

Supplementary information

Table S1: Frequency of the four bipyrimidine dinucleotides in a series of 99 bacteria with increasing GC content (% GC). These values were used to draw Figure 1.

Bacterial species	Accession number	%GC	TT	TC	CT	CC
<i>Buchnera aphidicola</i> str. Bp (<i>Baizongia pistaciae</i>)	NC 004545	25.3	0.1540	0.0449	0.0430	0.0196
<i>Buchnera aphidicola</i> str. APS (<i>Acyrtosiphon pisum</i>)	NC 002528	26.4	0.1522	0.0492	0.0452	0.0209
Aster yellows witches'-broom phytoplasma AYWB	NC 007716	26.8	0.1743	0.0449	0.0451	0.0250
<i>Arcobacter butzleri</i> RM4018	NC 009850	27.0	0.1546	0.0512	0.0543	0.0196
<i>Candidatus Blochmannia floridanus</i>	NC 005061	27.4	0.1301	0.0474	0.0416	0.0220
<i>Borrelia afzelii</i> PKo	NC 008277	27.8	0.1569	0.0524	0.0551	0.0262
<i>Clostridium perfringens</i> ATCC 13124	NC 008261	28.0	0.1413	0.0541	0.0649	0.0272
<i>Clostridium perfringens</i> SM101	NC 008262	28.0	0.1433	0.0550	0.0654	0.0274
<i>Clostridium perfringens</i> str13	NC 003366	28.0	0.1416	0.0553	0.0661	0.0279
<i>Candidatus Pelagibacter ubique</i> HTCC1062	NC 007205	29.0	0.1448	0.0511	0.0543	0.0266
<i>Campylobacter jejuni</i> RM1221	NC 003912	30.3	0.1513	0.0487	0.0579	0.0256
<i>Prochlorococcus marinus</i> MIT9312	NC 007577	31.2	0.1392	0.0583	0.0585	0.0308
<i>Methanococcus maripaludis</i> S2	NC 005791	33.0	0.1372	0.0556	0.0481	0.0332
<i>Baumannia cicadellincola</i> str. Hc (<i>Homalodisca coagulata</i>)	NC 007984	33.2	0.1096	0.0466	0.0583	0.0280
<i>Candidatus Protochlamydia amoebophila</i> UWE25	NC 005861	34.7	0.1337	0.0599	0.0577	0.0331
<i>Bacillus anthracis</i> str. 'Ames Ancestor'	NC 007530	35.2	0.1193	0.0572	0.0536	0.0321
<i>Bacillus thuringiensis</i> serovar konkukian str. 97-27	NC 005957	35.4	0.1191	0.0571	0.0536	0.0321
<i>Bacillus thuringiensis</i> str. Al Hakam	NC 008600	35.4	0.1185	0.0567	0.0533	0.0320
<i>Alkaliphilus oremlandii</i> OhLAs	NC 009922	36.2	0.1097	0.0596	0.0586	0.0392
<i>Alkaliphilus metalliredigens</i> QYMF	NC 009633	36.8	0.1091	0.0595	0.0594	0.0422
<i>Bartonella henselae</i> str. Houston-1	NC 005956	38.2	0.1150	0.0573	0.0532	0.0381
<i>Bartonella bacilliformis</i> KC583	NC 008783	38.2	0.1157	0.0561	0.0536	0.0343
<i>Bartonella quintana</i> str. Toulouse	NC 005955	38.8	0.1143	0.0579	0.0537	0.0390
<i>Acinetobacter baumannii</i> ATCC 17978	NC 009085	38.9	0.1105	0.0518	0.0542	0.0372
<i>Vibrio angustum</i> S14		39.6	0.1544	0.0779	0.0801	0.0520
<i>Pseudoalteromonas haloplanktis</i> TAC125 chromosome I	NC 007481	40.1	0.1063	0.0392	0.0570	0.0380
<i>Chlamydia muridarum</i> Nigg	NC 002620	40.3	0.1056	0.0675	0.0689	0.0433
<i>Acinetobacter</i> sp. ADP1	NC 005966	40.4	0.1020	0.0559	0.0516	0.0390
<i>Chlamydia pneumoniae</i> AR39	NC 002179	40.6	0.1019	0.0697	0.0716	0.0433
<i>Bacillus pumilus</i> SAFR-032	NC 009848	41.3	0.1017	0.0669	0.0580	0.0396
<i>Actinobacillus pleuropneumoniae</i> L20	NC 009053	41.3	0.1115	0.0546	0.0443	0.0424
<i>Chlamydia trachomatis</i> DIUW-3/CX	NC 000117	41.3	0.0993	0.0695	0.0713	0.0429
<i>Anabaena variabilis</i> ATCC 29413	NC 007413	41.4	0.0982	0.0553	0.0600	0.0474
<i>Anaplasma phagocytophilum</i> HZ	NC 007797	41.6	0.0870	0.0561	0.0668	0.0403
<i>Pyrococcus abyssi</i> GE5	NC 000868	42.0	0.0854	0.0714	0.0755	0.0612
<i>Carboxydotherrmus hydrogenoformans</i> Z-2901	NC 007503	42.0	0.1173	0.0506	0.0543	0.0644
<i>Psychrobacter arcticus</i> 273-4	NC 007204	42.7	0.0881	0.0524	0.0573	0.0397
<i>Bacteroides thetaiotaomicron</i> VPI-5482	NC 004663	42.9	0.0908	0.0657	0.0562	0.0477
<i>Bacteroides fragilis</i> NCTC 9343	NC 003228	43.1	0.0907	0.0646	0.0548	0.0505
<i>Bacteroides fragilis</i> YCH46	NC 006347	43.2	0.0907	0.0649	0.0549	0.0507
<i>Aquifex aeolicus</i> VF5	NC 000918	43.3	0.1019	0.0683	0.0722	0.0581
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	NC 000964	43.5	0.0989	0.0658	0.0564	0.0460
<i>Bacillus halodurans</i> C-125	NC 002570	43.7	0.0961	0.0670	0.0553	0.0466
<i>Campylobacter curvus</i> 525.92	NC 009715	44.5	0.1008	0.0604	0.0599	0.0410
<i>Bacillus clausii</i> KSM-K16	NC 006582	44.8	0.0989	0.0586	0.0543	0.0486
<i>Actinobacillus succinogenes</i> 130Z	NC 009655	44.9	0.1019	0.0550	0.0393	0.0522
<i>Vibrio vulnificus</i> CMCP6 chromosome I	NC 004459	45.0	0.0878	0.0574	0.0578	0.0541

<i>Bacillus licheniformis</i> ATCC 14580	NC 006270	46.2	0.0956	0.0695	0.0537	0.0508
<i>Chlorobium phaeobacteroides</i> DSM 266	NC 008639	48.4	0.0831	0.0717	0.0576	0.0587
<i>Archaeoglobus fulgidus</i> DSM 4304	NC 000917	48.6	0.0794	0.0735	0.0723	0.0607
<i>Anaplasma marginale</i> str. St. Maries	NC 004842	49.8	0.0684	0.0549	0.0629	0.0618
<i>Escherichia coli</i> K12	NC 000913	50.0	0.0732	0.0576	0.0509	0.0586
<i>Nitrosococcus oceani</i> ATCC 19707	NC 007484	50.3	0.0664	0.0624	0.0640	0.0741
<i>Bdellovibrio bacteriovorus</i> HD100	NC 005363	50.6	0.0797	0.0675	0.0562	0.0641
<i>Salmonella typhimurium</i> LT2	NC 003197	52.0	0.0698	0.0572	0.0500	0.0617
<i>Alcanivorax borkumensis</i> SK2	NC 008260	54.7	0.0612	0.0547	0.0555	0.0744
<i>Morella thermoacetica</i> ATCC 39073	NC 007644	55.0	0.0618	0.0516	0.0572	0.0968
<i>Pelobacter carbinolicus</i> DMS 2380	NC 007498	55.1	0.0643	0.0637	0.0504	0.0775
<i>Candidatus Desulfococcus oleovorans</i> Hxd3	NC 009943	56.1	0.0650	0.0575	0.0480	0.0906
<i>Chlorobium tepidum</i> TLS	NC 002932	56.5	0.0581	0.0723	0.0558	0.0679
<i>Brucella abortus</i> biovar 1 str. 9-941 chromosome I	NC 006932	57.2	0.0592	0.0640	0.0494	0.0720
<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	NC 009348	58.2	0.0449	0.0612	0.0613	0.0831
<i>Acidobacteria bacterium</i> Ellin345	NC 008009	58.4	0.0520	0.0713	0.0535	0.0681
<i>Agrobacterium tumefaciens</i> str. C58 chromosome circular	NC 003062	59.0	0.0509	0.0694	0.0498	0.0776
<i>Bifidobacterium adolescentis</i> ATCC 15703	NC 008618	59.2	0.0462	0.0694	0.0457	0.0724
<i>Synechococcus</i> sp WH8102	NC 005070	59.4	0.0453	0.0658	0.0607	0.0837
<i>Bifidobacterium longum</i> NCC2705	NC 004307	60.1	0.0419	0.0667	0.0487	0.0778
<i>Haloarcula marismortui</i> ATCC 43049 chromosome I	NC 006396	61.1	0.0515	0.0761	0.0610	0.0601
<i>Aeromonas hydrophila</i> subsp. <i>hydrophila</i> ATCC 7966	NC 008570	61.5	0.0381	0.0604	0.0603	0.0900
<i>Arthrobacter aurescens</i> TC1	NC 008711	62.4	0.0402	0.0638	0.0540	0.0920
<i>Erythrobacter litoralis</i> HTCC2594	NC 007722	63.0	0.0394	0.0700	0.0511	0.0789
<i>Natrosomonas pharaonis</i> DSM 2160	NC 007426	63.1	0.0346	0.0795	0.0501	0.0788
<i>Chromohalobacter salexigens</i> DSM 3043	NC 007963	63.9	0.0325	0.0658	0.0512	0.0829
<i>Azoarcus</i> sp. EbN1	NC 006513	64.7	0.0349	0.0691	0.0495	0.0773
<i>Chromobacterium violaceum</i> ATCC 12472	NC 005085	64.8	0.0375	0.0562	0.0531	0.0875
<i>Arthrobacter</i> sp. FB24 chromosome I	NC 008541	65.4	0.0377	0.0703	0.0585	0.1141
<i>Sphingopyxis alaskensis</i> RB2256	NC 008048	65.5	0.0351	0.0694	0.0446	0.0844
<i>Bradyrhizobium</i> sp. ORS278	NC 009445	65.5	0.0300	0.0718	0.0511	0.0797
<i>Acidovorax</i> sp. JS42	NC 008782	66.1	0.0315	0.0514	0.0528	0.0903
<i>Acidothermus cellolyticus</i> 11B	NC 008578	66.9	0.0296	0.0693	0.0441	0.0933
<i>Azorhizobium caulinodans</i> ORS 571	NC 009937	67.0	0.0278	0.0648	0.0520	0.0981
<i>Deinococcus radiodurans</i> R1 chromosome I	NC 001263	67.0	0.0402	0.0577	0.0604	0.0892
<i>Caulobacter crescentus</i> CB15	NC 002696	67.0	0.0292	0.0671	0.0526	0.0967
<i>Acidiphilum cryptum</i> JF-5	NC 009484	67.1	0.0284	0.0677	0.0463	0.0953
<i>Caulobacter crescentus</i> CB15	NC 002696	67.2	0.0292	0.0671	0.0526	0.0967
<i>Azoarcus</i> sp. BH72	NC 008702	67.9	0.0287	0.0592	0.0504	0.0906
<i>Bordetella parapertussis</i> 12822	NC 002928	68.1	0.0280	0.0524	0.0472	0.0922
<i>Burkholderia pseudomallei</i> K96243 chromosome I	NC 006350	68.1	0.0285	0.0683	0.0412	0.0714
<i>Acidovorax avenae</i> subsp. <i>citrulli</i> AAC00-1	NC 008752	68.5	0.0263	0.0551	0.0508	0.0980
<i>Nocardia farcinica</i> IFM 10152	NC 006361	70.0	0.0186	0.0632	0.0437	0.1028
<i>Frankia</i> sp Ccl3	NC 007777	70.1	0.0189	0.0672	0.0453	0.1100
<i>Rubrobacter xylanophilus</i> DSM 9941	NC 008148	70.5	0.0197	0.0663	0.0613	0.1223
<i>Saccharopolyspora erythraea</i> NRRL2338	NC 008142	71.1	0.0182	0.0641	0.0483	0.1038
<i>Nocardioides</i> sp. JS 614	NC 008699	71.4	0.0142	0.0691	0.0498	0.1059
<i>Streptomyces coelicolor</i> A3(2)	NC 003888	72.0	0.0159	0.0630	0.0482	0.1139
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPPB 382	NC 009480	72.0	0.0123	0.0725	0.0468	0.0998
<i>Frankia alni</i> ACN14a	NC 008278	72.5	0.0123	0.0725	0.0468	0.0998
<i>Anaeromyxobacter</i> sp. Fw109-5	NC 009675	72.8	0.0146	0.0640	0.0433	0.1142
<i>Anaeromyxobacter dehalogenans</i> 2CP-C	NC 007760	74.9	0.0136	0.0615	0.0489	0.1088

Table S2: Frequency of pyrimidine dinucleotides in the genomes of the studied species. All values are percentages (%). The proportion of C-containing bipyrimidine dinucleotides was also calculated.

	GC content	TT	TC	CT	CC	<i>C-containing dinucleotides</i>
<i>Clostridium perfringens</i>	28	16.5	6.0	6.6	2.3	<i>47.4</i>
<i>Vibrio angustum</i>	39.6	15.4	7.8	8.0	5.2	<i>57.6</i>
Calf Thymus	42	9.9	6.1	5.3	4.3	<i>61.2</i>
<i>Salmonella typhimurium</i>	52	7.0	5.7	5.0	6.2	<i>70.8</i>
<i>Sphingopyxis alaskensis</i>	65.5	3.5	6.9	4.5	8.4	<i>85.0</i>
<i>Micrococcus luteus</i>	72	1.5	6.6	4.5	10.6	<i>93.5</i>

Figure S1: Determination of the dinucleotide frequency for *M. luteus* (72% GC) in the high GC content region. Values were interpolated from those of eight species exhibiting a similar GC content: *Nocardia farcinica* IFM 10152 (70% GC), *Frankia* sp. Cc13 (70.1), *Saccharopolyspora erythraea* NRRL 2338 (71.1), *Nocardioides* sp. JS614 (71.4), *Streptomyces coelicolor* A3(2) (72), *Clavibacter michiganensis* subsp. *michiganensis* NCPPB 382 (72), *Frankia alni* ACN14a (72.5), *Anaeromyxobacter dehalogenans* 2CP-C (74.9).

