacteria	<i>Bartonella bacilliformis</i> KC583	1,445,021
NZ_CP010401	Proteobacteria	<i>Bartonella ancashensis</i>	1,467,695
NZ_CP010048	Proteobacteria	<i>Blochmannia</i> endosymbiont of <i>Polyrhachis (Hedomyrma) turneri</i>	749,321
NZ_CP010049	Proteobacteria	<i>Blochmannia</i> endosymbiont of <i>Camponotus (Colobopsis) obliquus</i>	773,940
NC_002528	Proteobacteria	<i>Buchnera aphidicola</i> str. APS (<i>Acyrtosiphon pisum</i>)	640,681
NC_008513	Proteobacteria	<i>Buchnera aphidicola</i> BCc	416,380
NC_017256	Proteobacteria	<i>Buchnera aphidicola</i> str. Ak (<i>Acyrtosiphon kondoi</i>)	641,794
NZ_CP011299	Proteobacteria	<i>Buchnera aphidicola</i> (<i>Schlechtendalia chinensis</i>)	607,835
NZ_LT635893	Proteobacteria	<i>Buchnera aphidicola</i> (<i>Cinara pseudotaxifoliae</i>)	446,627
NZ_CP007770	Proteobacteria	<i>Campylobacter insulaenigrae</i> NCTC 12927	1,465,081
NC_009465	Proteobacteria	Candidatus <i>Vesicomysocius okutanii</i> HA	1,022,154
NC_019814	Proteobacteria	Candidatus <i>Kinetoplastibacterium blastocritidii</i> (ex <i>Strigomonas culicis</i>)	820,037
NC_002971	Proteobacteria	<i>Coxiella burnetii</i> RSA 493	1,995,488
NZ_CP007541	Proteobacteria	<i>Coxiella</i> endosymbiont of <i>Amblyomma americanum</i>	656,901
NC_009446	Proteobacteria	<i>Dichelobacter nodosus</i> VCS1703A	1,389,350

NC_007354	Proteobacteria	<i>Ehrlichia canis</i> str. Jake	1,315,030
NC_007799	Proteobacteria	<i>Ehrlichia chaffeensis</i> str. Arkansas	1,176,248
NC_023063	Proteobacteria	<i>Ehrlichia muris</i> AS145	1,196,717
NC_008011	Proteobacteria	<i>Lawsonia intracellularis</i> PHE/MN1-00	1,457,619
NC_007798	Proteobacteria	<i>Neorickettsia sennetsu</i> str. Miyayama	859,006
NC_013009	Proteobacteria	<i>Neorickettsia risticii</i> str. Illinois	879,977
NZ_CP007481	Proteobacteria	<i>Neorickettsia helminthoeca</i> str. Oregon	884,232
NC_000963	Proteobacteria	<i>Rickettsia prowazekii</i> str. Madrid E	1,111,523
NC_003103	Proteobacteria	<i>Rickettsia conorii</i> str. Malish 7	1,268,755
NC_007109	Proteobacteria	<i>Rickettsia felis</i> URRWXCal2	1,485,148
NC_016050	Proteobacteria	<i>Rickettsia japonica</i> YH	1,283,087
NC_004344	Proteobacteria	<i>Wigglesworthia glossinidia</i> endosymbiont of <i>Glossina brevipalpis</i>	697,724
NC_002978	Proteobacteria	<i>Wolbachia endosymbiont</i> of <i>Drosophila melanogaster</i>	1,267,782
NC_021084	Proteobacteria	<i>Wolbachia endosymbiont</i> of <i>Drosophila simulans</i> wNo	1,301,823
NC_001318	Spirochaetes	<i>Borrelia burgdorferi</i> B31	910,724
NC_018887	Spirochaetes	<i>Borrelia afzelii</i> HLJ01	905,471
NZ_CP015780	Spirochaetes	<i>Borrelia mayonii</i>	904,387
NZ_CP009117	Spirochaetes	<i>Borrelia valaisiana</i> Tom4006	912,160
NZ_CP018744	Spirochaetes	<i>Borrelia garinii</i>	905,638
NC_006055	Tenericutes	<i>Mesoplasma florum</i> L1	793,224
NC_000908	Tenericutes	<i>Mycoplasma genitalium</i> G37	580,076
NC_000912	Tenericutes	<i>Mycoplasma pneumoniae</i> M129	816,394
NC_005364	Tenericutes	<i>Mycoplasma mycoides</i> subsp. <i>mycoides</i> SC str. PG1	1,211,703
NC_014760	Tenericutes	<i>Mycoplasma bovis</i> PG45	1,003,404
NC_022575	Tenericutes	<i>Mycoplasma parvum</i> str. Indiana	564,395
NC_005303	Tenericutes	Onion yellows phytoplasma OY-M	853,092
NC_021284	Tenericutes	<i>Spiroplasma syrphidicola</i> EA-1	1,107,344
NC_021846	Tenericutes	<i>Spiroplasma taiwanense</i> CT-1	1,075,140
NZ_CP006720	Tenericutes	<i>Spiroplasma mirum</i> ATCC 29335	1,132,608
NZ_CP012328	Tenericutes	<i>Spiroplasma turonicum</i>	1,261,374
NC_010503	Tenericutes	<i>Ureaplasma parvum</i> serovar 3 str. ATCC 27815	751,679

Supplementary Table S4. Continued

Supplementary Table S5. List of query sequences (COG sequences) for ribosomal proteins

Protein name	COG ID	Number of sequences	Protein name	GOG ID	Number of sequences
bS1	COG0539	729	uL1	COG0539	709
uS2	COG0052	709	uL2	COG0052	707
uS3	COG0092	707	uL3	COG0092	710
uS4	COG0522	744	uL4	COG0522	709
uS5	COG0098	709	uL5	COG0098	710
bS6	COG0360	628	uL6	COG0360	707
uS7	COG0049	712	bL9	COG0049	640
uS8	COG0096	709	uL10	COG0096	623
uS9	COG0103	712	uL11	COG0103	707
uS10	COG0051	723	bL12	COG0051	716
uS11	COG0100	704	uL13	COG0100	707
uS12	COG0048	711	uL14	COG0048	709
uS13	COG0099	717	uL15	COG0099	709
uS14	COG0199	663	uL16	COG0199	710
uS15	COG0184	710	bL17	COG0184	628
bS16	COG0228	624	uL18	COG0228	703
uS17	COG0186	706	bL19	COG0186	625
bS18	COG0238	631	bL20	COG0238	627
uS19	COG0185	704	bL21	COG0185	624
bS20	COG0268	611	uL22	COG0268	709
bS21	COG0828	505	uL23	COG0828	706
			uL24	COG0539	703
			bL25	COG0052	563
			bL27	COG0092	635
			bL28	COG0522	634
			uL29	COG0098	659
			uL30	COG0360	594
			bL31	COG0049	688
			bL32	COG0096	579
			bL33	COG0103	673
			bL34	COG0051	468
			bL35	COG0100	598
			bL36	COG0048	559

See the following URL for each COG ID:
<https://www.ncbi.nlm.nih.gov/research/cog-project/>

Supplementary Table S6. List of queries (Pfam domain data) for ribosomal proteins

Protein name	Pfam ID	Domain name	Protein name	Pfam ID	Domain name
bS1	PF00575	S1	uL1	PF00687	Ribosomal_L1
uS2	PF00318	Ribosomal_S2	uL2	PF00181	Ribosomal_L2
uS3	PF00189	Ribosomal_S3_C	uL2	PF03947	Ribosomal_L2_C
uS4	PF00163	Ribosomal_S4	uL3	PF00297	Ribosomal_L3
uS5	PF00333	Ribosomal_S5	uL4	PF00573	Ribosomal_L4
bS5	PF03719	Ribosomal_S5_C	uL5	PF00281	Ribosomal_L5
bS6	PF01250	Ribosomal_S6	uL5	PF00673	Ribosomal_L5_C
uS7	PF00177	Ribosomal_S7	uL6	PF00347	Ribosomal_L6
uS8	PF00410	Ribosomal_S8	bL9	PF03948	Ribosomal_L9_C
uS9	PF00380	Ribosomal_S9	bL9	PF01281	Ribosomal_L9_N
uS10	PF00338	Ribosomal_S10	uL10	PF00466	Ribosomal_L10
uS11	PF00411	Ribosomal_S11	uL11	PF00298	Ribosomal_L11
uS12	PF00164	Ribosom_S12_S23	uL11	PF03946	Ribosomal_L11_N
uS13	PF00416	Ribosomal_S13	bL12	PF00542	Ribosomal_L12
uS14	PF00253	Ribosomal_S14	bL12	PF16320	Ribosomal_L12_N
uS15	PF00312	Ribosomal_S15	uL13	PF00572	Ribosomal_L13
bS16	PF00886	Ribosomal_S16	uL14	PF00238	Ribosomal_L14
uS17	PF00366	Ribosomal_S17	uL15	PF00828	Ribosomal_L27A
bS18	PF01084	Ribosomal_S18	uL16	PF00252	Ribosomal_L16
uS19	PF00203	Ribosomal_S19	bL17	PF01196	Ribosomal_L17
bS20	PF01649	Ribosomal_S20p	uL18	PF00861	Ribosomal_L18p
bS21	PF01165	Ribosomal_S21	bL19	PF01245	Ribosomal_L19
			bL20	PF00453	Ribosomal_L20
			bL21	PF00829	Ribosomal_L21p
			uL22	PF00237	Ribosomal_L22
			uL23	PF00276	Ribosomal_L23
			uL24	PF17136	ribosomal_L24
			bL25	PF01386	Ribosomal_L25p
			bL25	PF14693	Ribosomal_TL5_C
			bL27	PF01016	Ribosomal_L27
			bL28	PF00830	Ribosomal_L28
			uL29	PF00831	Ribosomal_L29
			uL30	PF00327	Ribosomal_L30
			bL31	PF01197	Ribosomal_L31
			bL32	PF01783	Ribosomal_L32p
			bL33	PF00471	Ribosomal_L33
			bL34	PF00468	Ribosomal_L34
			bL35	PF01632	Ribosomal_L35p
			bL36	PF00444	Ribosomal_L36

See the following URL for each COG ID:
<http://pfam.xfam.org/>

Supplementary Table S7. Statistical comparison of ribosomal protein length in CPR and non-CPR bacteria.

Protein name	Length in CPR (aa)	Length in non-CPR (aa)	<i>p</i>-value (Mann–Whitney U test)
bS1	389.7 ± 46.5	519.9 ± 90.0	< 0.001
uS2	260.1 ± 33.0	266.9 ± 27.7	< 0.001
uS3	226.3 ± 19.6	239.7 ± 22.6	< 0.001
uS4	210.4 ± 20.4	204.4 ± 6.8	< 0.001
uS5	185.9 ± 28.9	181.4 ± 23.9	< 0.001
bS6	153.5 ± 43.0	119.5 ± 28.5	< 0.001
uS7	162.4 ± 27.5	156.4 ± 2.7	< 0.001
uS8	137.8 ± 19.1	130.9 ± 7.7	< 0.001
uS9	153.5 ± 22.1	140.4 ± 19.3	< 0.001
uS10	120.1 ± 24.2	103.3 ± 8.5	< 0.001
uS11	144.9 ± 19.9	130.7 ± 5.4	< 0.001
uS12	149.0 ± 15.5	127.5 ± 6.2	< 0.001
uS13	134.2 ± 10.4	121.7 ± 4.2	< 0.001
uS14	69.0 ± 13.4	83.1 ± 18.7	< 0.001
uS15	102.4 ± 21.4	89.1 ± 2.3	< 0.001
bS16	126.1 ± 35.7	109.3 ± 34.9	< 0.001
uS17	91.8 ± 17.0	88.6 ± 8.7	0.9958
bS18	77.8 ± 13.1	83.0 ± 11.5	< 0.001
uS19	110.2 ± 19.5	92.2 ± 3.9	< 0.001
bS20	99.9 ± 19.3	88.5 ± 7.4	< 0.001
bS21	81.1 ± 20.2	67.1 ± 8.7	< 0.001
uL1	265.9 ± 47.9	232.2 ± 4.6	< 0.001
uL2	274.9 ± 18.8	276.5 ± 5.6	< 0.001
uL3	219.3 ± 38.4	217.9 ± 16.5	0.0108
uL4	229.3 ± 25.9	210.2 ± 16.5	< 0.001
uL5	187.7 ± 17.3	183.3 ± 7.2	< 0.001
uL6	187.9 ± 15.4	178.9 ± 3.0	< 0.001
bL9	153.0 ± 17.2	159.6 ± 44.0	0.0561
uL10	183.6 ± 19.3	173.1 ± 21.2	< 0.001
uL11	146.1 ± 13.1	143.7 ± 7.8	0.3477
bL12	141.0 ± 21.7	125.5 ± 23.6	< 0.001
uL13	136.0 ± 18.2	146.5 ± 6.7	< 0.001
uL14	126.4 ± 10.8	122.1 ± 1.6	< 0.001
uL15	150.5 ± 24.1	150.4 ± 12.4	< 0.001
uL16	140.7 ± 13.6	139.7 ± 3.6	0.0671
bL17	132.9 ± 21.0	141.5 ± 27.3	< 0.001
uL18	121.4 ± 13.9	120.1 ± 4.9	0.0076

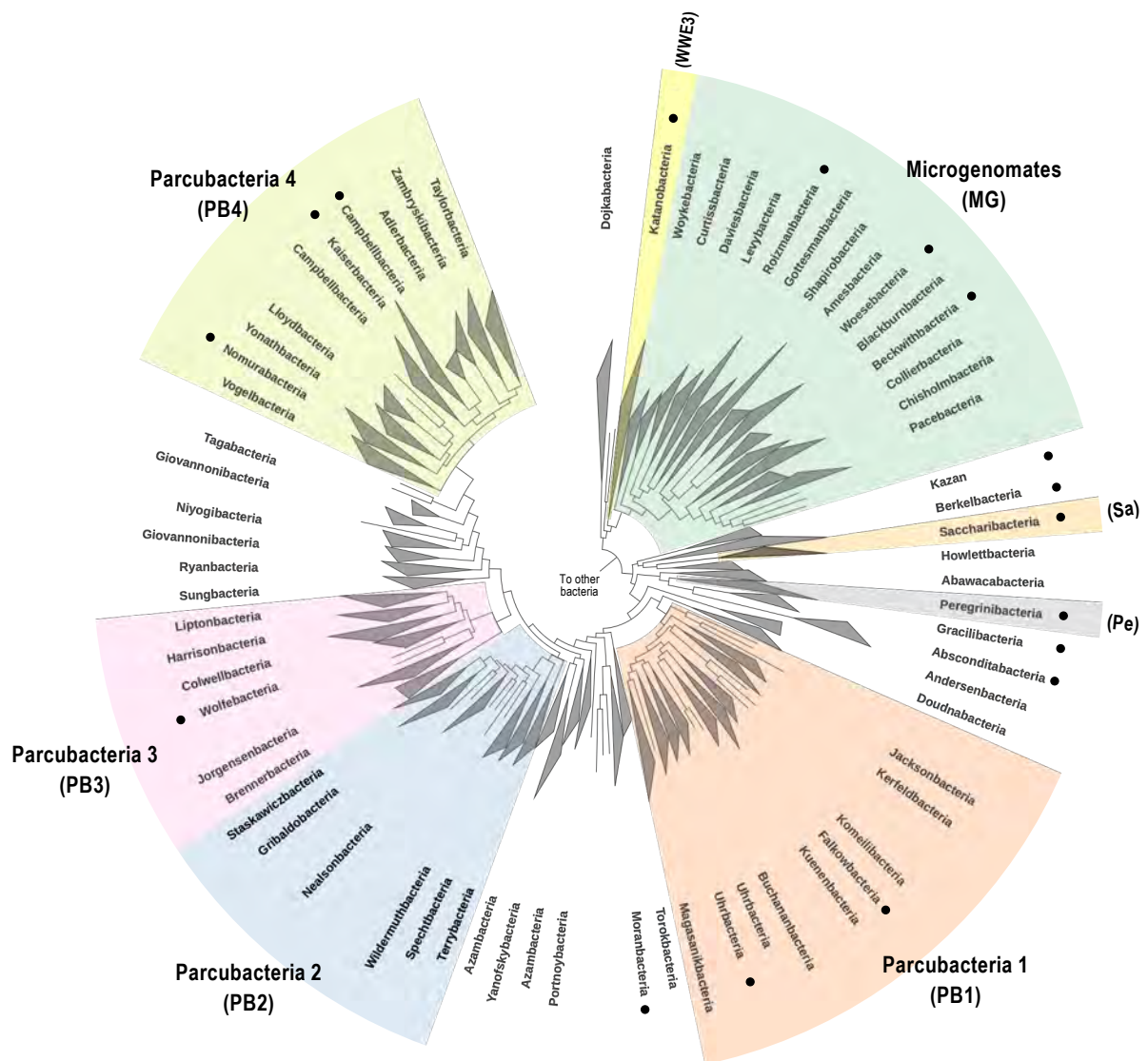
bL19	145.5 ± 32.5	121.8 ± 12.4	< 0.001
bL20	116.7 ± 9.3	120.9 ± 10.3	< 0.001
bL21	115.5 ± 17.1	112.0 ± 26.5	< 0.001
uL22	149.9 ± 28.9	120.6 ± 16.6	< 0.001
uL23	121.5 ± 26.8	100.2 ± 10.3	< 0.001
uL24	110.2 ± 21.4	105.6 ± 11.7	0.0013
bL25	238.7 ± 27.5	194.4 ± 37.7	< 0.001
bL27	93.7 ± 15.1	89.0 ± 6.9	< 0.001
bL28	83.4 ± 24.1	74.8 ± 11.9	< 0.001
uL29	73.4 ± 11.7	71.7 ± 28.0	< 0.001
bL31	103.7 ± 24.5	78.6 ± 15.9	< 0.001
bL32	78.5 ± 23.3	60.0 ± 5.9	< 0.001
bL33	60.8 ± 11.1	53.2 ± 5.1	< 0.001
bL34	54.6 ± 14.6	45.7 ± 4.4	< 0.001
bL35	70.2 ± 11.3	65.6 ± 11.3	< 0.001
bL36	44.0 ± 12.9	38.6 ± 2.6	< 0.001

Supplementary Table S7. Continued

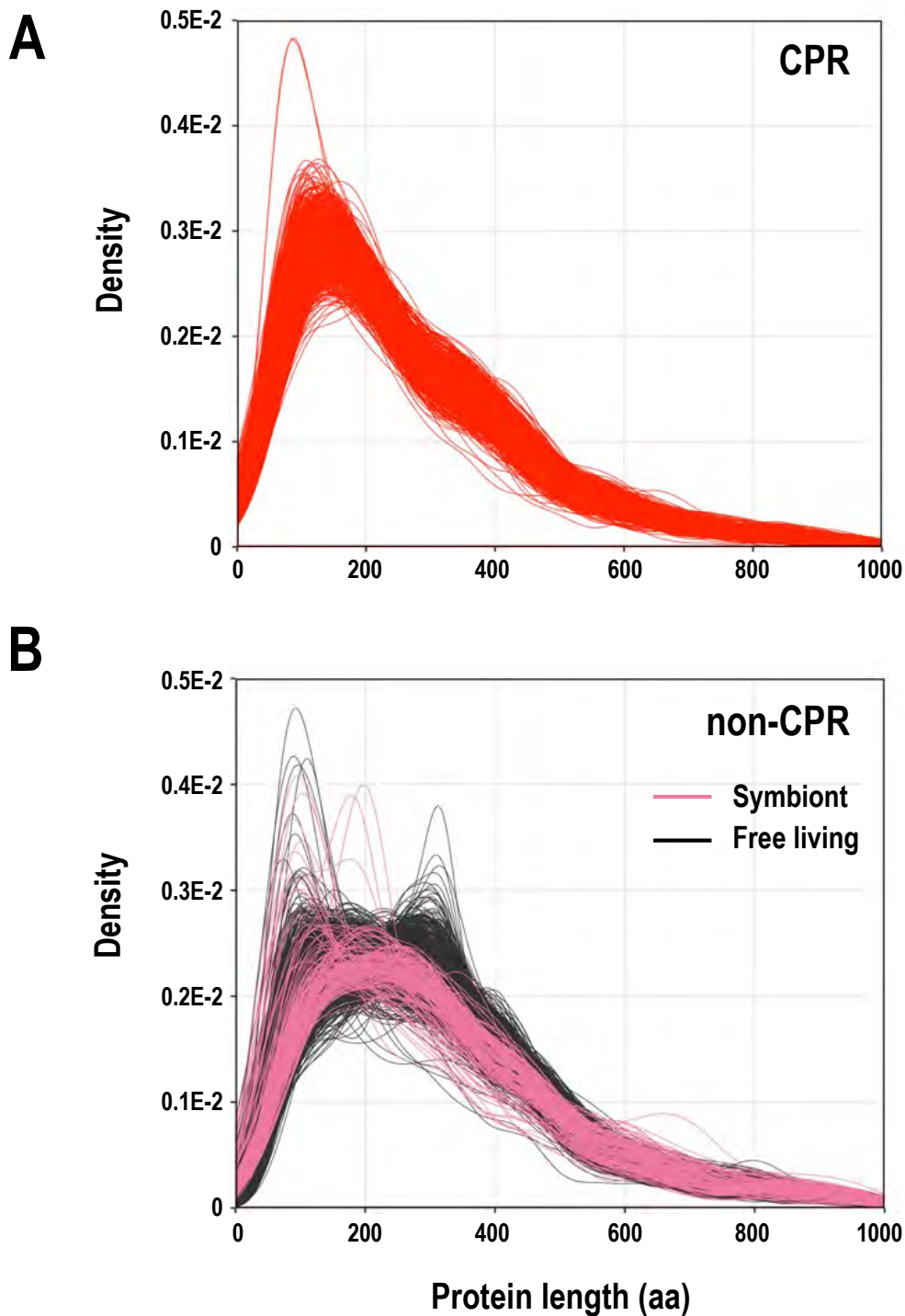
Supplementary Table S8. Ribosomal protein sequence alignments (FASTA format) used to generate Supplementary Figure S4.

Supplementary Table S9. rRNA gene sequence alignments (FASTA format) used to generate Supplementary Figure S5.

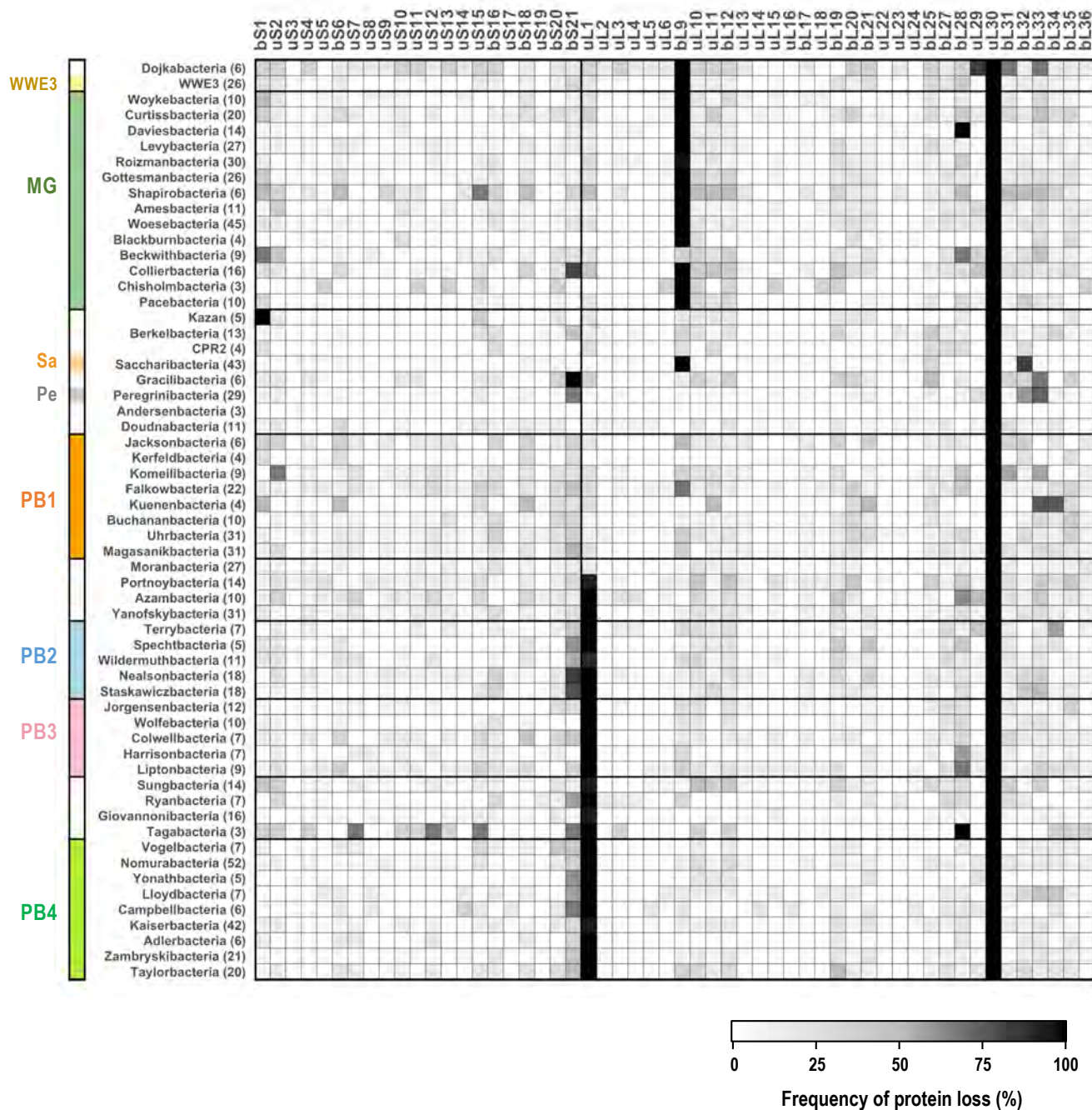
These datasets are supplied as the separate files.



Supplementary Figure S1. Phylogenetic relationships of CPR bacteria. The phylogenetic tree data were acquired from a previous report (Jaffe *et al.* 2020). Parcubacteria 1-4 (PB1-4), which are monophyletic subgroups within the Parcubacteria superphylum, are highlighted in different colors, as is the Microgenomates superphylum (MG). Three representative phyla outside Parcubacteria and Microgenomates are also highlighted: Katanobacteria (formerly known as WWE3); Saccharibacteria (Sa); and Peregrinibacteria (Pe). Phyla with complete genomes included in our study are marked with black dots.



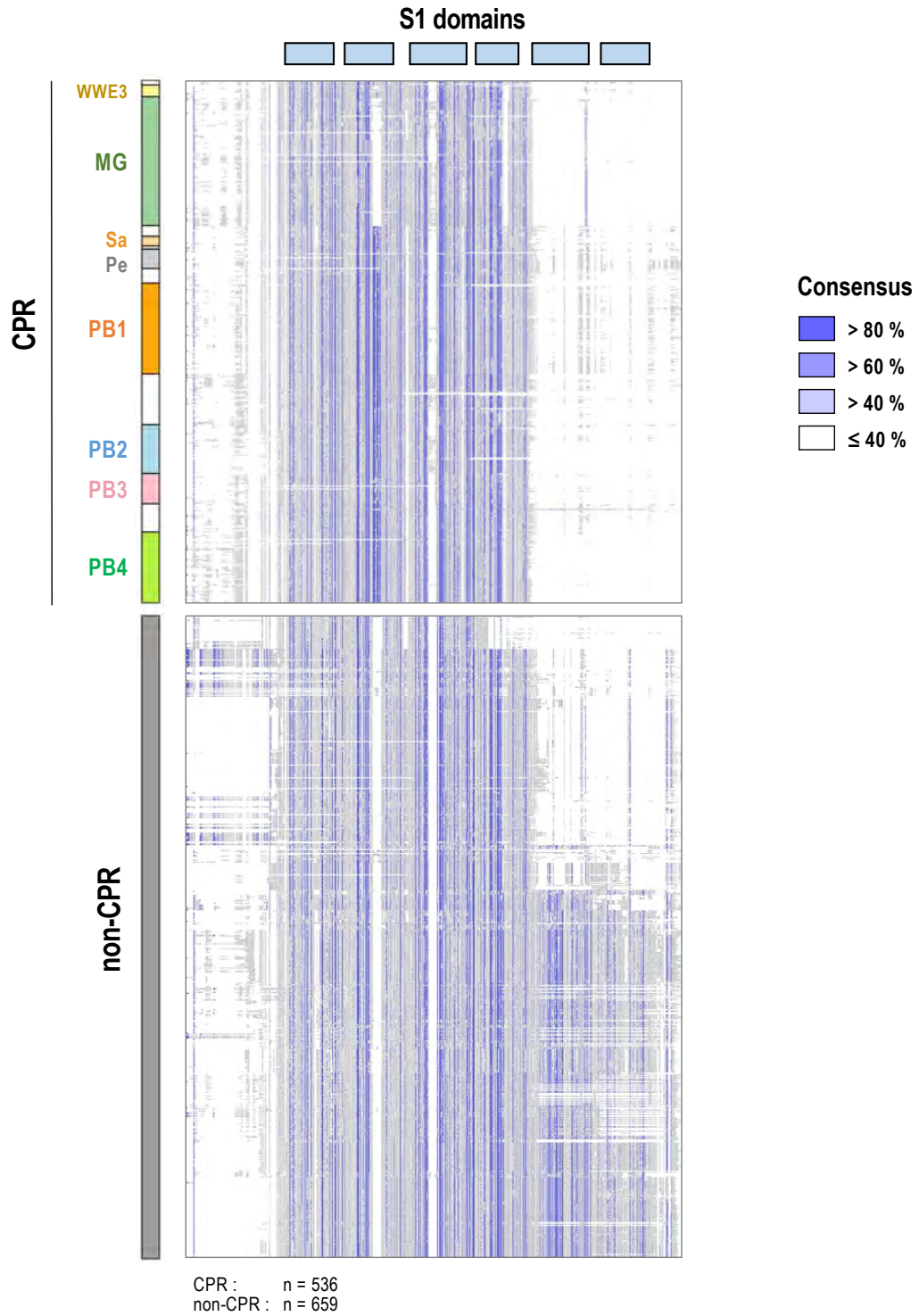
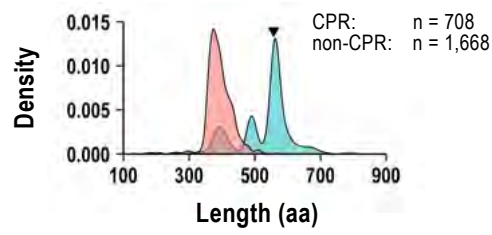
Supplementary Figure S2. Distribution of lengths of proteins encoded in 897 CPR and 1,661 non-CPR bacterial genomes. (A) Distribution of the lengths of proteins encoded in 897 CPR genomes (69 complete and 828 draft genomes). (B) Distribution of the lengths of proteins encoded in 1,661 non-CPR genomes (167 symbiotic [pink lines] and 1,494 free-living [black lines] species). Distributions of the lengths of proteins are shown as density curves based on the deduced amino acid sequences of the protein genes in each genome.



Supplementary Figure S3. Phylum-level distribution of each ribosomal protein in CPR bacteria.

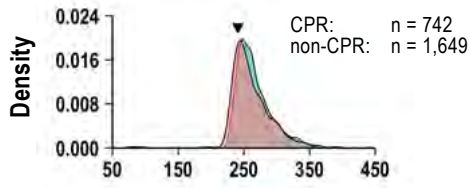
The distributions of 21 SSU and 33 LSU ribosomal proteins (columns) across 59 CPR phyla (rows) for which at least three genomes were available. The white-to-black gradient indicates the frequency of the lack of a ribosomal protein in each phylum. The rows were sorted based on a phylogenetic tree (Jaffe *et al.* 2020), and the number given to the right of each phylum name is the number of genomes belonging to that phylum. Left panel indicates CPR subgroups, with the colors corresponding to those in Supplementary Figure S1. MG: Microgenomates; Sa: Saccharibacteria; Pe: Peregrinibacteria; PB: Parcubacteria.

bS1

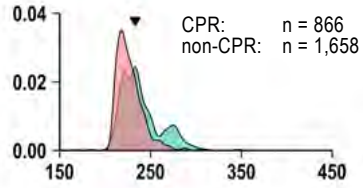


Supplementary Figure S4

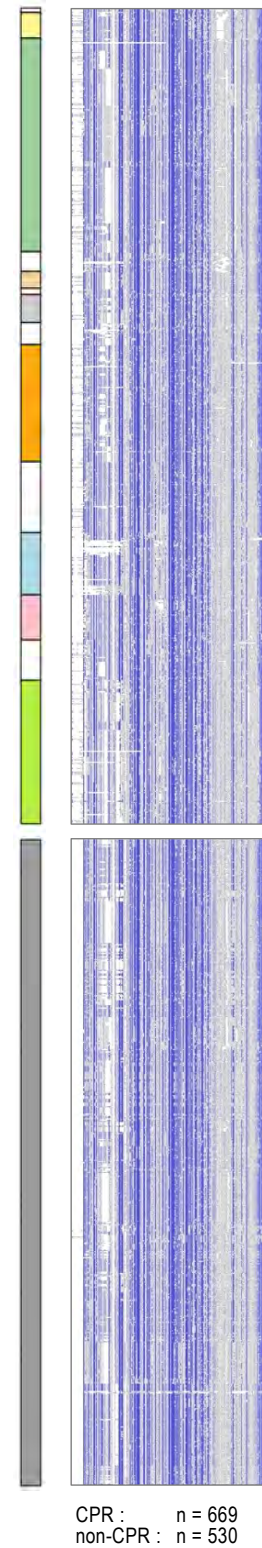
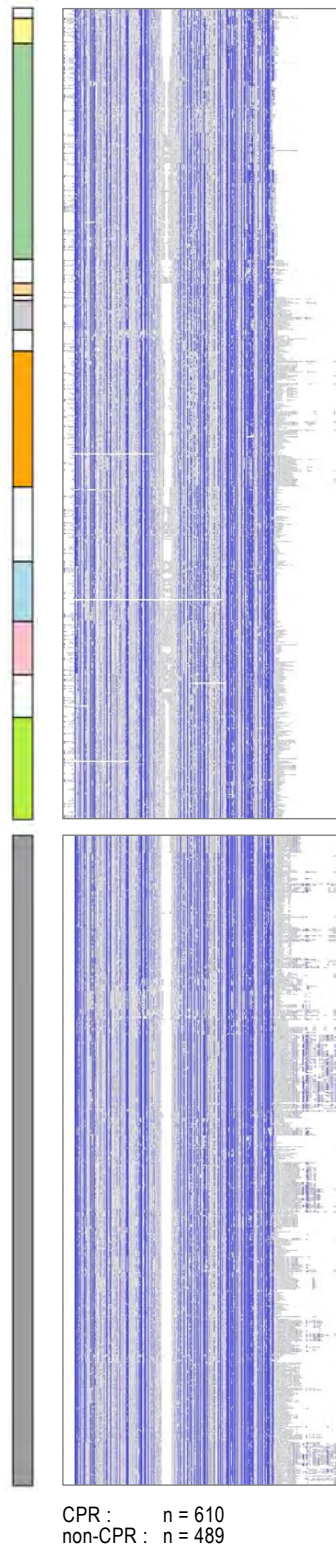
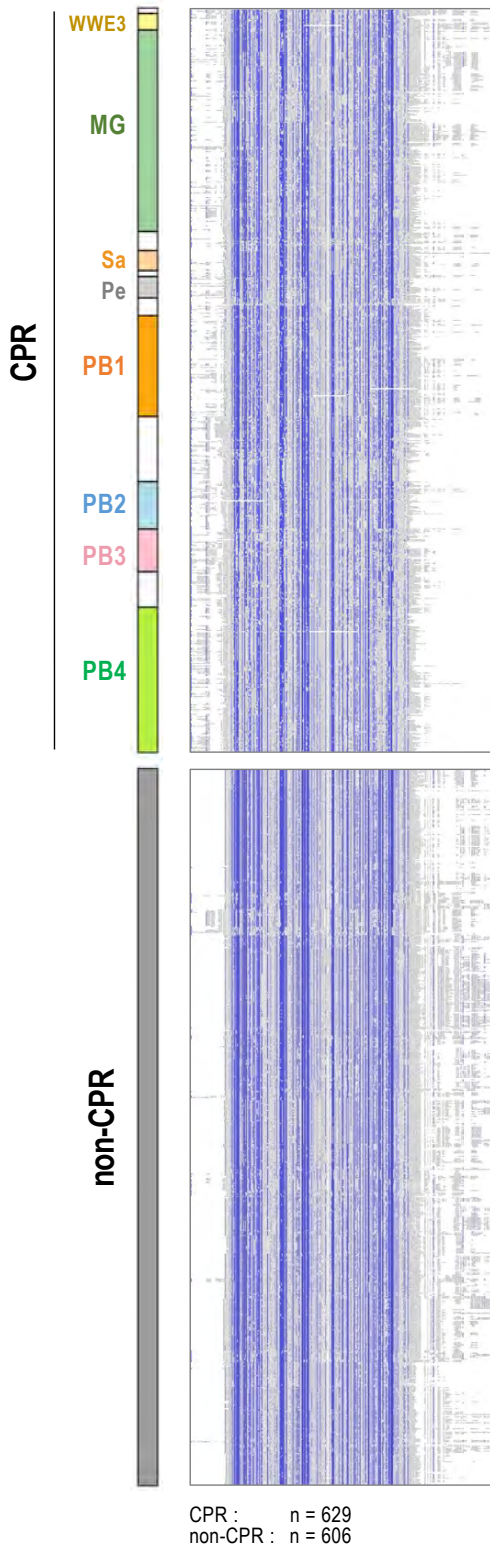
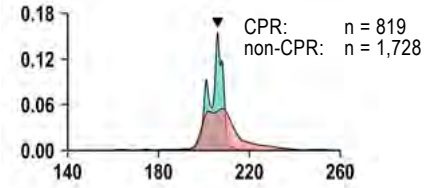
uS2

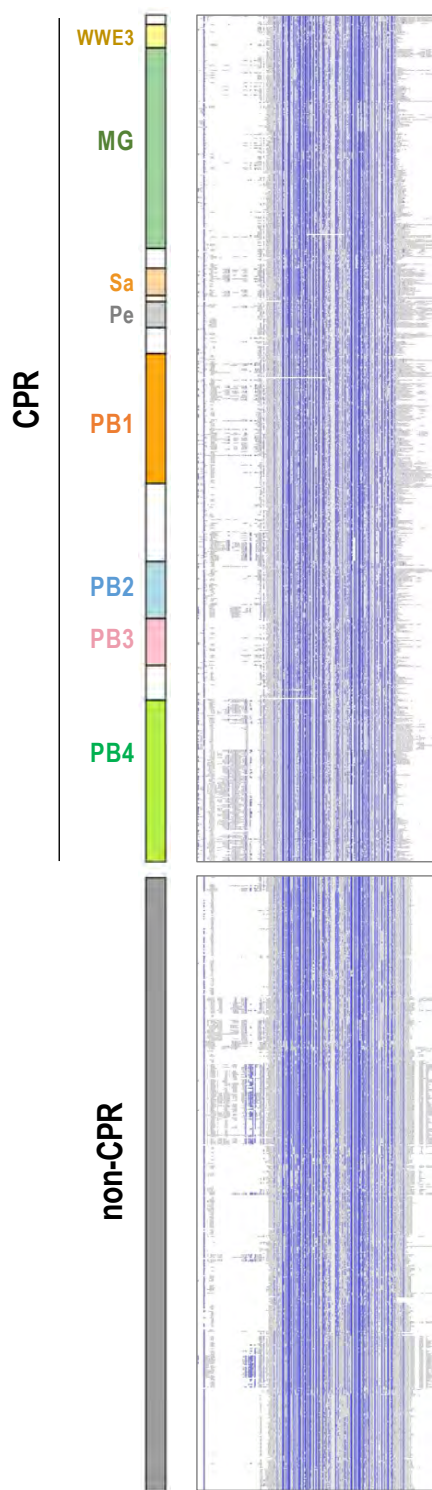
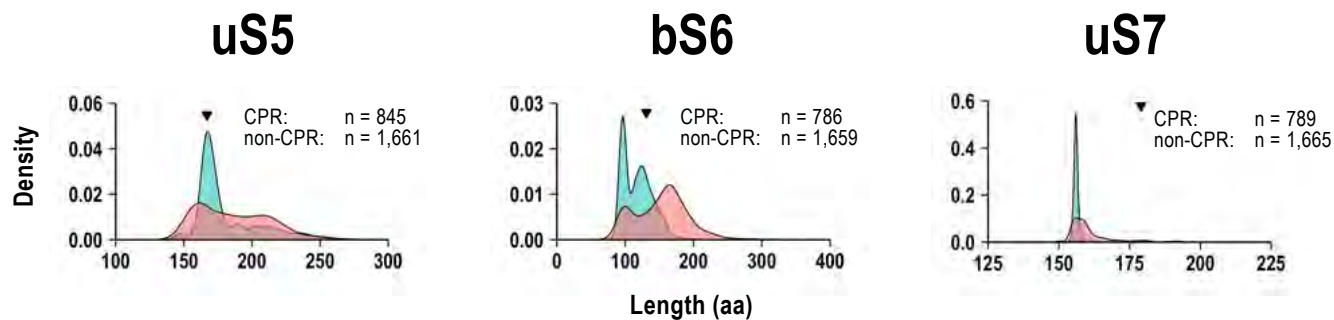


uS3

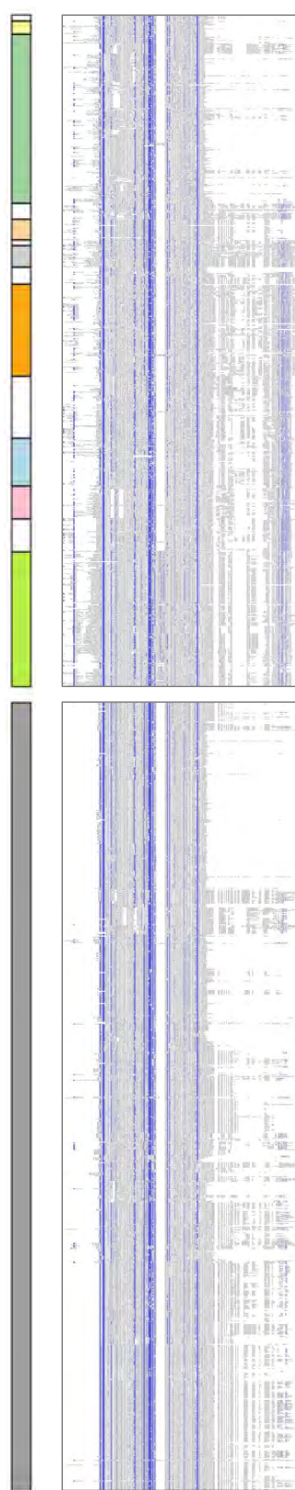


uS4

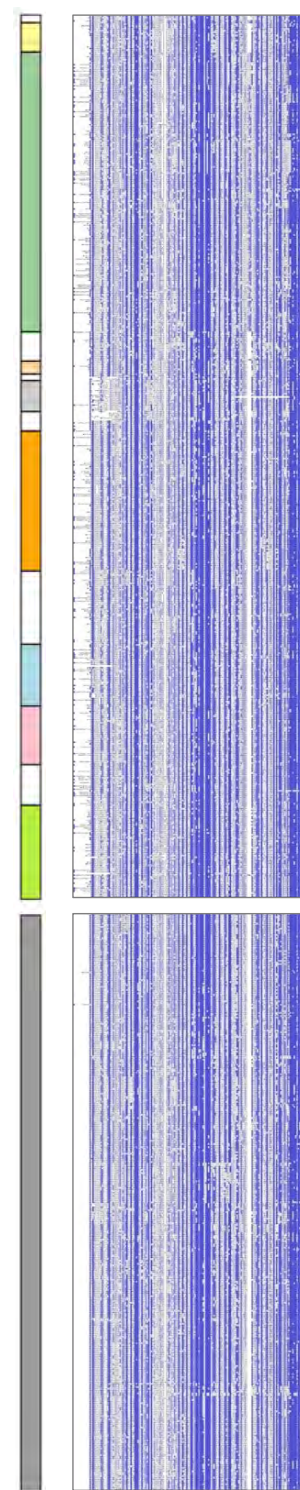




CPR : n = 685
 non-CPR : n = 497

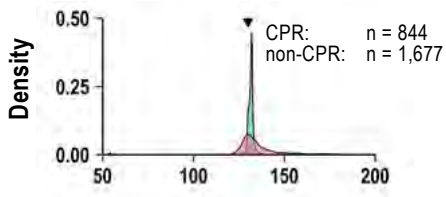


CPR : n = 716
 non-CPR : n = 841

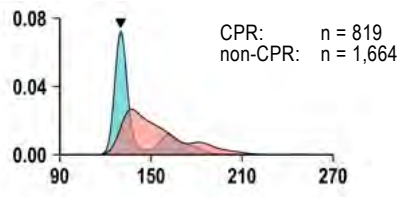


CPR : n = 544
 non-CPR : n = 355

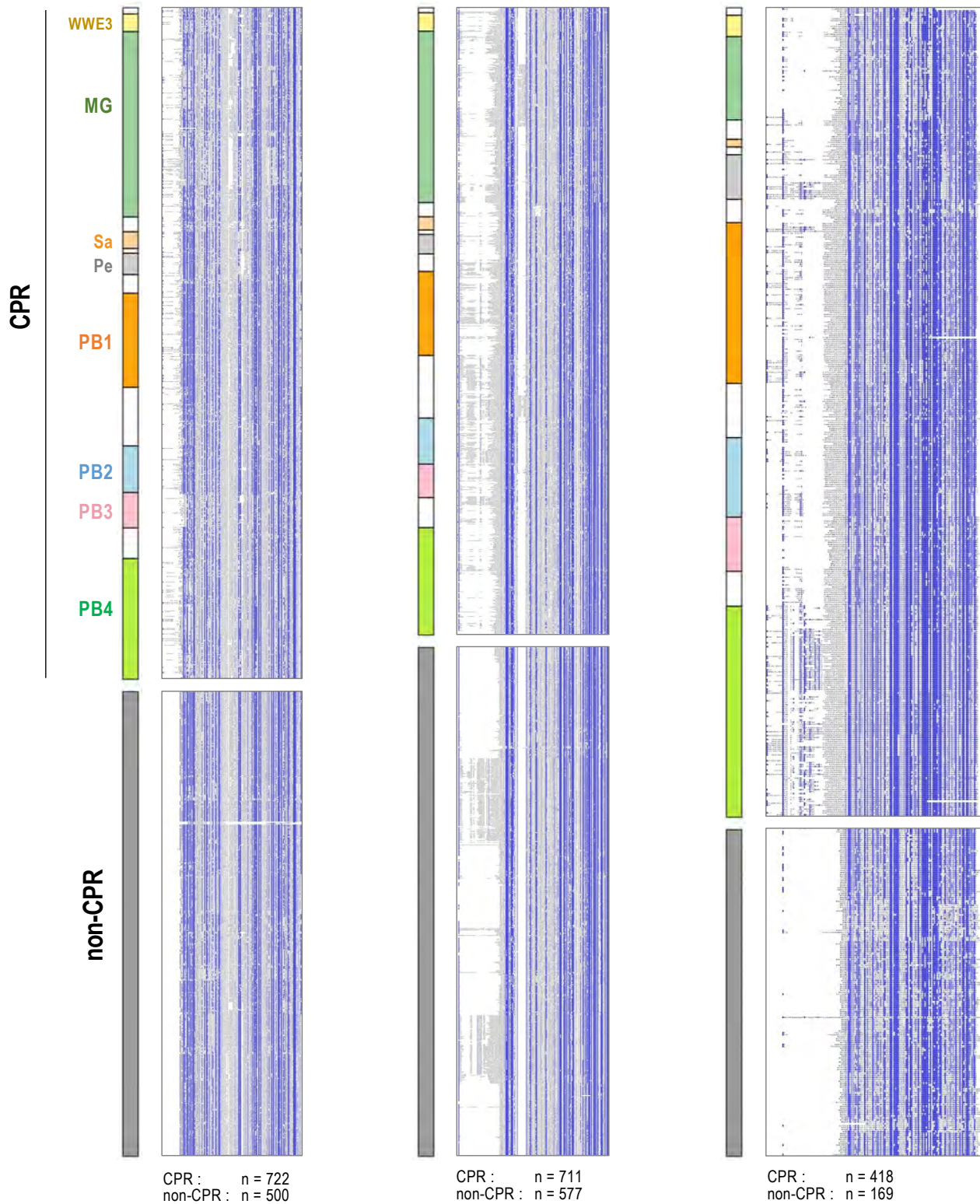
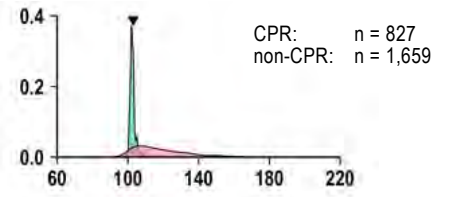
uS8



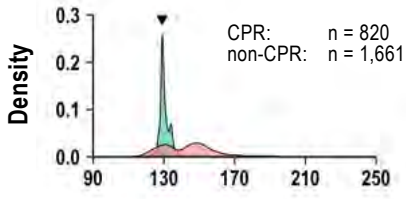
uS9



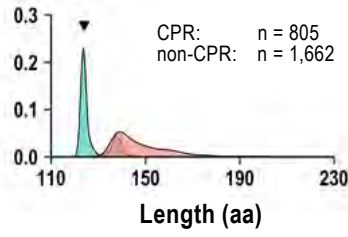
uS10



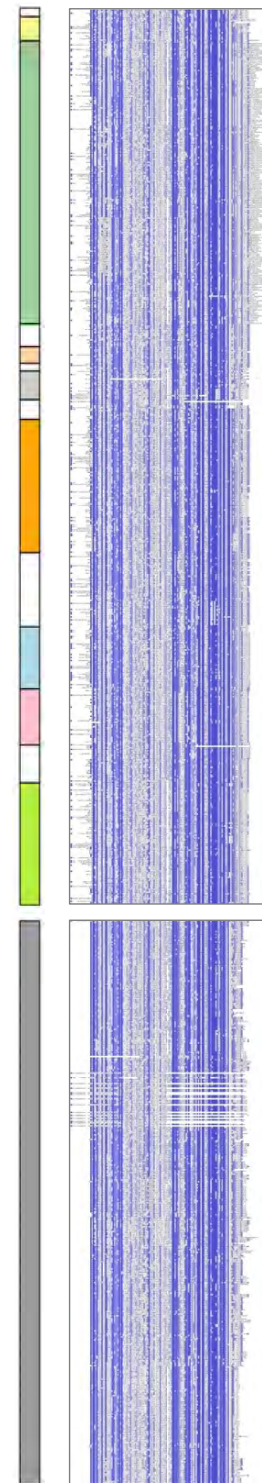
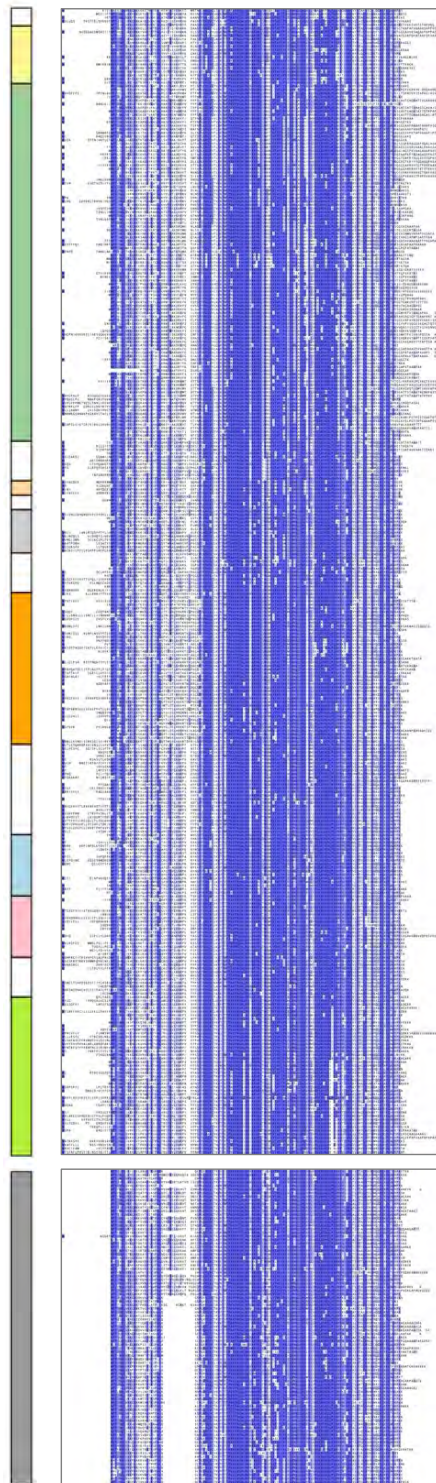
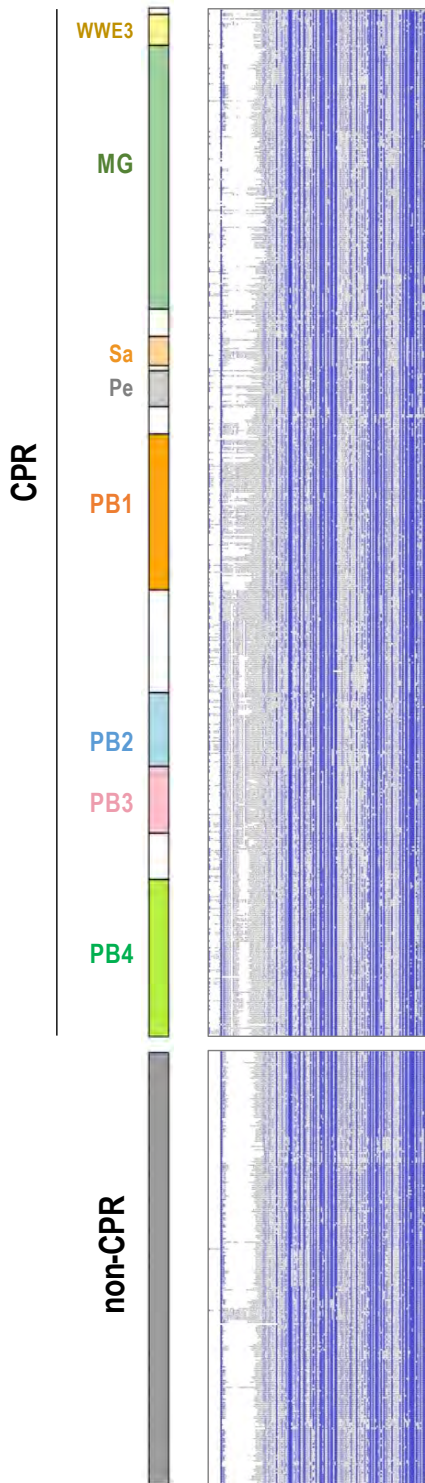
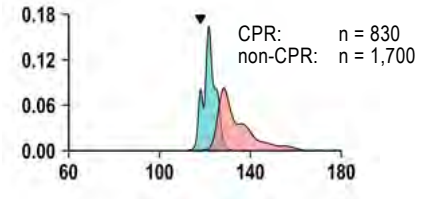
uS11



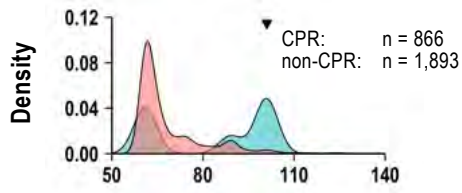
uS12



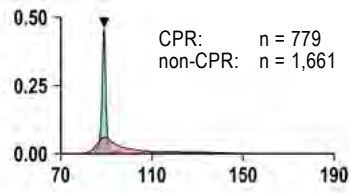
uS13



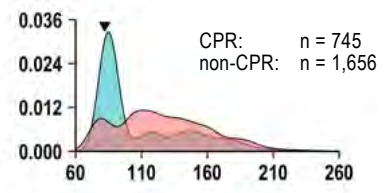
uS14



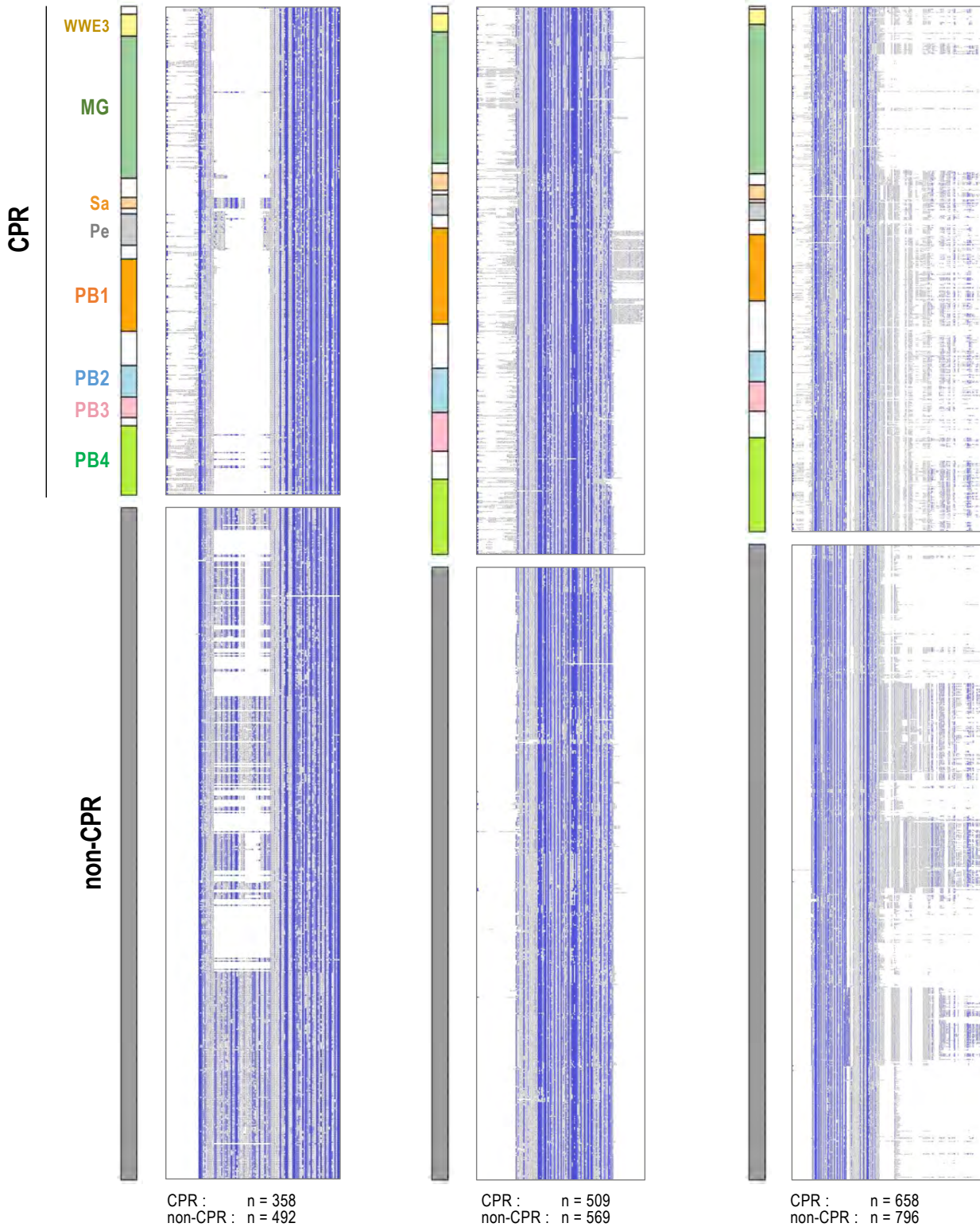
uS15



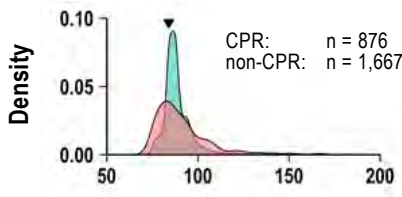
bS16



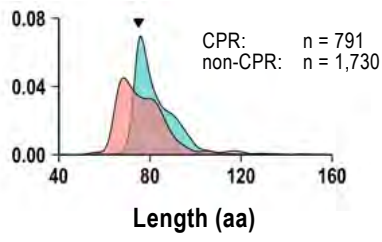
Length (aa)



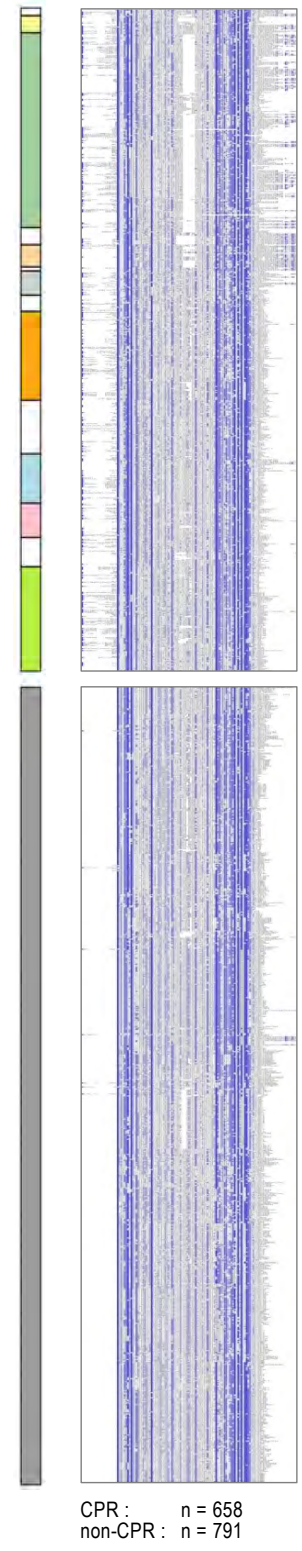
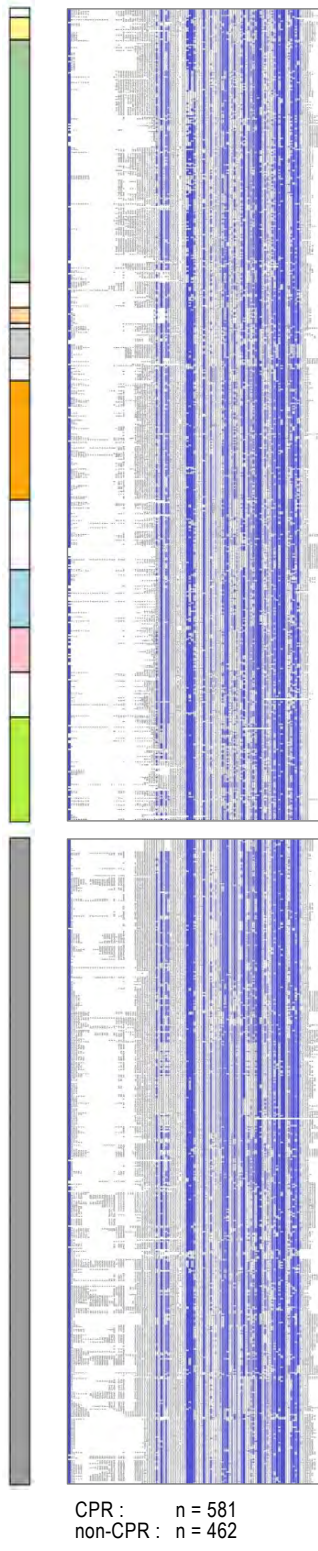
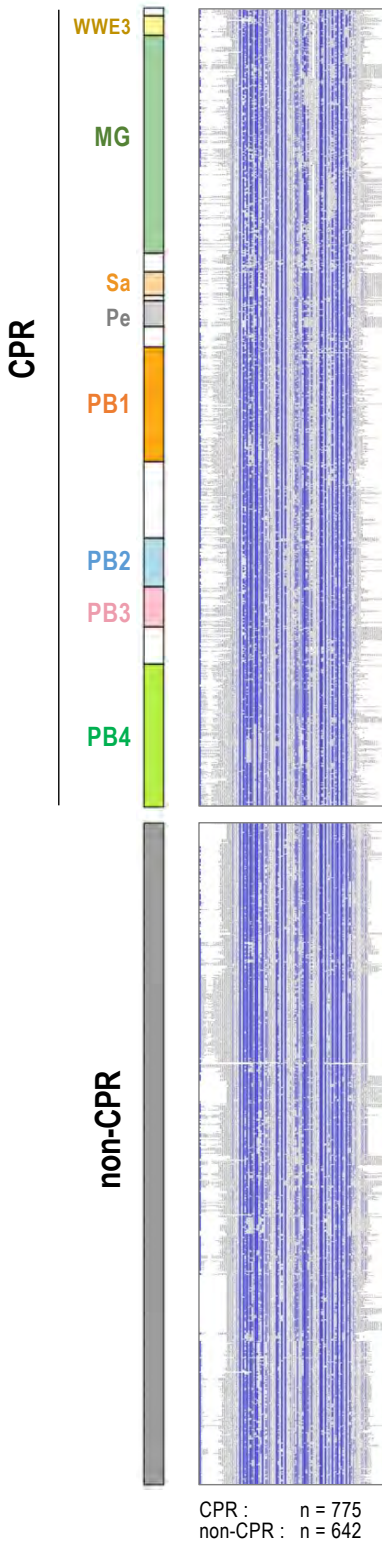
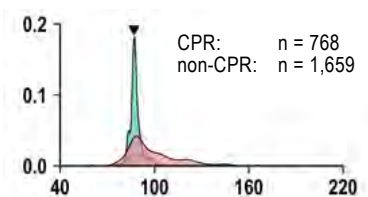
uS17



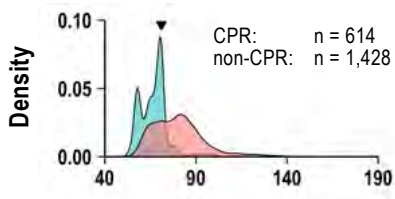
bS18



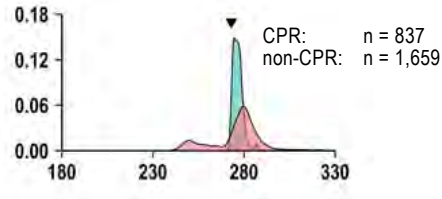
bS20



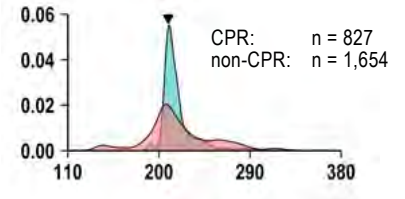
bS21



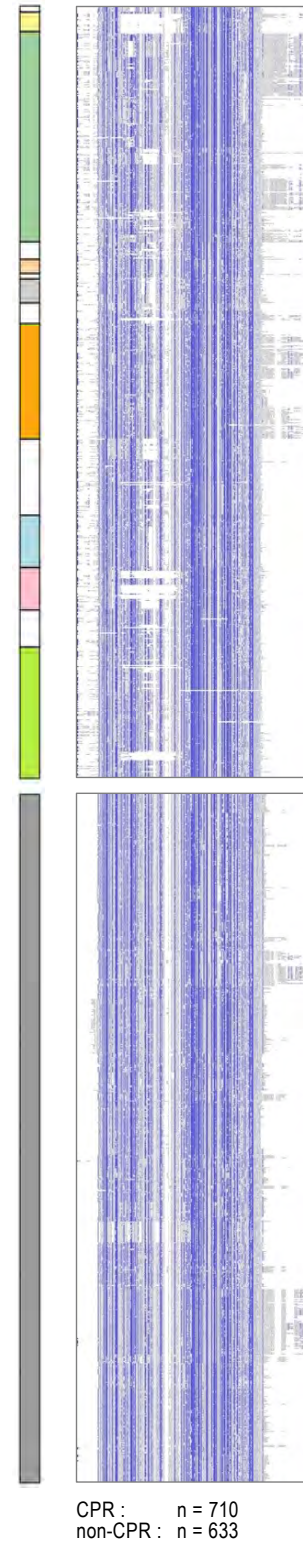
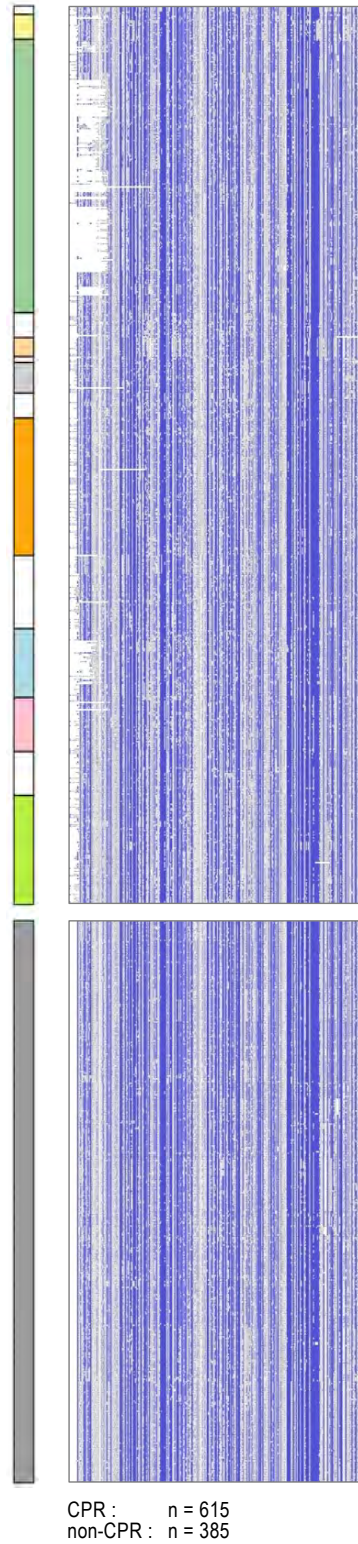
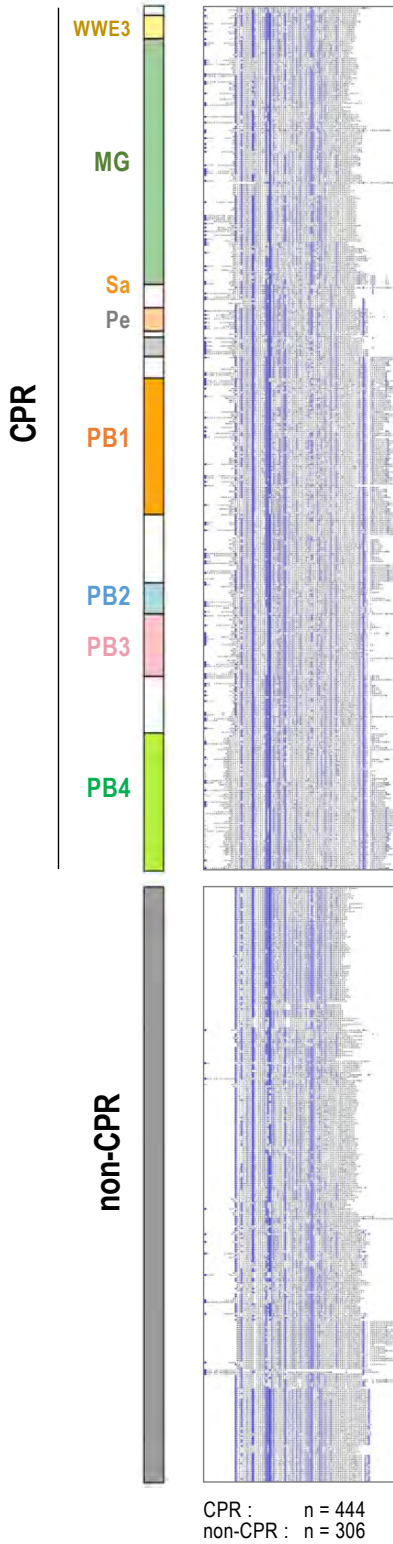
uL2



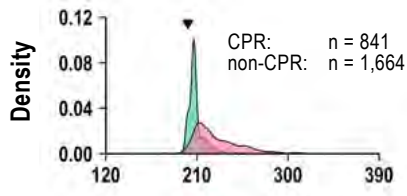
uL3



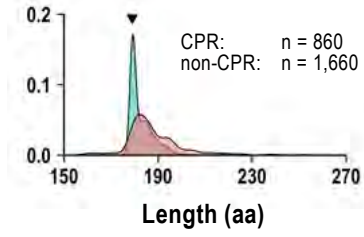
Length (aa)



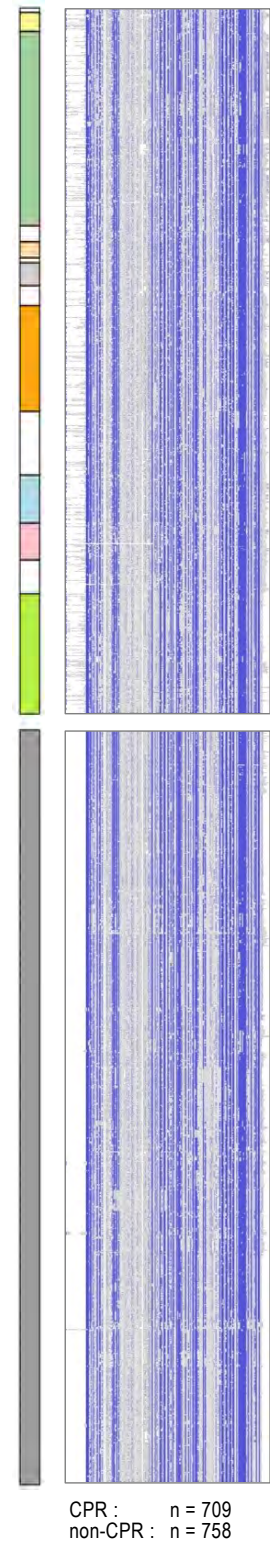
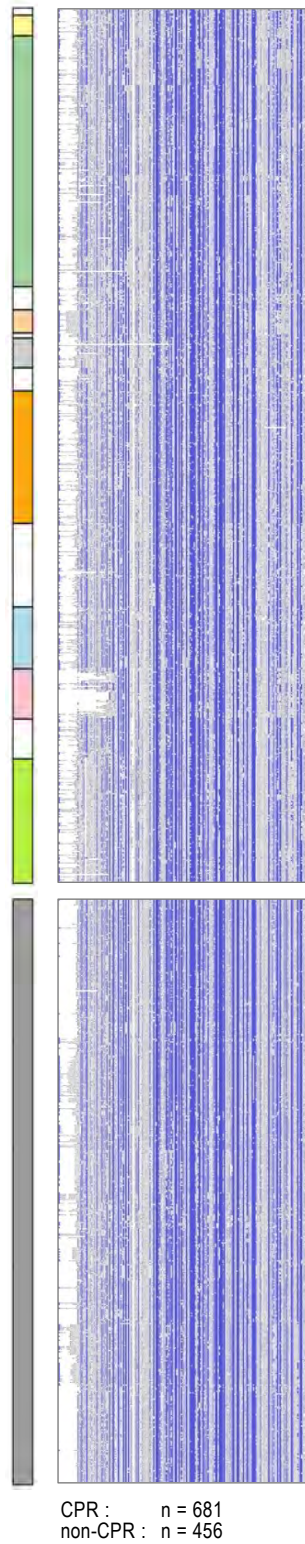
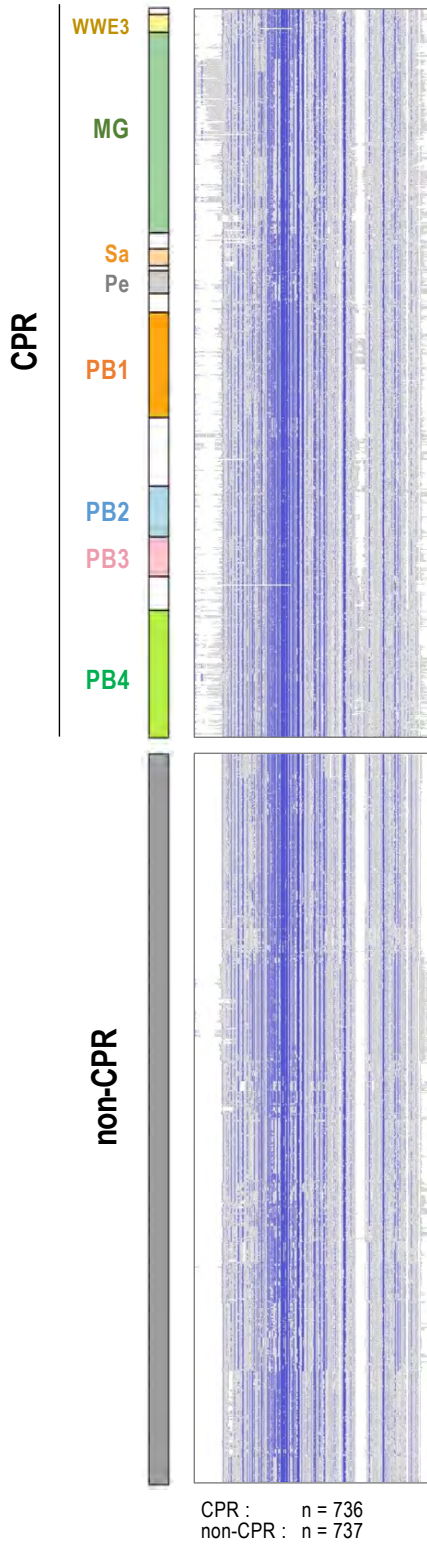
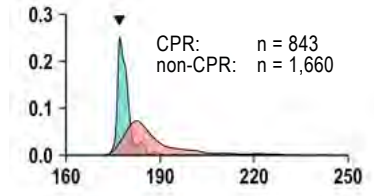
uL4

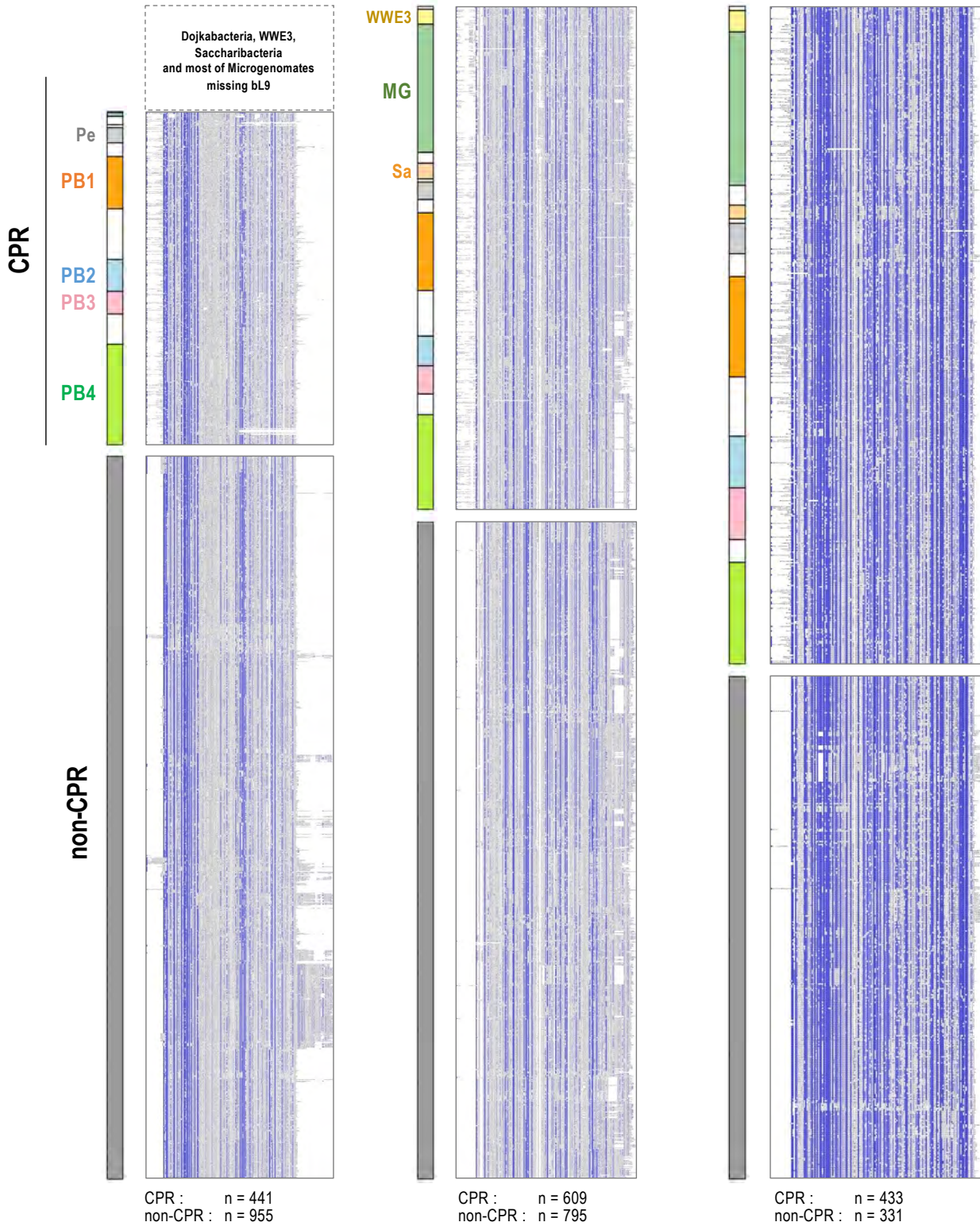
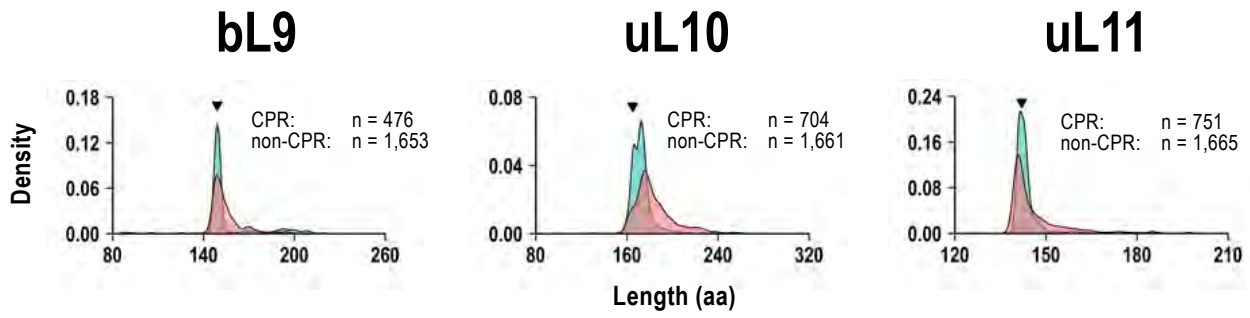


uL5

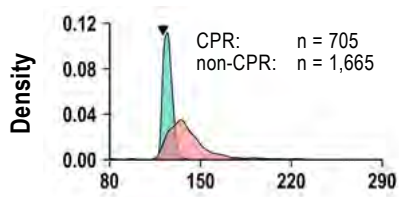


uL6

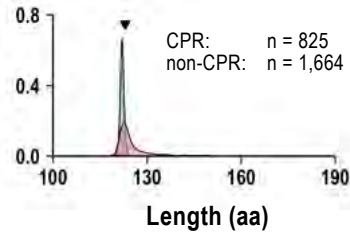




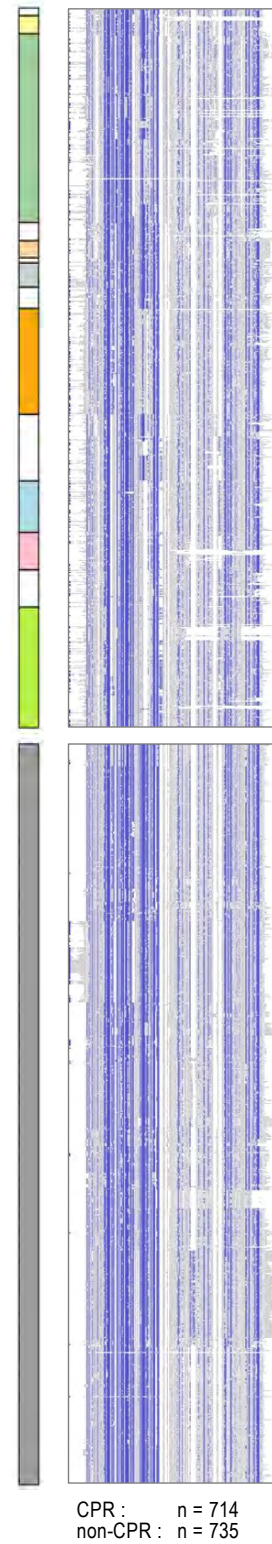
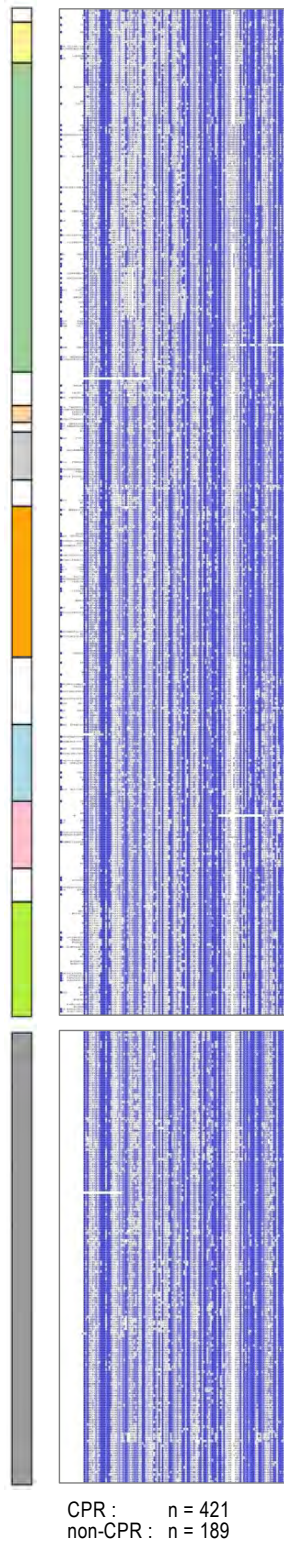
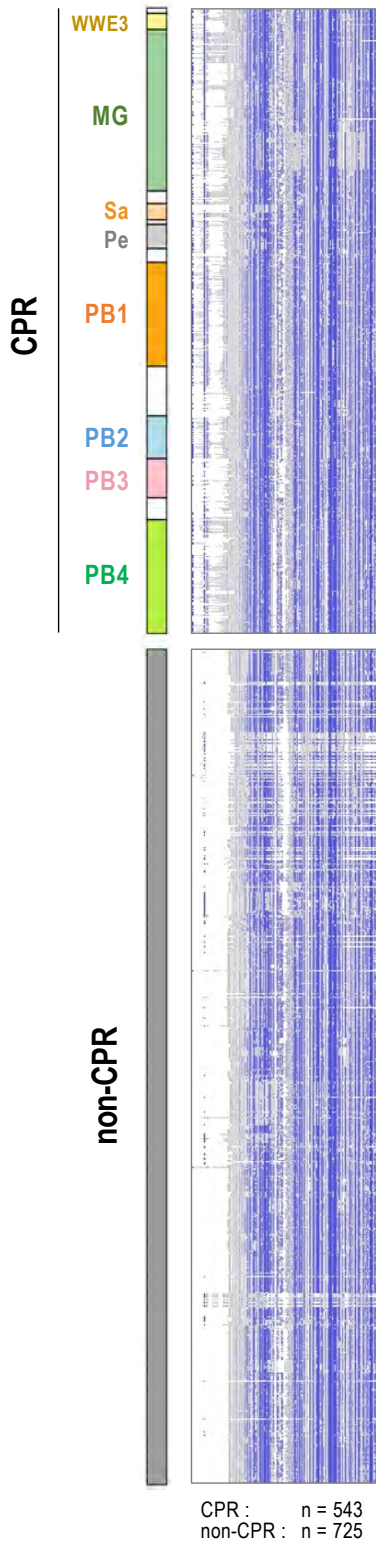
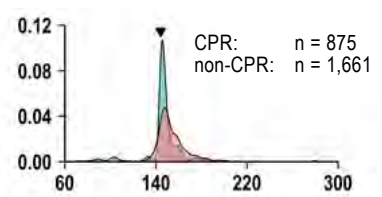
bL12



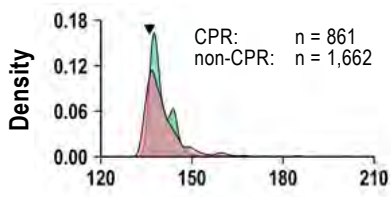
uL14



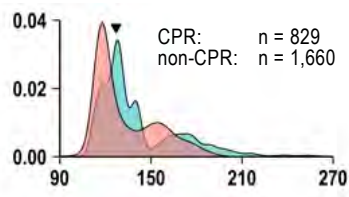
uL15



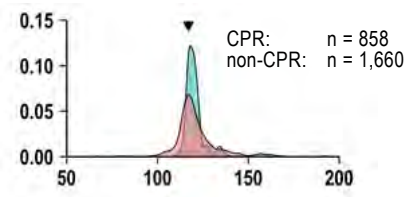
uL16



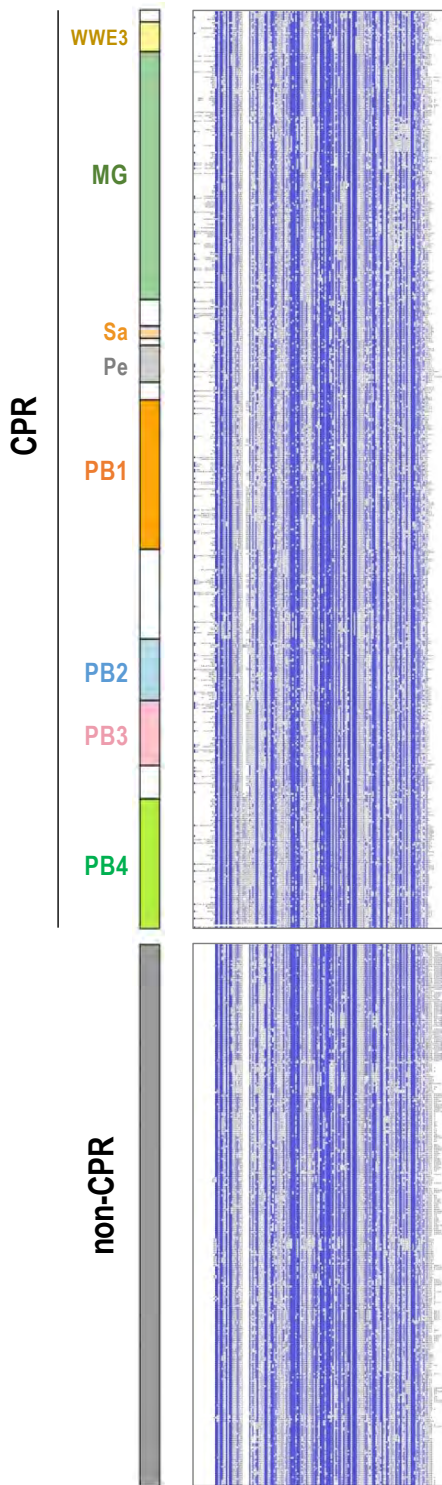
bL17



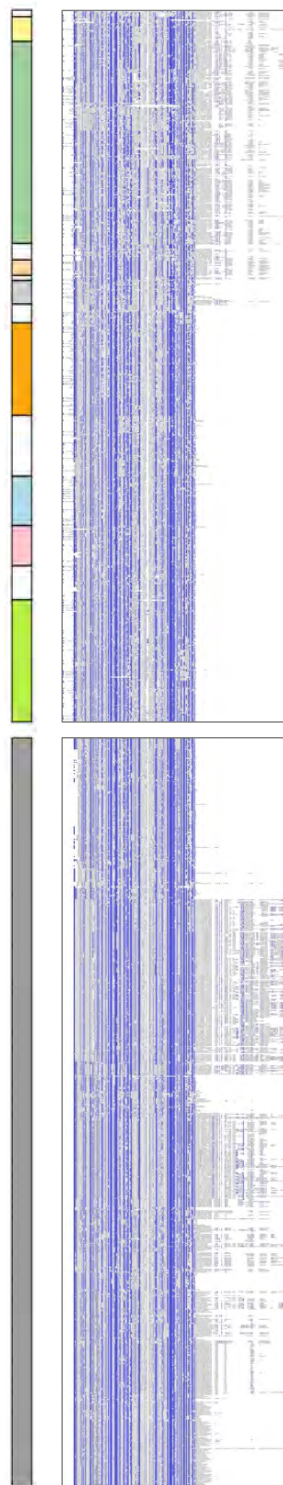
uL18



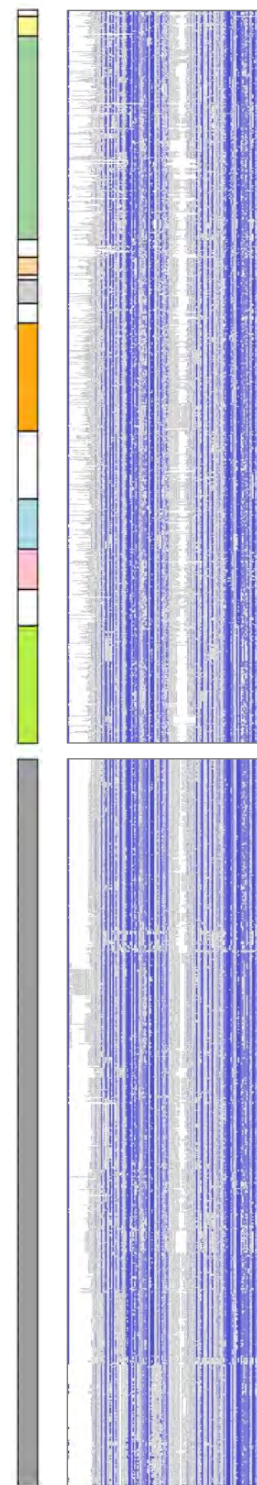
Length (aa)



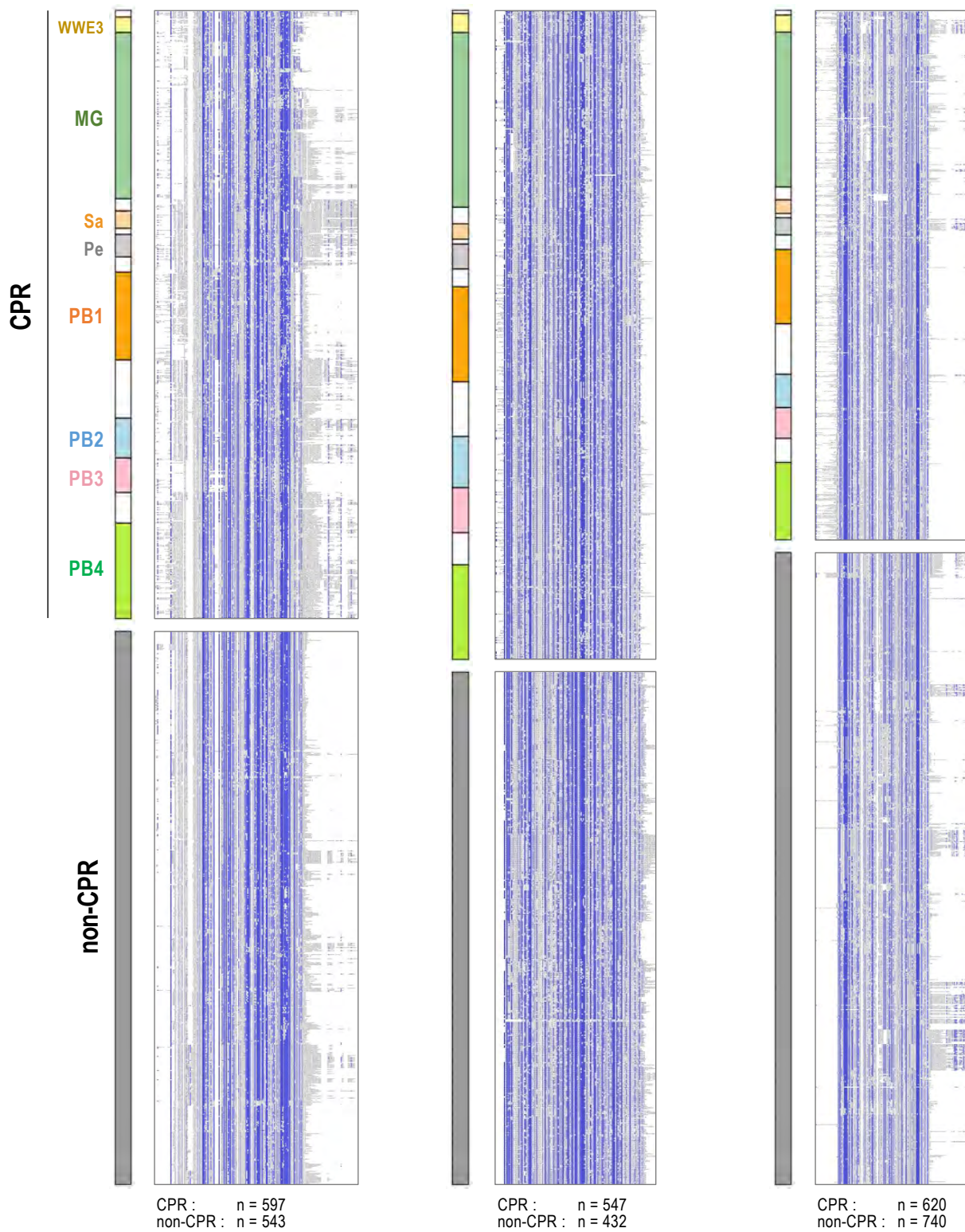
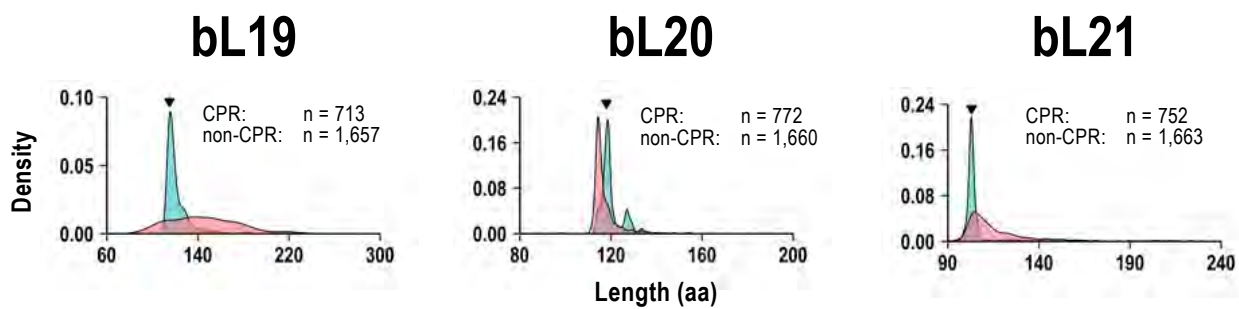
CPR : n = 528
non-CPR : n = 309



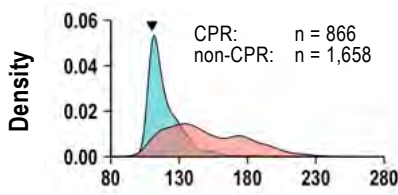
CPR : n = 692
non-CPR : n = 728



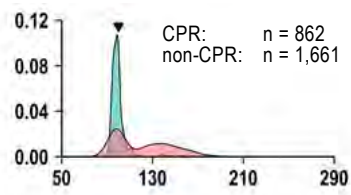
CPR : n = 747
non-CPR : n = 741



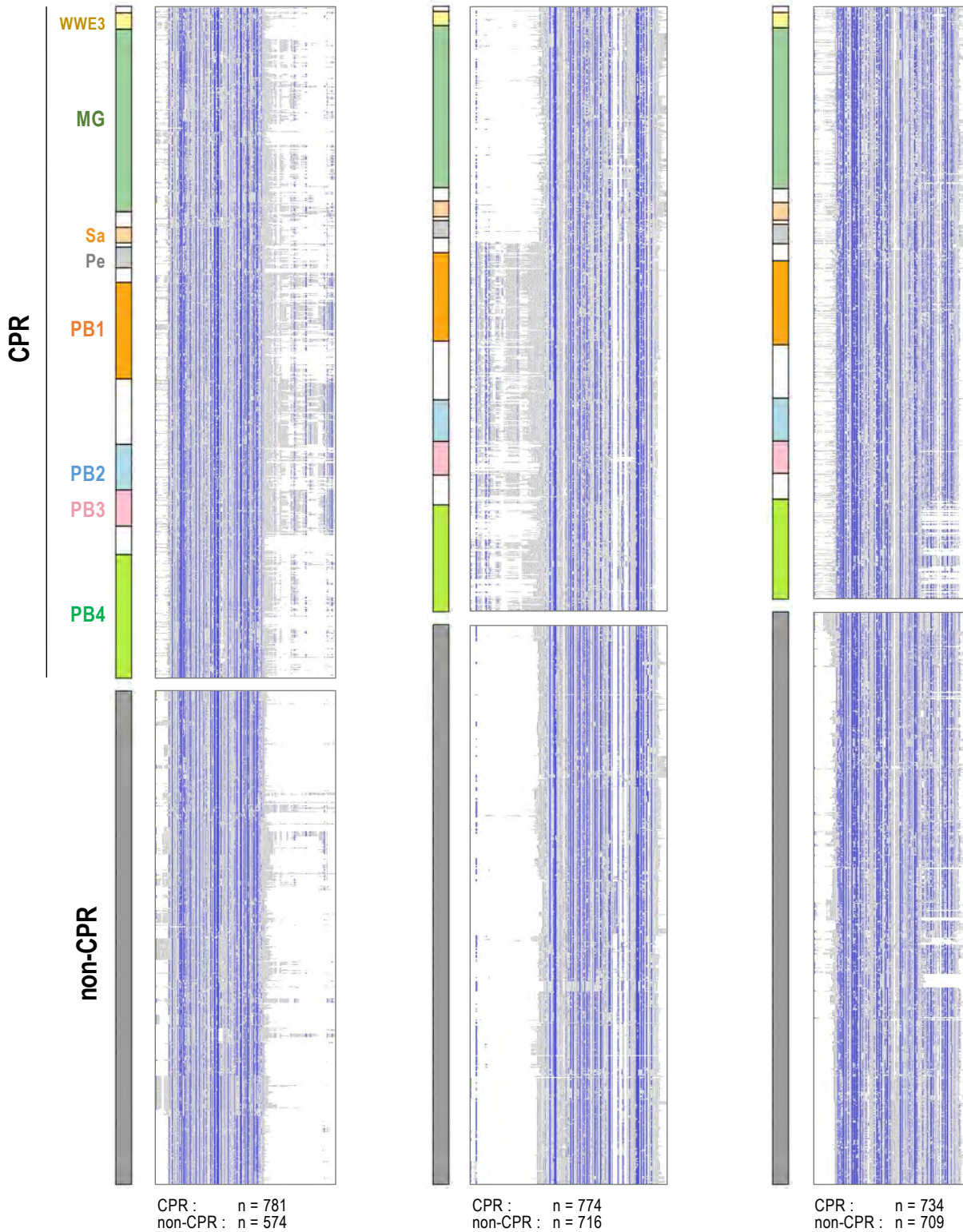
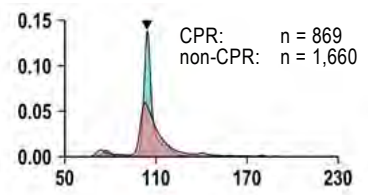
uL22



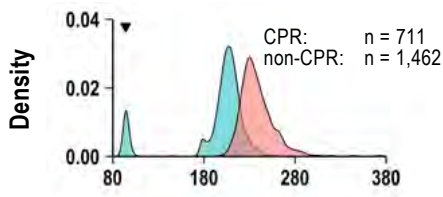
uL23



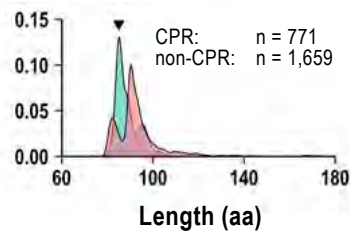
uL24



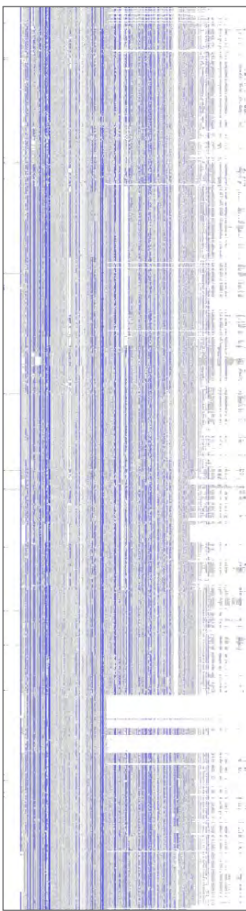
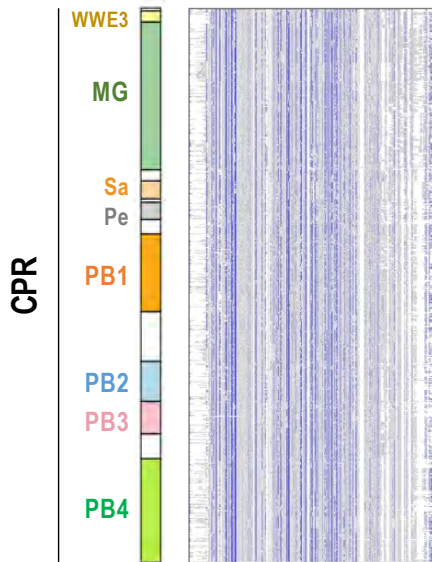
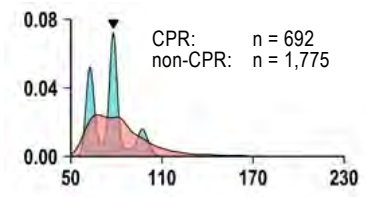
bL25



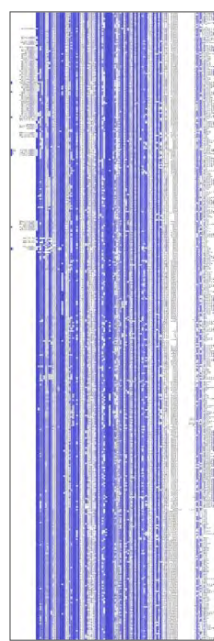
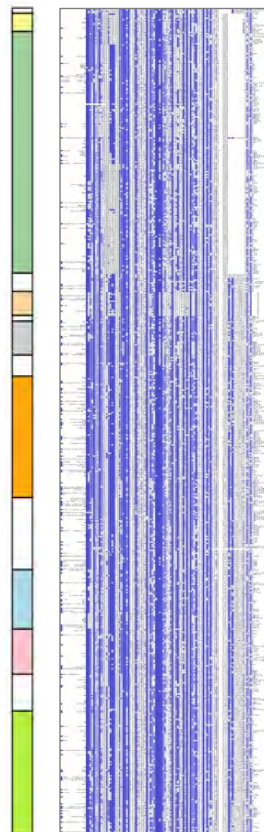
bL27



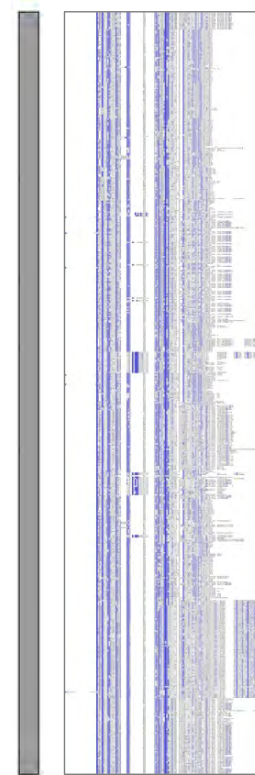
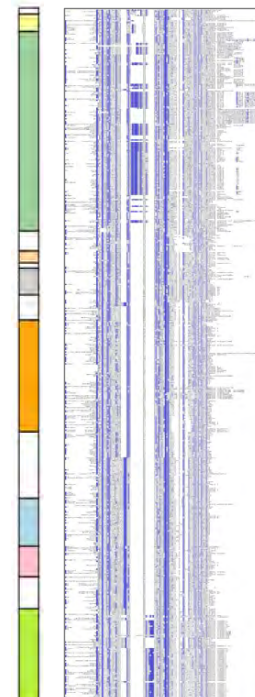
bL28



CPR : n = 650
non-CPR : n = 1059

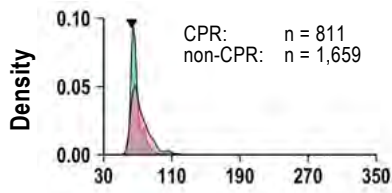


CPR : n = 588
non-CPR : n = 446

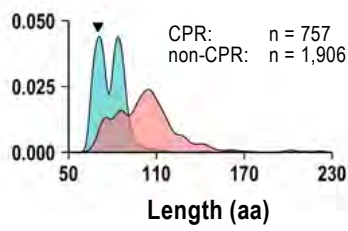


CPR : n = 524
non-CPR : n = 574

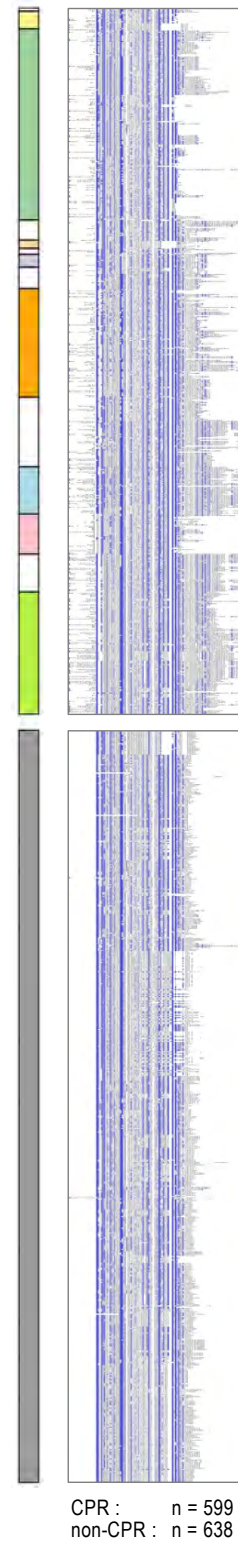
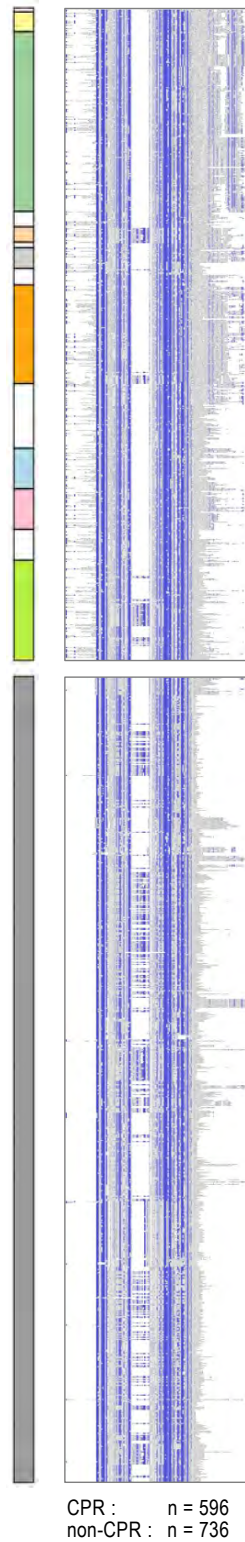
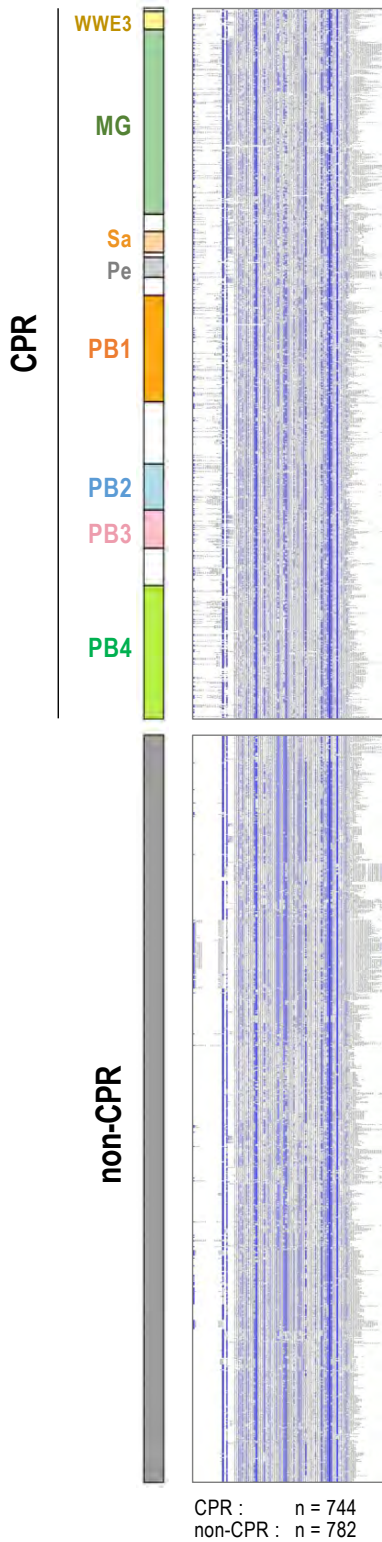
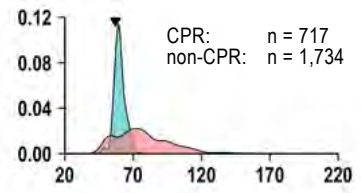
uL29

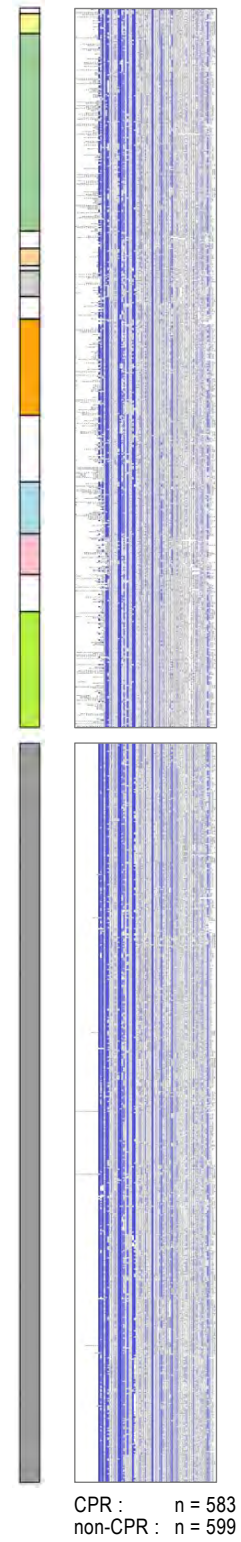
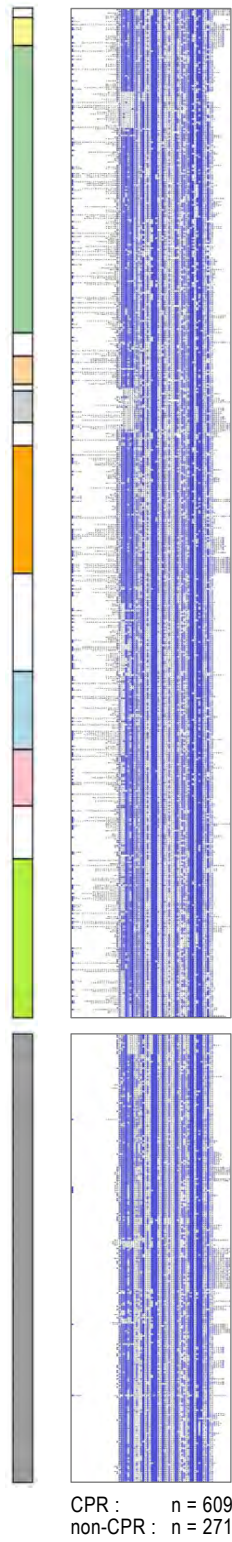
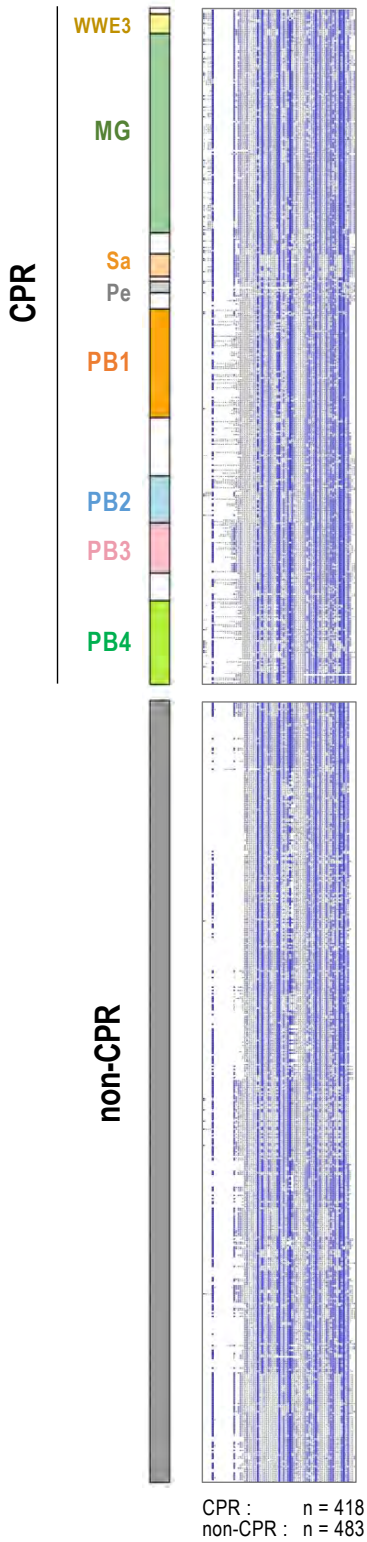
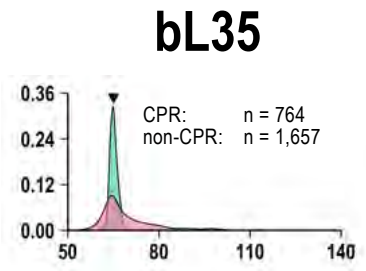
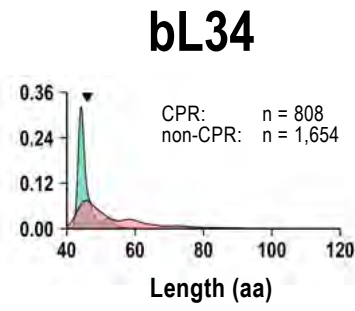
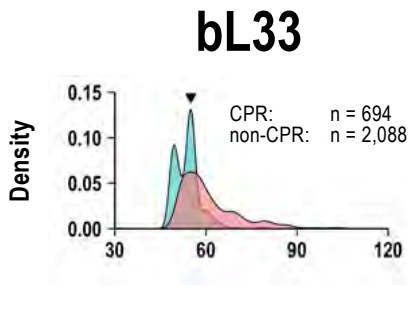


bL31

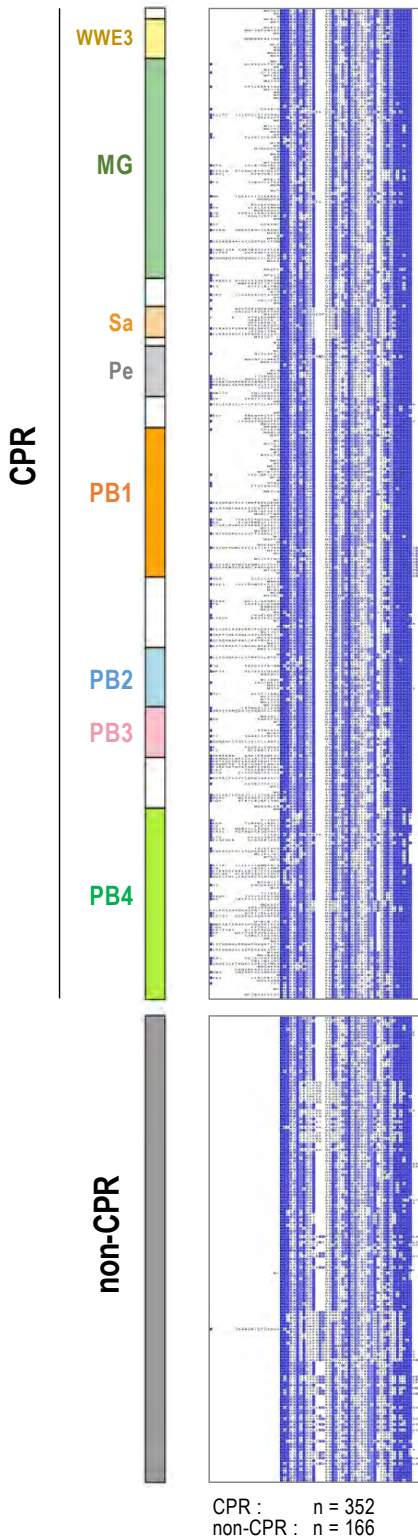
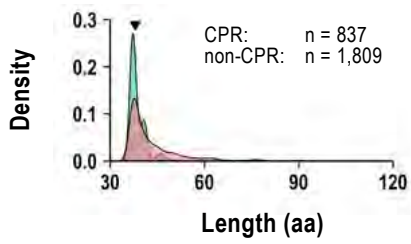


bL32

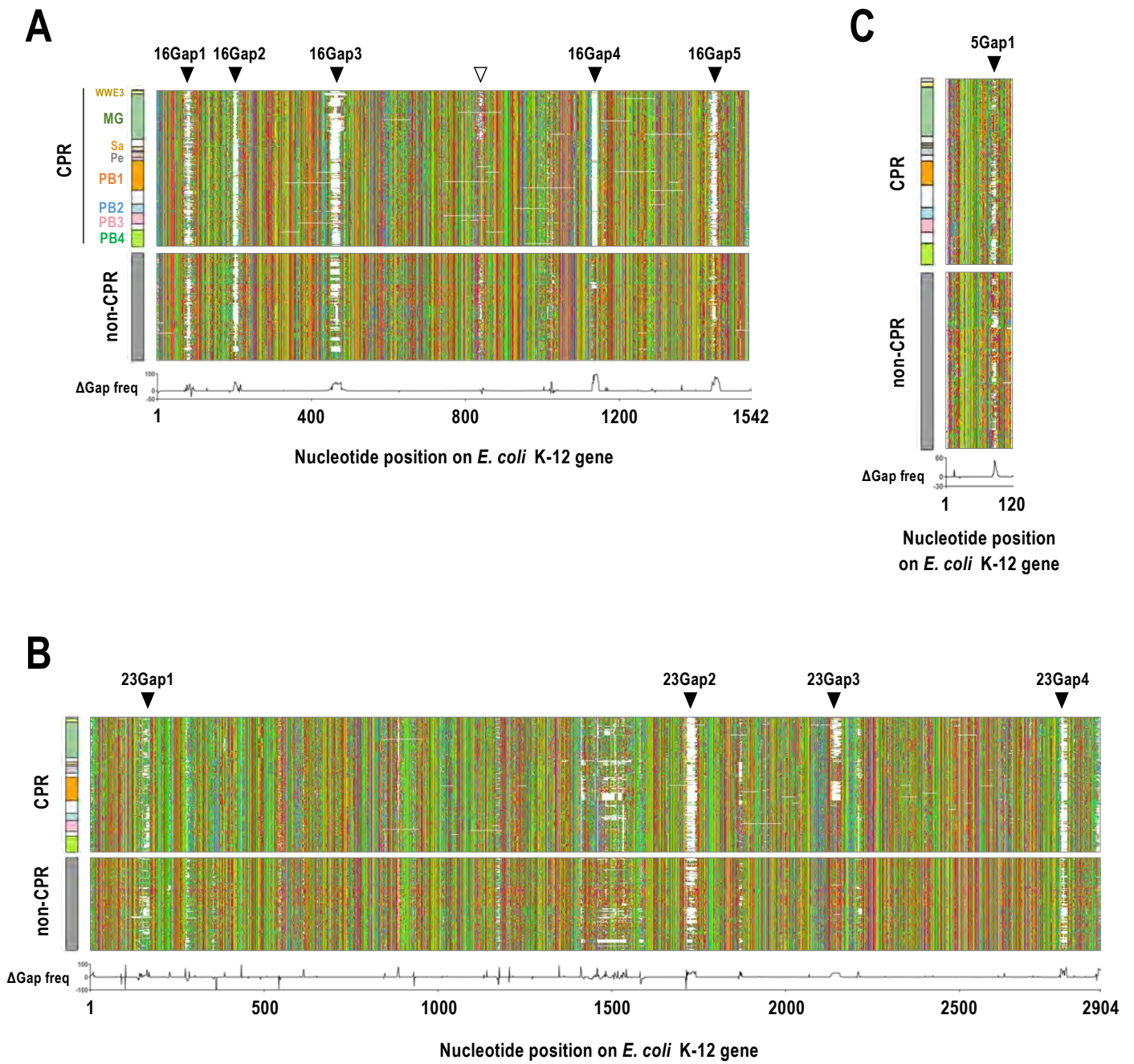




bL36



Supplementary Figure S4. Amino acid sequence alignments of another 50 ribosomal proteins (completion of Figure 3). Density curves in the top panel show the distributions of ORF lengths (CPR, light red; non-CPR, light blue). See Figure 3 legend for details. Aligned sequences are ordered based on their phylogenetic relationships (Jaffe et al. 2020).



Supplementary Figure S5

The legend for this figure is placed on the next page.

Supplementary Figure S5. Nucleotide sequence alignment of the three types of rRNA genes in CPR and non-CPR bacteria. Multiple alignments of (A) 16S, (B) 23S, and (C) 5S rRNA gene sequences from CPR and non-CPR bacterial genomes. Gene sequences of each type of rRNA (16S, 23S, and 5S) are clustered by sequence identity, and representative sequences for 16S (CPR: n = 278; non-CPR, n = 191), 23S (CPR, n = 256; non-CPR, n = 176), and 5S rRNA genes (CPR, n = 250; non-CPR, n = 237) were aligned with MAFFT (L-INS-i algorithm). Insertion sequences with respect to the *Escherichia coli* K-12 genes were removed. Aligned sequences are shown in the order based largely on the phylogenetic relationships (Jaffe et al. 2020). Δ Gap frequency (Δ Gap freq.) is defined as follows: (gap frequency in CPR bacteria) – (gap frequency in non-CPR bacteria). Δ Gap freq. at each site was calculated based on all the sequences before clustering and is shown at the bottom. Black inverted triangles at the top of each alignment indicate regions with higher Δ Gap freq. values. A white inverted triangle indicates a region highly conserved in CPR bacteria. Bases are colored as follows: adenine (A), light green; thymine (T), blue; guanine (G), red; cytosine (C), orange. Panels on the left side of the alignment are colored according to the taxonomic group of each sequence (see Supplementary Figure S1). MG: Microgenomates; Sa: Saccharibacteria; Pe: Peregrinibacteria; PB: Parcubacteria.

Supplementary Figure S6. Mapping the lost rRNA regions and lost ribosomal protein-binding regions in CPR bacteria to 2D *Escherichia coli* ribosomal structures (supplementary analysis of Figure 6). The regions were mapped with the RiboVision2 server (<http://apollo.chemistry.gatech.edu/RiboVision2/>) (Bernier *et al.* 2014), based on *E. coli* ribosome structure data (PDB ID: 4V9D). Ribonucleotides contacting the ribosomal proteins bS21, bL9, bL25, bL28, uL29, uL30, bL32, and bL33, which are lost in all or some CPR bacteria, are shown with green circles on the secondary structures of *E. coli* rRNAs. The approximate binding regions on the rRNAs for each ribosomal protein are surrounded by a broken line. The rRNA regions lacking in CPR bacteria (see Figure 5) are highlighted with red circles.

References

Jaffe AL, Castelle CJ, Matheus Carnevali PB, Gribaldo S, Banfield JF. 2020. The rise of diversity in metabolic platforms across the Candidate Phyla Radiation. *BMC Biol* **18**:1–15.

Bernier CR., Petrov AS, Waterbury CC, Jett J, Li F, Freil LE, Xiong X, Wang L, Migliozzi BLR, Hershkovits E, Xue Y, Hsiao C, Bowman JC, Harvey SC, Grover MA, Wartell ZJ, Williams LD. 2014. RiboVision Suite for Visualization and Analysis of Ribosomes. *Faraday Discussions* **169**(1):195-207.