



Original article

Information theoretic perspective on genome clustering

Alaguraj Veluchamy^{a,b,*}, Preeti Mehta^a, K.V. Srividhya^a, Hirendra Vikram^a, M.K. Govind^a,
Ramneek Gupta^a, Abdul Aziz Bin Dukhyil^c, Raed Abdullah Alharbi^d, Saleh Abdullah Aloyuni^d,
Mohamed M. Hassan^{e,f}, S. Krishnaswamy^a

^a Centre of Excellence in Bioinformatics, School of Biotechnology, Madurai Kamaraj University, Madurai 625021, India

^b Department of Computational Biology, St. Jude Children's Research Hospital, Danny Thomas Place, Memphis 38105, Tennessee, United States of America

^c Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Majmaah University, Al Majmaah 11952, Saudi Arabia

^d Department of Public Health, College of Applied Medical Sciences, Majmaah University, Al Majmaah 11952, Saudi Arabia

^e Department of Biology, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

^f Department of Genetics, Faculty of Agriculture, Menoufia University, Sheben El-Kom, Menoufia Government, Egypt



ARTICLE INFO

Article history:

Received 16 November 2020

Revised 15 December 2020

Accepted 23 December 2020

Available online 31 December 2020

Keywords:

Information theory
Genome evolution
Genome clustering
Nucleotide distribution
Shannon redundancy
Genome arrangement

ABSTRACT

Shannon's information theoretic perspective of communication helps one to understand the storage and processing of information in one-dimensional sequences. An information theoretic analysis of 937 available completely sequenced prokaryotic genomes and 238 eukaryotic chromosomes is presented. Information content (Id) values were used to cluster these chromosomes. Chargaff's second parity rule i.e. compositional self-complementarity, an empirical fact is observed in all the genomes, except for the proteobacteria *Candidatus* Hodgkinia cicadicola. High information content, arising out of biased base composition in all the 14 chromosomes of *Plasmodium falciparum* is found among two other genomes of prokaryotes viz. *Buchnera aphidicola* str. Cc (*Cinara cedri*) and *Candidatus* Carsonella ruddii PV. Despite size and compositional variations, both prokaryotic and eukaryotic genomes do not deviate significantly from an equiprobable and random situation. Eukaryotic chromosomes of an organism tend to have similar informational restraints as seen when a simple distance based method is used to cluster them. In eukaryotes, in certain cases, Id values are also similar for the two arms (p and q arm) of the chromosomes. The results of this current study confirm that the information content can provide insights into the clustering of genomes and the evolution of messaging strategies of the genomes. An efficient and robust Perl CGI standalone tool is created based on this information theory algorithm for the analysis of the whole genomes and is made available at <https://github.com/AlagurajVeluchamy/InformationTheory>.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

A genome can be viewed as a message string from an alphabet of four symbols. The analysis of symbol arrangements and their groupings has been used in linguistics (Zipf, 1935; Zipf, 1942;

* Corresponding author at: Department of Computational Biology, St. Jude Children's Research Hospital, Danny Thomas Place, Memphis 38105, Tennessee, United States of America

E-mail address: avelucha@stjude.org (A. Veluchamy).

Peer review under responsibility of King Saud University.



Dewey, 1923). Any text using a language has a symbol composition bias at the letter or word level. The ordering of symbols or syntax also differs between authors and languages. Such a syntax-based approach was used in biological systematics (Brooks and Wiley, 1986; Chomsky, 1965). Information theory provides a utilizable implement for for investigating any one-dimensional arrangement of symbols (Shannon and Weaver, 1949; Shannon, 1951). Shannon, who developed this concept of information theory, relied on it to quantify information transfer and error management. He had also applied the theory to an analysis of symbol arrangements within the English text. Formal language theory has been applied on a continuous basis to different biological aspects. Rather than routine regular expression, which is difficult for genes, grammatical analysis and gene patterns have been applied to genomes. (Brendel and Busse, 1984). Information theory has been used in varied biological contexts, including bioinformatics, neurobiology,

<https://doi.org/10.1016/j.sjbs.2020.12.039>

1319-562X/© 2020 The Author(s). Published by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ecology and evolutionary biology (Staden, 1984; Schneider, 1997; Wallace and Wallace, 1998; Forsdyke, 1999). Information theory applied to DNA sequences dates back to the beginning of the 1970s and since then has been discussed in a number of studies on different aspects of DNA sequence and structure. (Gatlin, 1972; Yockey, 1974; Tanaka, 1980; Doig, 1997; Reddy et al., 2006a,b). Shannon information has been used to discover patterns such as conserved DNA motif (Pesole et al., 1994), to implement sequence alignment (Clote and Backofen, 2000) and to complete genomes of selected prokaryotes and eukaryotes (Chang et al., 2005). Earlier work on mouse and human genomes for the synonymous codon usage biases in coding regions from a Shannon information theory perspective suggest that evolution may target and position a given gene in a region of a particular GC content so as to adjust the information to the appropriate value for that gene in that organism (Zeeberg, 2002).

For applying Shannon-Weaver formalism to biological sequences, Gatlin proposed modifications to the concept of this SW formalism (Gatlin, 1972). In our study, we took advantage of this formalism, to calculate information density and redundancy measures for completely sequenced genomes. Similar work was done by Gatlin herself, when she used the nearest neighbour data for 15 organisms and tissues to calculate information density, redundancy and D-indices, treating DNA as a Markov process of order one. The use of the Shannon information metric and Gatlin's formalism in the biological context for functional characterization was disputed (Hariri et al., 1990; Sibbald et al., 1989). However, functional interpretations for instance on applications of Information content on Transcription Factor Binding Sites has been made (Reddy et al., 2006a,b). Our current work presented here looks at the clustering of genomes and the evolution of messaging strategies in genomes in the context of information theory.

2. Methods and databases

Completed genomes of 937 prokaryotes and 238 eukaryotic chromosomes were obtained <https://ftp.ncbi.nlm.nih.gov/genbank> p and q arm data for these eukaryotic chromosomes were obtained from <https://www.ncbi.nlm.nih.gov/genome/gdv>. Preliminary compositional studies were done using GCG (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, USA), EMBOSS (European Molecular Biology Software Suite) and Perl scripts. Additional computations were carried out using UNIX shell scripts. Only the standard bases (A, T, G and C) were taken into consideration when defining the sampling space. Unassigned nucleotides (N) were omitted. In prokaryotic genomes, the number of such inputs is negligible compared to the length of the genome. In eukaryotes individual chromosomes were considered as separate entities. In this study, the genomes were analyzed only up to a second order Markov process. Perl CGI was used for implementation of the algorithm and Information theory based analysis is made available at <https://github.com/AlagurajVeluchamy/InformationTheory>

2.1. Algorithm

Information density (Id) for a second order Markov process was calculated as given below:

$$Id = D_1 + D_2 + D_3$$

$$D_1 = \log_2 a - H_1,$$

$$D_2 = H_1 - H_M^{(1)} \text{ and}$$

$$D_3 = H_M^{(1)} - H_M^{(2)}$$

$$H_1 = - \sum p_i \log_2 p_i,$$

$$H_M^{(1)} = - \sum p_i p_{ij} \log_2 p_{ij},$$

$$H_M^{(2)} = - \sum p_i p_{ij} p_{ijk} \log_2 p_{ijk}$$

where, 'a' is the number of letters in the alphabet (4 in the case of DNA), 'D₁' is a measure of divergence from equiprobability (as $\log_2 a$ is the maximum uncertainty that can be obtained from a sequence with equiprobable occurrence of each element), 'D₂', 'D₃' provide a measure of divergence from independent occurrence of symbols for the dinucleotide and trinucleotide probabilities of occurrence and 'H_M⁽¹⁾', 'H_M⁽²⁾' are the entropy of the Markov process arising from dinucleotide and trinucleotide distributions. 'p_{ij}' is the conditional probability of occurrence of nucleotide 'j', given 'i' has occurred and 'p_{ijk}' is the conditional probability of occurrence of 'k', given the doublet 'ij' has occurred.

Redundancy was calculated as in Gatlin's formalism (Gatlin, 1972) and is given by $Id/\log_2 a$. The D-indices, RD1, RD2 and RD3 were calculated as defined by Gatlin (Gatlin, 1972). RD1 (=D₁/Id) is a measure of the contribution of the composition to the total redundancy. RD2 (=D₂/Id) and RD3 (=D₃/Id) are measures of the Shannon redundancy arising from Markov processes of order one and two and are defined, similarly, in terms of D₂ and D₃.

2.2. Clustering of genomes using Id values

A matrix of dimension n × n is created with Id values, where each element entry represents the absolute value of their difference. A graphical tree akin to that of the phylogenetic tree is created with the UPGMA method (Sneath and Snokal, 1973) of clustering, using the PHYLIP package (Felsenstein, 1989). Above method was applied to a dataset of 937 prokaryotes and the 238 eukaryotic chromosomes. Results of this analyses are compared to the standard taxonomy tree of different groups (<http://ncbi.nlm.nih.gov/Taxonomy>). The purpose of this dendrogram construction is to find the relationship of various genomes in terms of information content.

3. Results

3.1. Chargaff's rule

Chargaff's second parity rule of A = T, G = C even within the single strand of a double stranded genome is found to be valid for all the genome sequences studied except for alpha proteobacteria *Candidatus* Hodgkinia cicadicola which shows A/T ratio of 0.74 and a G/C ratio of 1.11 (Tab.1). Similarly the G/C ratio is found to be less (0.88) for *Candidatus* Liberibacter asiaticus str. psy62.

3.2. Prokaryotes and eukaryotes show similar nucleotide distribution

The information density analysis was carried out for complete sequences of 937 prokaryotic genomes (Table 1) and 238 eukaryotic chromosomes (Table 2). For prokaryotes, information density varied from 0.016 bits to 0.383 bits (0.083 ± 0.052). *Candidatus* Carsonella ruddii PV has the highest and *Anaplasma marginale* str. Florida has the least information density (Id) value. For eukaryotes Id values varied between 0.011 and 0.354 bits (0.098 ± 0.078) with *Plasmodium falciparum* showing highest and *Eremothecium gossypii* ATCC 10,895 chromosome-V having the lowest value. Genomes of *Homo sapiens*, *Rattus norvegicus* (except chromosome 12 and 20), *Arabidopsis thaliana*, *Mus musculus* and *Pan troglodytes* tend to show similar values of information density. For prokaryotes and eukaryotes, the Id values varied over a restricted range of values.

3.3. Nucleotide distribution until markov third order is equiprobable

Theoretically the Id values can vary between zero bits (for a random and equiprobable case) to a maximum of 2.00 bits (for a monotonic case) (Fig. 1a and Fig. 1b), considering that the range

Table 1
Information density values for prokaryotic genomes.

Organism	Length	Id (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
<i>Escherichia coli</i> str. K-12 substr. W3110	4,646,332	0.037	0.9985563	1.0018308	0.005	0.5	0.495	0	0.018	0.018
<i>Chlamydia trachomatis</i> D/UW-3/CX	1,042,519	0.05	1.0051087	1.0007992	0.438	0.392	0.17	0.022	0.02	0.009
<i>Thermotoga maritima</i> MSB8	1,860,725	0.07	1.0069953	1.031547	0.059	0.587	0.354	0.004	0.041	0.025
<i>Aeropyrum pernix</i> K1	1,669,696	0.045	0.9744891	0.9862076	0.257	0.489	0.254	0.012	0.022	0.011
<i>Pyrococcus abyssii</i> GE5	1,765,118	0.044	0.9956219	0.9931989	0.184	0.523	0.292	0.008	0.023	0.013
<i>Haemophilus influenzae</i> Rd KW20	1,830,138	0.072	1.0059937	0.9906276	0.57	0.345	0.085	0.041	0.025	0.006
<i>Mycoplasma genitalium</i> G37	580,076	0.148	1.0246947	1.0086433	0.672	0.219	0.11	0.099	0.032	0.016
<i>Methanocaldococcus jannaschii</i> DSM 2661	1,664,970	0.143	1.0090428	1.0228403	0.713	0.211	0.075	0.102	0.03	0.011
<i>Synechocystis</i> sp. PCC 6803	3,573,470	0.042	0.9962869	1.0027565	0.036	0.809	0.155	0.002	0.034	0.006
<i>Mycoplasma pneumoniae</i> M129	816,394	0.067	1.0359619	1.0048061	0.436	0.406	0.158	0.029	0.027	0.011
<i>Escherichia coli</i> str. K-12 substr. MG1655	4,639,675	0.037	1.0011028	0.9977698	0.005	0.501	0.494	0	0.018	0.018
<i>Helicobacter pylori</i> 26,695	1,667,867	0.103	0.9831194	0.9822459	0.351	0.433	0.216	0.036	0.044	0.022
<i>Methanothermobacter thermautotrophicus</i> str. Delta H	1,751,377	0.049	0.9887473	1.0033266	0.001	0.639	0.359	0	0.032	0.018
<i>Archaeoglobus fulgidus</i> DSM 4304	2,178,400	0.038	1.0073386	1.0070227	0.016	0.694	0.29	0.001	0.026	0.011
<i>Aquifex aeolicus</i> VF5	1,551,335	0.069	1.0107407	1.0051697	0.18	0.527	0.293	0.012	0.036	0.02
<i>Treponema pallidum</i> subsp. pallidum str. Nichols	1,138,011	0.034	0.9940553	1.0138022	0.066	0.363	0.571	0.002	0.012	0.019
<i>Helicobacter pylori</i> J99	1,643,831	0.1	0.9947974	0.9896772	0.339	0.438	0.223	0.034	0.044	0.022
<i>Chlamydia pneumoniae</i> CWL029	1,230,230	0.051	1.0094918	0.9971552	0.502	0.38	0.118	0.026	0.02	0.006
<i>Pyrococcus horikoshii</i> OT3	1,738,505	0.058	0.9954245	0.9750859	0.333	0.483	0.185	0.019	0.028	0.011
<i>Mycobacterium tuberculosis</i> H37Rv	4,411,532	0.108	1.0002455	0.9962882	0.663	0.159	0.178	0.072	0.017	0.019
<i>Rickettsia prowazekii</i> str. Madrid E	1,111,523	0.143	0.9929545	1.0167612	0.918	0.052	0.03	0.131	0.007	0.004
<i>Bacillus subtilis</i> subsp. subtilis str. 168	4,214,630	0.051	0.9957352	0.9954971	0.239	0.477	0.284	0.012	0.024	0.014
<i>Deinococcus radiodurans</i> R1	2,648,638	0.112	0.9985151	0.9978972	0.758	0.139	0.103	0.085	0.016	0.012
<i>Borrelia burgdorferi</i> B31	910,724	0.173	0.9874173	0.9914806	0.789	0.193	0.018	0.137	0.034	0.003
<i>Ureaplasma parvum</i> serovar 3 str. ATCC 700,970	751,719	0.206	1.0025137	1.0307232	0.876	0.087	0.037	0.181	0.018	0.008
<i>Campylobacter jejuni</i> subsp. jejuni NCTC 11,168	1,641,481	0.16	1.005905	0.9908884	0.701	0.234	0.065	0.112	0.037	0.01
<i>Chlamydia pneumoniae</i> AR39	1,229,853	0.051	0.9906761	1.0027495	0.502	0.38	0.119	0.026	0.02	0.006
<i>Xylella fastidiosa</i> 9a5c	2,679,306	0.027	0.9097728	1.1117256	0.146	0.64	0.214	0.004	0.017	0.006
<i>Chlamydia pneumoniae</i> J138	1,226,565	0.051	1.0102885	0.9965905	0.501	0.38	0.118	0.026	0.02	0.006
<i>Vibrio cholerae</i> O1 biovar El Tor str. N16961	2,961,149	0.031	0.9867453	1.0078863	0.05	0.578	0.372	0.002	0.018	0.012
<i>Pseudomonas aeruginosa</i> PAO1	6,264,404	0.128	1.0165397	0.9828533	0.632	0.127	0.241	0.081	0.016	0.031
<i>Buchnera aphidicola</i> str. APS (<i>Acyrtosiphon pisum</i>)	640,681	0.187	1.0125413	1.0111539	0.903	0.059	0.038	0.169	0.011	0.007
<i>Bacillus halodurans</i> C-125	4,202,352	0.033	1.0046633	1.0169247	0.349	0.481	0.17	0.012	0.016	0.006
<i>Thermoplasma acidophilum</i> DSM 1728	1,564,906	0.038	1.0136883	1.007707	0.123	0.388	0.489	0.005	0.015	0.019
<i>Halobacterium</i> sp. NRC-1	2,014,239	0.146	1.0011207	0.9969854	0.649	0.281	0.07	0.095	0.041	0.01
<i>Chlamydia muridarum</i> Nigg	1,072,950	0.056	1.0023399	1.0064671	0.48	0.384	0.136	0.027	0.022	0.008
<i>Escherichia coli</i> O157:H7 EDL933	5,528,445	0.036	1.0025756	0.9986057	0.002	0.478	0.52	0	0.017	0.018
<i>Lactococcus lactis</i> subsp. lactis I11403	2,365,589	0.092	1.0026012	1.012585	0.686	0.262	0.053	0.063	0.024	0.005
<i>Pasteurella multocida</i> subsp. multocida str. Pm70	2,257,487	0.051	1.0037264	1.0267011	0.523	0.391	0.086	0.027	0.02	0.004
<i>Mycobacterium leprae</i> TN	3,268,203	0.04	0.9924797	1.0122391	0.44	0.234	0.326	0.018	0.009	0.013
<i>Mesorhizobium loti</i> MAFF303099	7,036,071	0.104	1.0010453	0.9843826	0.455	0.335	0.209	0.047	0.035	0.022
<i>Thermoplasma volcanium</i> GSS1	1,584,804	0.047	1.0081152	1.0052746	0.629	0.167	0.204	0.03	0.008	0.01
<i>Escherichia coli</i> O157:H7 str. Sakai	5,498,450	0.036	1.0024258	1.0045313	0.002	0.479	0.519	0	0.017	0.018
<i>Caulobacter crescentus</i> CB15	4,016,947	0.139	1.0079974	0.9962353	0.629	0.153	0.219	0.087	0.021	0.03
<i>Streptococcus pyogenes</i> M1 GAS	1,852,441	0.06	1.0099953	1.0190775	0.64	0.248	0.111	0.038	0.015	0.007
<i>Staphylococcus aureus</i> subsp. aureus N315	2,814,816	0.098	0.9900767	1.0011949	0.882	0.087	0.031	0.087	0.009	0.003
<i>Sulfolobus solfataricus</i> P2	2,992,245	0.075	0.9893945	1.0048268	0.784	0.153	0.063	0.059	0.012	0.005
<i>Mycobacterium tuberculosis</i> CDC1551	4,403,837	0.108	0.9996511	0.9954969	0.662	0.16	0.178	0.071	0.017	0.019
<i>Staphylococcus aureus</i> subsp. aureus Mu50	2,878,529	0.098	0.994033	1.0060729	0.883	0.086	0.031	0.086	0.008	0.003
<i>Mycoplasma pulmonis</i> UAB CTIP	963,879	0.215	1.016606	0.9980227	0.764	0.214	0.023	0.164	0.046	0.005
<i>Bordetella bronchiseptica</i> RB50	5,339,179	0.15	0.9962368	1.0087376	0.642	0.183	0.175	0.096	0.027	0.026
<i>Bordetella parapertussis</i> 12,822	4,773,551	0.151	0.9950462	1.0156752	0.64	0.184	0.176	0.097	0.028	0.027
<i>Bordetella pertussis</i> Tohama 1	4,086,189	0.145	0.9995515	1.0078314	0.639	0.181	0.18	0.093	0.026	0.026
<i>Chlorobium tepidum</i> TLS	2,154,946	0.057	1.0135893	0.9875008	0.216	0.505	0.279	0.012	0.029	0.016
<i>Corynebacterium diphtheriae</i> NCTC 13,129	2,488,635	0.02	0.9982119	1.0006266	0.176	0.541	0.284	0.003	0.011	0.006
<i>Dehalococcoides ethenogenes</i> 195	1,469,720	0.051	0.9903203	1.0188511	0.008	0.387	0.605	0	0.02	0.031
<i>Desulfovibrio vulgaris</i> str. Hildenborough	3,570,858	0.084	1.0029051	0.999543	0.598	0.176	0.225	0.05	0.015	0.019
<i>Geobacter sulfurreducens</i> PCA	3,814,139	0.058	0.9930198	1.0037476	0.598	0.228	0.174	0.035	0.013	0.01
<i>Haemophilus ducreyi</i> 35000HP	1,698,955	0.069	0.9769756	1.0645451	0.593	0.31	0.097	0.041	0.021	0.005
<i>Legionella pneumophila</i> subsp. pneumophila str. Philadelphia 1	3,397,754	0.064	1.0004093	1.0047456	0.627	0.317	0.056	0.04	0.02	0.004
<i>Mycobacterium avium</i> subsp. paratuberculosis K-10	4,829,781	0.154	1.0052821	0.9972724	0.717	0.134	0.148	0.11	0.021	0.023
<i>Mycobacterium bovis</i> AF2122/97	4,345,492	0.108	0.9992664	0.9957569	0.663	0.159	0.178	0.072	0.017	0.019
<i>Neisseria gonorrhoeae</i> FA 1090	2,153,922	0.065	0.9879401	1.0029617	0.032	0.718	0.249	0.002	0.047	0.016
<i>Pseudomonas putida</i> KT2440	6,181,863	0.082	0.9953694	1.0123958	0.473	0.228	0.299	0.039	0.019	0.025
<i>Porphyromonas gingivalis</i> W83	2,343,476	0.026	0.9973137	1.0052028	0.033	0.394	0.574	0.001	0.01	0.015
<i>Staphylococcus aureus</i> subsp. aureus COL	2,809,422	0.099	0.9994226	1.0130897	0.881	0.088	0.031	0.087	0.009	0.003
<i>Staphylococcus aureus</i> subsp. aureus MRSA252	2,902,619	0.099	0.987711	0.9984954	0.883	0.087	0.03	0.087	0.009	0.003
<i>Staphylococcus aureus</i> subsp. aureus MSSA476	2,799,802	0.098	0.996749	1.0126332	0.881	0.088	0.031	0.087	0.009	0.003
<i>Treponema denticola</i> ATCC 35,405	2,843,201	0.091	0.9962816	0.9825042	0.472	0.331	0.198	0.043	0.03	0.018
<i>Coxiella burnetii</i> RSA 493	1,995,281	0.048	1.0039307	1.0005806	0.326	0.628	0.046	0.016	0.03	0.002
<i>Listeria monocytogenes</i> str. 4b F2365	2,905,187	0.064	1.0070784	0.9872494	0.652	0.26	0.088	0.042	0.017	0.006
<i>Staphylococcus epidermidis</i> RP62A	2,616,530	0.102	0.9672591	0.9710915	0.923	0.052	0.025	0.094	0.005	0.003
<i>Methylococcus capsulatus</i> str. Bath	3,304,561	0.091	1.003071	0.9942682	0.59	0.218	0.193	0.054	0.02	0.018

(continued on next page)

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Wolbachia endosymbiont of Drosophila melanogaster	1,267,782	0.082	0.9906056	0.9914362	0.776	0.167	0.056	0.064	0.014	0.005
Streptococcus pneumoniae TIGR4	2,160,842	0.054	1.0072571	1.0098479	0.568	0.307	0.125	0.031	0.017	0.007
Clostridium acetobutylicum ATCC 824	3,940,880	0.13	1.0017765	1.0052911	0.826	0.117	0.057	0.108	0.015	0.007
Sinorhizobium meliloti 1021	3,654,135	0.102	1.0002866	0.9922367	0.465	0.379	0.156	0.047	0.038	0.016
Agrobacterium tumefaciens str. C58	2,841,580	0.082	1.0140306	0.9808372	0.31	0.45	0.24	0.026	0.037	0.02
Streptococcus pneumoniae R6	2,038,615	0.055	1.0025761	1.0058961	0.564	0.309	0.128	0.031	0.017	0.007
Rickettsia conorii str. Malish 7	1,268,755	0.109	0.9941004	1.0127489	0.833	0.1	0.067	0.091	0.011	0.007
Sulfolobus tokodaii str. 7	2,694,756	0.105	0.9896391	1.0132709	0.83	0.131	0.04	0.087	0.014	0.004
Neisseria meningitidis MC58	2,272,360	0.057	0.9973706	1.0162275	0.012	0.733	0.255	0.001	0.042	0.014
Neisseria meningitidis Z2491	2,184,406	0.058	0.9918147	1.0009087	0.016	0.734	0.249	0.001	0.043	0.014
Yersinia pestis CO92	4,653,728	0.027	1.0017803	1.0104426	0.059	0.49	0.451	0.002	0.013	0.012
Streptomyces avermitilis MA-4680	9,025,608	0.159	1.0096042	0.999179	0.801	0.103	0.096	0.128	0.016	0.015
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	4,857,432	0.042	1.0008623	0.9998487	0.034	0.495	0.471	0.001	0.021	0.02
Salmonella enterica subsp. enterica serovar Typhi str. CT18	4,809,037	0.04	0.9969912	1.0030481	0.031	0.491	0.478	0.001	0.02	0.019
Listeria monocytogenes EGD-e	2,944,528	0.064	1.0024542	0.985372	0.656	0.256	0.087	0.042	0.016	0.006
Listeria innocua Clip11262	3,011,208	0.067	0.999426	0.9841823	0.684	0.24	0.077	0.046	0.016	0.005
Bacteroides fragilis NCTC 9343	5,205,140	0.035	0.9988927	1.0023691	0.386	0.215	0.399	0.013	0.007	0.014
Nostoc sp. PCC 7120	6,413,771	0.038	0.9974227	1.0032553	0.569	0.265	0.165	0.022	0.01	0.006
Ralstonia solanacearum GM11000	3,716,413	0.136	0.9873271	1.0117041	0.628	0.227	0.146	0.085	0.031	0.02
Brucella melitensis 16 M	2,117,144	0.072	0.9970786	1.0064415	0.207	0.566	0.226	0.015	0.041	0.016
Chlamydomonas reinhardtii GPC	1,173,390	0.057	1.0038721	0.9968588	0.598	0.265	0.137	0.034	0.015	0.008
Pyrobaculum aerophilum str. IM2	2,222,430	0.032	1.0146492	1.0206437	0.018	0.34	0.642	0.001	0.011	0.021
Clostridium perfringens str. 13	3,031,430	0.169	0.9591412	0.9384325	0.813	0.149	0.038	0.137	0.025	0.006
Pyrococcus furiosus DSM 3638	1,908,256	0.064	1.0000885	1.0018939	0.389	0.482	0.129	0.025	0.031	0.008
Corynebacterium glutamicum ATCC 13,032	3,309,401	0.029	1.0001256	0.990736	0.147	0.59	0.263	0.004	0.017	0.008
Fusobacterium nucleatum subsp. nucleatum ATCC 25,586	2,174,500	0.196	0.9685618	0.9400734	0.801	0.156	0.043	0.157	0.031	0.008
Streptococcus pyogenes MGAS8232	1,895,017	0.059	0.9963178	1.0057968	0.643	0.247	0.11	0.038	0.015	0.007
Methanopyrus kandleri AV19	1,694,969	0.072	1.0046042	0.990608	0.5	0.324	0.176	0.036	0.023	0.013
Methanosarcina acetivorans C2A	5,751,492	0.051	0.987523	0.9983458	0.305	0.397	0.298	0.016	0.02	0.015
Thermoanaerobacter tengcongensis MB4	2,689,445	0.082	0.9962358	0.99778	0.548	0.321	0.13	0.045	0.026	0.011
Streptomyces coelicolor A3(2)	8,667,507	0.179	0.9921675	1.002687	0.815	0.089	0.095	0.146	0.016	0.017
Methanosarcina mazei Go1	4,096,345	0.062	1.004251	1.0015888	0.338	0.357	0.305	0.021	0.022	0.019
Xanthomonas campestris pv. campestris str. ATCC 33,913	5,076,188	0.121	1.0060501	1.0007644	0.548	0.283	0.169	0.067	0.034	0.021
Bacillus cereus ATCC 10,987	5,224,283	0.075	1.0029069	1.0033618	0.811	0.098	0.091	0.061	0.007	0.007
Colwellia psychrerythraea 34H	5,373,180	0.06	1.0022915	1.0037993	0.695	0.181	0.125	0.042	0.011	0.008
Ruegeria pomeroyi DSS-3	4,109,442	0.11	0.9997742	1.0053829	0.539	0.243	0.218	0.059	0.027	0.024
Campylobacter jejuni RM1221	1,777,831	0.163	1.006554	0.9913149	0.706	0.231	0.063	0.115	0.038	0.01
Xanthomonas axonopodis pv. citri str. 306	5,175,554	0.12	1.000915	0.9974568	0.532	0.29	0.178	0.064	0.035	0.021
Staphylococcus aureus subsp. aureus MW2	2,820,462	0.099	0.9894798	1.0031888	0.882	0.088	0.031	0.087	0.009	0.003
Bacillus anthracis str. Ames	5,227,293	0.077	0.9956907	0.988427	0.816	0.094	0.09	0.063	0.007	0.007
Buchnera aphidicola str. Sg (Schizaphis graminum)	641,454	0.206	1.0076877	1.0211461	0.89	0.07	0.039	0.184	0.015	0.008
Streptococcus pyogenes MGAS315	1,900,521	0.059	0.9856757	0.9914424	0.641	0.247	0.112	0.038	0.015	0.007
Yersinia pestis KIM	4,600,755	0.027	0.9931358	1.0100093	0.059	0.491	0.45	0.002	0.013	0.012
Thermosynechococcus elongatus BP-1	2,593,857	0.033	1.004207	1.0016175	0.136	0.668	0.196	0.004	0.022	0.006
Streptococcus agalactiae 2603 V/R	2,160,267	0.074	1.0057755	0.993636	0.809	0.134	0.057	0.06	0.01	0.004
Pseudomonas fluorescens Pf-5	7,074,893	0.094	1.011285	0.99349	0.552	0.19	0.258	0.052	0.018	0.024
Oceanobacillus iheyensis HTE831	3,630,528	0.072	0.9965764	0.9981369	0.831	0.091	0.078	0.06	0.007	0.006
Bifidobacterium longum NCC2705	2,256,640	0.065	1.0034729	1.0002269	0.457	0.303	0.24	0.03	0.02	0.016
Brucella suis 1330	2,107,794	0.072	1.0014202	0.9934476	0.209	0.565	0.226	0.015	0.041	0.016
Shigella flexneri 2a str. 301	4,607,203	0.035	1.0043879	0.9926449	0.007	0.478	0.515	0	0.017	0.018
Leptospira interrogans serovar Lai str. 56,601	4,332,241	0.107	1.0022464	1.0108732	0.616	0.317	0.068	0.066	0.034	0.007
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	697,724	0.263	1.0026622	0.9951025	0.878	0.081	0.041	0.231	0.021	0.011
Shewanella oneidensis MR-1	4,969,803	0.029	1.0017395	1.0025611	0.163	0.523	0.313	0.005	0.015	0.009
Streptococcus mutans UA159	2,030,921	0.076	0.9921186	0.9864988	0.668	0.256	0.076	0.051	0.019	0.006
Streptococcus agalactiae NEM316	2,211,485	0.075	1.0041333	0.9980752	0.806	0.139	0.055	0.06	0.01	0.004
Corynebacterium efficiens YS-314	3,147,090	0.084	0.9970386	1.0044731	0.597	0.232	0.171	0.05	0.02	0.014
Escherichia coli CFT073	5,231,428	0.035	1.0034855	0.9984131	0.002	0.481	0.518	0	0.017	0.018
Mycoplasma penetrans HF-2	1,358,633	0.212	0.9950223	1.001541	0.837	0.133	0.03	0.178	0.028	0.006
Vibrio vulnificus CMCP6	3,281,944	0.031	0.9921782	1.0081064	0.116	0.582	0.302	0.004	0.018	0.009
Staphylococcus epidermidis ATCC 12,228	2,499,279	0.103	0.9753836	0.9784581	0.923	0.053	0.024	0.095	0.005	0.002
Bradyrhizobium japonicum USDA 110	9,105,828	0.111	0.9985642	1.000769	0.52	0.354	0.126	0.058	0.039	0.014
Buchnera aphidicola str. Bp (Baizongia pistaciae)	615,980	0.196	0.9886102	1.0020778	0.936	0.033	0.031	0.183	0.007	0.006
Pectobacterium atrosepticumSCRI1043	5,064,019	0.031	1.005306	1.0005107	0.009	0.518	0.473	0	0.016	0.015
Tropheryma whippelii TW08/27	925,938	0.022	0.9729673	1.0040515	0.179	0.381	0.44	0.004	0.008	0.01
Chlamydomonas reinhardtii S26/3	1,144,377	0.048	1.0014222	0.9926883	0.622	0.243	0.135	0.03	0.012	0.006
Xylella fastidiosa Temecula1	2,519,802	0.024	0.9925391	1.0220674	0.04	0.718	0.242	0.001	0.017	0.006
Clostridium tetani E88	2,799,251	0.164	0.9826141	0.962502	0.82	0.127	0.054	0.135	0.021	0.009
Lactobacillus plantarum WCFS1	3,308,274	0.036	0.9976934	0.9958227	0.246	0.333	0.421	0.009	0.012	0.015
Tropheryma whippelii str. Twist	927,303	0.022	1.0274875	0.9855715	0.178	0.38	0.442	0.004	0.009	0.01
Pseudomonas syringae pv. tomato str. DC3000	6,397,126	0.062	1.0009552	0.9982795	0.331	0.37	0.299	0.02	0.023	0.018
Vibrio parahaemolyticus RIMD 2,210,633	3,288,558	0.031	0.9896739	0.9990689	0.196	0.502	0.302	0.006	0.016	0.01
Streptococcus pyogenes SSI-1	1,894,275	0.06	1.0110311	1.0239799	0.642	0.246	0.112	0.038	0.015	0.007

Table 1 (continued)

Organism	Length	Id (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Salmonella enterica subsp. enterica serovar Typhi str. Ty2	4,791,961	0.04	0.9965077	1.0056716	0.03	0.493	0.477	0.001	0.02	0.019
Bacteroides thetaiotaomicron VPI-5482	6,260,361	0.037	0.9912433	1.0068864	0.402	0.183	0.416	0.015	0.007	0.015
Enterococcus faecalis V583	3,218,031	0.069	0.9923057	0.9941034	0.658	0.287	0.055	0.045	0.02	0.004
Bacillus cereus ATCC 14,579	5,411,809	0.077	1.0009111	1.0020531	0.823	0.093	0.084	0.063	0.007	0.006
Shigella flexneri 2a str. 2457 T	4,599,354	0.036	1.0037394	0.9954713	0.007	0.478	0.516	0	0.017	0.018
Nitrosomonas europaea ATCC 19,718	2,812,094	0.042	1.016088	0.96223	0.007	0.442	0.55	0	0.018	0.023
Mycoplasma gallisepticum R	996,422	0.134	1.0125193	1.0003765	0.758	0.072	0.169	0.102	0.01	0.023
Anaplasma marginale str. St. Maries	1,197,687	0.016	1.0032095	0.9816891	0.003	0.627	0.37	0	0.01	0.006
Helicobacter hepaticus ATCC 51,449	1,799,146	0.106	1.0100057	0.9700619	0.548	0.255	0.198	0.058	0.027	0.021
Rhodopirellula baltica SH 1	7,145,576	0.072	1.0119857	0.9846517	0.117	0.675	0.207	0.008	0.049	0.015
Prochlorococcus marinus subsp. marinus str. CCMP1375	1,751,080	0.082	1.0059353	1.0029218	0.658	0.309	0.033	0.054	0.025	0.003
Chlamydomydia pneumoniae TW-183	1,225,935	0.051	1.009708	0.9960916	0.502	0.379	0.119	0.026	0.019	0.006
Candidatus Blochmannia floridanus	705,557	0.165	1.007173	1.0760112	0.928	0.037	0.036	0.153	0.006	0.006
Synechococcus sp. WH 8102	2,434,428	0.067	0.9906538	1.000596	0.382	0.374	0.243	0.026	0.025	0.016
Prochlorococcus marinus str. MIT 9313	2,410,873	0.036	1.0859257	0.9371186	0.031	0.735	0.233	0.001	0.027	0.009
Prochlorococcus marinus subsp. pastoris str. CCMP1986	1,657,990	0.136	0.9923474	0.9934538	0.8	0.185	0.015	0.109	0.025	0.002
Chromobacterium violaceum ATCC 12,472	4,751,080	0.119	0.9948244	0.9997736	0.541	0.231	0.228	0.064	0.027	0.027
Wolinella succinogenes DSM 1740	2,110,355	0.073	0.9928309	1.0233353	0.01	0.56	0.43	0.001	0.041	0.032
Gloeobacter violaceus PCC 7421	4,659,019	0.07	1.0060675	0.9995833	0.595	0.205	0.199	0.042	0.014	0.014
Photorhabdus luminescens subsp. laumondii TTO1	5,688,987	0.037	0.9999545	0.9780328	0.408	0.335	0.257	0.015	0.012	0.009
Vibrio vulnificus YJ016	3,354,505	0.031	1.000935	0.9973915	0.119	0.575	0.306	0.004	0.018	0.01
Nanoarchaeum equitans Kin4-M	490,885	0.145	0.999988	1.0025718	0.692	0.181	0.127	0.1	0.026	0.008
Ehrlichia ruminantium str. Welgevonden	1,516,355	0.165	1.0097869	0.972954	0.918	0.059	0.023	0.152	0.01	0.004
Rhodospseudomonas palustris CGA009	5,459,213	0.131	0.9967338	1.0032538	0.507	0.317	0.176	0.066	0.041	0.023
Onion yellows phytoplasma OY-M	860,631	0.187	1.0219583	0.9915568	0.791	0.187	0.022	0.148	0.035	0.004
Lactobacillus johnsonii NCC 533	1,992,676	0.089	1.0098534	1.0225216	0.782	0.148	0.069	0.069	0.013	0.006
Bdellovibrio bacteriovorus HD100	3,782,950	0.046	0.9968789	1.0038807	0.003	0.661	0.336	0	0.03	0.015
Mycoplasma mycoides subsp. mycoides SC str. PG1	1,211,703	0.232	0.9820856	0.9877882	0.885	0.084	0.031	0.206	0.019	0.007
Pseudomonas syringae pv. phaseolicola 1448A	5,928,787	0.059	0.9927902	1.0049097	0.316	0.365	0.318	0.019	0.022	0.019
Methanococcus maripaludis S2	1,661,137	0.116	1.0124756	1.0052074	0.725	0.198	0.077	0.084	0.023	0.009
Yersinia pestis biovar Microtus str. 91,001	4,595,065	0.028	0.9977252	1.0062748	0.058	0.491	0.451	0.002	0.014	0.012
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	4,277,185	0.107	1.0013271	0.9938727	0.613	0.318	0.07	0.065	0.034	0.007
Thermus thermophilus HB27	1,894,877	0.196	1.0074801	1.0035861	0.572	0.221	0.207	0.112	0.043	0.04
Candidatus Protochlamydia amoebophila UWE25	2,414,465	0.097	0.9954413	1.0050425	0.705	0.283	0.012	0.068	0.027	0.001
Picrophilus torridus DSM 9790	1,545,895	0.1	0.9980824	0.9842738	0.574	0.225	0.201	0.058	0.023	0.02
Bacillus anthracis str. Sterne	5,228,663	0.077	0.9955426	0.9882736	0.816	0.094	0.09	0.063	0.007	0.007
Bartonella quintana str. Toulouse	1,581,384	0.067	1.0137593	0.9791448	0.547	0.355	0.098	0.037	0.024	0.007
Bartonella henselae str. Houston-1	1,931,047	0.071	1.0281556	0.9796686	0.573	0.322	0.105	0.04	0.023	0.007
Bacillus thuringiensis serovar konkukian str. 97–27	5,237,682	0.076	0.9960835	0.9912455	0.816	0.094	0.09	0.062	0.007	0.007
Acinetobacter sp. ADP1	3,598,621	0.056	1.0013482	1.0097146	0.474	0.352	0.174	0.027	0.02	0.01
Mesoplasma florum L1	793,224	0.186	0.9953844	0.9892991	0.852	0.118	0.029	0.158	0.022	0.005
Propionibacterium acnes KPA171202	2,560,265	0.051	0.9953327	0.997722	0.573	0.253	0.175	0.029	0.013	0.009
Streptococcus pyogenes MGAS10394	1,899,877	0.058	0.9903109	0.9950162	0.638	0.252	0.11	0.037	0.015	0.006
Leifsonia xyli subsp. xyli str. CTCB07	2,584,158	0.13	1.0049734	0.9990214	0.71	0.214	0.076	0.092	0.028	0.01
Desulfotalea psychrophila LSV54	3,523,383	0.032	1.0253377	0.9799715	0.096	0.556	0.348	0.003	0.018	0.011
Rickettsia typhi str. Wilmington	1,111,496	0.144	0.9995342	0.9724842	0.919	0.052	0.029	0.132	0.007	0.004
Yersinia pseudotuberculosis IP 32,953	4,744,671	0.028	0.997524	1.0099691	0.059	0.493	0.448	0.002	0.014	0.013
Borrelia garinii PBI	904,246	0.175	0.9893674	0.9846848	0.804	0.181	0.014	0.14	0.032	0.002
Symbiobacterium thermophilum IAM 14,863	3,566,135	0.129	1.0032647	0.9977418	0.796	0.061	0.143	0.103	0.008	0.019
Bacillus licheniformis ATCC 14,580	4,222,597	0.058	0.9957311	1.0005322	0.072	0.597	0.331	0.004	0.034	0.019
Bacillus cereus E33L	5,300,915	0.077	0.9973078	0.9915873	0.817	0.093	0.09	0.063	0.007	0.007
Mannheimia succiniciproducens MBEL55E	2,314,078	0.064	1.0001803	0.9824105	0.25	0.599	0.151	0.016	0.039	0.01
Bacillus licheniformis ATCC 14,580	4,222,645	0.058	0.9954302	1.0004595	0.072	0.597	0.331	0.004	0.034	0.019
Bacteroides fragilis YCH46	5,277,274	0.034	0.9950568	0.9988584	0.383	0.219	0.398	0.013	0.007	0.014
Burkholderia mallei ATCC 23,344	2,325,379	0.208	0.9937663	1.0006337	0.513	0.421	0.066	0.107	0.088	0.014
Burkholderia pseudomallei K96243	4,074,542	0.186	0.9890853	0.9954101	0.498	0.433	0.069	0.093	0.081	0.013
Mycoplasma hyopneumoniae 232	892,758	0.19	0.9884704	0.9983068	0.72	0.245	0.035	0.137	0.047	0.007
Nocardia farcinica IFM 10,152	6,021,225	0.176	1.0002359	1.0026211	0.732	0.135	0.134	0.129	0.024	0.024
Legionella pneumophila str. Paris	3,503,610	0.063	1.0055661	0.999332	0.624	0.321	0.055	0.039	0.02	0.003
Legionella pneumophila str. Lens	3,345,687	0.063	1.0105402	1.01008	0.618	0.322	0.06	0.039	0.02	0.004
Photobacterium profundum SS9	4,085,304	0.038	0.9949729	1.0055164	0.492	0.271	0.237	0.019	0.01	0.009
Haloarcula marismortui ATCC 43,049	3,131,724	0.086	1.0021502	0.9995248	0.516	0.343	0.141	0.045	0.03	0.012
Streptococcus thermophilus LMG 18,311	1,796,846	0.054	1.0109713	1.0083175	0.642	0.226	0.132	0.035	0.012	0.007
Streptococcus thermophilus CNRZ1066	1,796,226	0.054	1.0108772	1.0091963	0.642	0.227	0.131	0.035	0.012	0.007
Thermus thermophilus HB8	1,849,742	0.197	1.0178364	0.9897948	0.573	0.22	0.207	0.113	0.043	0.041
Geobacillus kaustophilus HTA426	3,544,776	0.06	1.0031004	0.9935314	0.021	0.768	0.211	0.001	0.046	0.013
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150	4,585,229	0.041	0.9987569	0.9929584	0.033	0.499	0.468	0.001	0.021	0.019
Idiomarina loihiensis L2TR	2,839,318	0.029	1.003053	0.9437947	0.098	0.468	0.435	0.003	0.013	0.013
Aromatoleum aromaticum EbN1	4,296,230	0.122	1.0028039	1.0014208	0.549	0.332	0.12	0.067	0.041	0.015
Zymomonas mobilis subsp. mobilis ZM4	2,056,416	0.053	0.9618914	1.0652478	0.082	0.587	0.331	0.004	0.031	0.018
Francisella tularensis subsp. tularensis SCHU S4	1,892,819	0.115	1.0093636	1.0270083	0.808	0.112	0.08	0.093	0.013	0.009
Synechococcus elongatus PCC 6301	2,696,255	0.048	0.9975172	0.9998598	0.181	0.44	0.378	0.009	0.021	0.018
Bacillus clausii KSM-K16	4,303,871	0.039	0.9981479	0.9957733	0.206	0.657	0.137	0.008	0.025	0.005

(continued on next page)

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
<i>Thermococcus kodakarensis</i> KOD1	2,088,737	0.033	1.0083964	0.9968248	0.035	0.597	0.367	0.001	0.02	0.012
<i>Gluconobacter oxydans</i> 621H	2,702,173	0.08	0.9845226	1.0294764	0.446	0.306	0.248	0.036	0.025	0.02
<i>Lactobacillus acidophilus</i> NCFM	1,993,560	0.089	1.0012948	1.0333303	0.773	0.128	0.099	0.069	0.011	0.009
<i>Ehrlichia ruminantium</i> str. Gardel	1,499,920	0.165	1.009814	0.9753999	0.918	0.059	0.023	0.151	0.01	0.004
<i>Ehrlichia ruminantium</i> str. Welgevonden	1,512,977	0.165	1.0100649	0.973049	0.918	0.059	0.023	0.152	0.01	0.004
<i>Wolbachia endosymbiont</i> strain TRS of <i>Brugia malayi</i>	1,080,084	0.091	1.0186528	0.9340865	0.81	0.143	0.046	0.074	0.013	0.004
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	4,941,439	0.109	1.0020618	0.9987087	0.504	0.319	0.176	0.055	0.035	0.019
<i>Vibrio fischeri</i> ES114	2,897,536	0.054	0.9963479	0.9969743	0.655	0.179	0.166	0.036	0.01	0.009
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Choleraesuis</i> str. SC-B67	4,755,700	0.041	1.0017753	1.0018546	0.033	0.496	0.471	0.001	0.02	0.019
<i>Mycoplasma mobile</i> 163 K	777,079	0.23	0.9963169	0.9922422	0.823	0.162	0.015	0.189	0.037	0.003
<i>Brucella abortus</i> bv. 1 str. 9–941	2,124,241	0.072	1.00248	0.9940711	0.207	0.567	0.226	0.015	0.041	0.016
<i>Corynebacterium glutamicum</i> ATCC 13,032	3,282,708	0.029	0.9995854	0.9932762	0.148	0.588	0.263	0.004	0.017	0.008
<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	6,093,698	0.066	0.9963244	0.9976985	0.372	0.327	0.301	0.025	0.022	0.02
<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. 8004	5,148,708	0.12	0.9977897	0.9948346	0.547	0.284	0.168	0.066	0.034	0.02
<i>Rickettsia felis</i> URRWXCa2	1,485,148	0.11	1.0159796	1.0203068	0.825	0.107	0.067	0.091	0.012	0.007
<i>Haemophilus influenzae</i> 86-028NP	1,914,490	0.071	0.998218	1.0008878	0.576	0.342	0.082	0.041	0.024	0.006
<i>Corynebacterium jeikeium</i> K411	2,462,499	0.055	0.9862124	0.9931446	0.686	0.148	0.166	0.038	0.008	0.009
<i>Staphylococcus haemolyticus</i> JCS1435	2,685,015	0.096	1.0117558	1.0129354	0.91	0.053	0.037	0.087	0.005	0.004
<i>Sulfolobus acidocaldarius</i> DSM 639	2,225,959	0.069	0.9938228	1.0024996	0.752	0.197	0.05	0.052	0.014	0.003
<i>Psychrobacter arcticus</i> 273–4	2,650,701	0.047	0.9972273	1.0031784	0.316	0.316	0.368	0.015	0.015	0.017
<i>Candidatus Pelagibacter ubique</i> HTCC1062	1,308,759	0.152	1.0085579	1.0093729	0.806	0.164	0.03	0.123	0.025	0.005
<i>Candidatus Blochmannia pennsylvanicus</i> str. BPEN	791,654	0.135	1.0321513	0.9949877	0.921	0.049	0.03	0.124	0.007	0.004
<i>Mycoplasma synoviae</i> 53	799,476	0.171	0.9597804	0.9977026	0.808	0.158	0.034	0.138	0.027	0.006
<i>Mycoplasma hyopneumoniae</i> J	897,405	0.19	0.9797011	1.001204	0.725	0.242	0.033	0.138	0.046	0.006
<i>Streptococcus pyogenes</i> MGAS6180	1,897,573	0.061	0.9861238	0.9947864	0.65	0.242	0.109	0.04	0.015	0.007
<i>Streptococcus pyogenes</i> MGAS5005	1,838,554	0.06	0.9925996	0.9987842	0.64	0.249	0.111	0.038	0.015	0.007
<i>Dechloromonas aromatica</i> RCB	4,501,104	0.071	0.9940266	1.0003802	0.35	0.35	0.3	0.025	0.025	0.021
<i>Mycoplasma hyopneumoniae</i> 7448	920,079	0.191	0.974584	0.9959183	0.724	0.242	0.034	0.138	0.046	0.006
<i>Thermobifida fusca</i> YX	3,642,249	0.11	1.0007456	1.0018318	0.82	0.096	0.084	0.09	0.011	0.009
<i>Prochlorococcus marinus</i> str. NATL2A	1,842,899	0.091	1.0045856	1.0040494	0.71	0.263	0.027	0.065	0.024	0.002
<i>Ralstonia eutropha</i> JMP134	3,806,533	0.114	0.993171	0.9960305	0.555	0.271	0.174	0.063	0.031	0.02
<i>Staphylococcus saprophyticus</i> subsp. <i>saprophyticus</i> ATCC 15,305	2,516,575	0.094	1.0043007	1.0120701	0.88	0.08	0.039	0.083	0.008	0.004
<i>Ehrlichia canis</i> str. Jake	1,315,030	0.148	1.0037343	0.9880604	0.892	0.08	0.028	0.132	0.012	0.004
<i>Methanosarcina barkeri</i> str. <i>Fusaro</i>	4,837,408	0.062	1.0095434	0.9922195	0.54	0.282	0.178	0.033	0.018	0.011
<i>Dehalococcoides</i> sp. CBDB1	1,395,502	0.047	1.0011559	1.0050871	0.055	0.384	0.561	0.003	0.018	0.026
<i>Shigella sonnei</i> Ss046	4,825,265	0.036	1.0014438	1.0020055	0.008	0.472	0.52	0	0.017	0.019
<i>Thiobacillus denitrificans</i> ATCC 25,259	2,909,809	0.131	1.0005495	1.0012325	0.578	0.314	0.108	0.076	0.041	0.014
<i>Nitrobacter winogradskyi</i> Nb-255	3,402,093	0.097	1.0064739	0.9941422	0.436	0.422	0.142	0.042	0.041	0.014
<i>Anabaena variabilis</i> ATCC 29,413	6,365,727	0.038	1.0063913	1.0040554	0.566	0.266	0.168	0.021	0.01	0.006
<i>Natronomonas pharaonis</i> DSM 2160	2,595,221	0.108	0.9965094	1.0018336	0.487	0.335	0.179	0.053	0.036	0.019
<i>Chlamydia trachomatis</i> A/HAR-13	1,044,459	0.05	1.0046498	1.00135	0.439	0.391	0.17	0.022	0.02	0.008
<i>Streptococcus agalactiae</i> A909	2,127,839	0.075	1.0156281	1.011202	0.812	0.132	0.056	0.061	0.01	0.004
<i>Burkholderia pseudomallei</i> 1710b	4,126,292	0.183	0.990953	0.9971393	0.498	0.434	0.068	0.091	0.08	0.012
<i>Pseudoalteromonas haloplanktis</i> TAC125	635,328	0.074	1.002896	1.0638946	0.445	0.438	0.117	0.033	0.033	0.009
<i>Nitrosococcus oceanii</i> ATCC 19,707	3,481,691	0.023	1.0106761	0.9951119	0.002	0.819	0.179	0	0.019	0.004
<i>Pseudomonas fluorescens</i> Pf0-1	6,438,405	0.079	1.0089075	0.9955855	0.406	0.314	0.28	0.032	0.025	0.022
<i>Rhodobacter sphaeroides</i> 2.4.1	943,016	0.156	1.0191466	0.999659	0.687	0.212	0.101	0.107	0.033	0.016
<i>Pelobacter carbinolicus</i> DSM 2380	3,665,893	0.043	0.9958275	1.0044753	0.174	0.45	0.376	0.008	0.02	0.016
<i>Carboxydotherrmus hydrogenoformans</i> Z-2901	2,401,520	0.077	1.0020729	0.9983414	0.238	0.583	0.179	0.018	0.045	0.014
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> str. 85–10	5,178,466	0.12	1.0026866	1.002185	0.529	0.291	0.179	0.064	0.035	0.022
<i>Burkholderia</i> sp. 383	1,395,069	0.137	0.9998267	0.9982422	0.499	0.385	0.115	0.068	0.053	0.016
<i>Pelodictyon luteolum</i> DSM 273	2,364,842	0.051	1.0238815	0.9939392	0.305	0.339	0.356	0.016	0.017	0.018
<i>Synechococcus</i> sp. CC9902	2,234,828	0.042	0.9923272	0.9962415	0.121	0.643	0.236	0.005	0.027	0.01
<i>Chlorobium chlorochromatii</i> Cad3	2,572,079	0.053	1.0060002	1.0154355	0.181	0.586	0.233	0.009	0.031	0.012
<i>Synechococcus</i> sp. CC9605	2,510,659	0.067	1.0017317	1.0024902	0.369	0.413	0.217	0.025	0.028	0.015
<i>Geobacter metallireducens</i> GS-15	3,997,420	0.051	1.0122818	0.9918288	0.518	0.315	0.167	0.026	0.016	0.008
<i>Desulfovibrio desulfuricans</i> subsp. <i>desulfuricans</i> str. G20	3,730,232	0.068	1.0037915	1.0064468	0.263	0.324	0.413	0.018	0.022	0.028
<i>Thiomicrospira crunogena</i> XCL-2	2,427,734	0.046	1.0059643	1.0201907	0.301	0.528	0.171	0.014	0.024	0.008
<i>Bacillus anthracis</i> str. 'Ames Ancestor'	5,227,419	0.077	0.9956929	0.9884153	0.816	0.094	0.09	0.063	0.007	0.007
<i>Sulfurimonas denitrificans</i> DSM 1251	2,201,561	0.11	1.0146272	1.0127363	0.643	0.185	0.172	0.071	0.02	0.019
<i>Lactobacillus sakei</i> subsp. <i>sakei</i> 23 K	1,884,661	0.051	1.0018822	0.9940605	0.438	0.322	0.24	0.022	0.016	0.012
<i>Prochlorococcus marinus</i> str. MIT 9312	1,709,204	0.133	0.9958971	0.999239	0.783	0.2	0.017	0.104	0.027	0.002
<i>Synechococcus elongatus</i> PCC 7942	2,695,903	0.048	1.0028412	1.0019305	0.18	0.441	0.379	0.009	0.021	0.018
<i>Shigella dysenteriae</i> Sd197	4,369,232	0.036	0.9962924	0.9899755	0.013	0.475	0.513	0	0.017	0.018
<i>Shigella boydii</i> Sb227	4,519,823	0.036	1.0004543	1.0050152	0.012	0.468	0.521	0	0.017	0.019
<i>Nitrosospora multififormis</i> ATCC 25,196	3,184,243	0.034	1.0019956	0.9984259	0.131	0.579	0.29	0.004	0.02	0.01
<i>Brucella melitensis</i> biovar <i>Abortus</i> 2308	2,121,359	0.072	1.0022559	0.9943304	0.207	0.567	0.226	0.015	0.041	0.016
<i>Staphylococcus aureus</i> RF122	2,742,531	0.099	0.997352	1.0048772	0.884	0.086	0.03	0.087	0.008	0.003
<i>Magnetospirillum magneticum</i> AMB-1	4,967,148	0.108	1.0067967	0.9942584	0.618	0.19	0.192	0.067	0.021	0.021
<i>Mycoplasma capricolum</i> subsp. <i>capricolum</i> ATCC 27,343	1,010,023	0.238	1.009191	1.0171647	0.878	0.088	0.034	0.209	0.021	0.008
<i>Rhodospirillum rubrum</i> ATCC 11,170	4,352,825	0.127	1.0005906	1.0008071	0.549	0.226	0.224	0.07	0.029	0.029
<i>Moorella thermoacetica</i> ATCC 39,073	2,628,784	0.055	1.0051125	1.003266	0.178	0.417	0.405	0.01	0.023	0.022

Table 1 (continued)

Organism	Length	Id (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Hahella chejuensis KCTC 2396	7,215,267	0.032	0.9988357	0.9949118	0.134	0.484	0.382	0.004	0.016	0.012
Burkholderia thailandensis E264	3,809,201	0.182	0.9861224	1.0252233	0.485	0.446	0.07	0.088	0.081	0.013
Salinibacter ruber DSM 13,855	3,551,823	0.1	1.0058108	0.9987313	0.773	0.141	0.086	0.077	0.014	0.009
Methanospaera stadtmannae DSM 3091	1,767,403	0.173	1.0181048	0.9870503	0.867	0.088	0.045	0.15	0.015	0.008
Xanthomonas oryzae pv. oryzae MAFF 311,018	4,940,217	0.109	1.0050393	0.9963931	0.504	0.319	0.177	0.055	0.035	0.019
Sodalis glossinidius str. 'morsitans'	4,171,146	0.041	1.0020586	1.0162415	0.155	0.492	0.353	0.006	0.02	0.015
Aster yellows witches'-broom phytoplasma AYW8	706,569	0.2	0.9864075	1.0146349	0.801	0.182	0.017	0.16	0.036	0.003
Erythrobacter litoralis HTCC2594	3,052,398	0.102	1.0026218	1.0034699	0.489	0.357	0.154	0.05	0.036	0.016
Syntrophus aciditrophicus SB	3,179,300	0.049	1.0029371	1.0129492	0.013	0.649	0.338	0.001	0.032	0.016
Anaeromyxobacter dehalogenans 2CP-C	5,013,479	0.24	0.990481	0.9950139	0.779	0.123	0.098	0.187	0.03	0.023
Rhizobium etli CFN 42	4,381,608	0.097	1.0035236	0.9767416	0.382	0.418	0.201	0.037	0.041	0.019
Synechococcus sp. JA-3-3Ab	2,932,766	0.07	0.9997702	1.0023212	0.436	0.373	0.191	0.03	0.026	0.013
Synechococcus sp. JA-2-3B'a(2-13)	3,046,682	0.062	1.0061934	0.9975479	0.334	0.453	0.213	0.021	0.028	0.013
Frankia sp. Cc13	5,433,628	0.154	1.0026263	0.9996567	0.778	0.118	0.105	0.12	0.018	0.016
Rhodospseudomonas palustris HaA2	5,331,656	0.146	1.0013631	0.9962664	0.516	0.336	0.147	0.076	0.049	0.022
Staphylococcus aureus subsp. aureus USA300_FPR3757	2,872,769	0.099	0.9885182	0.9989636	0.883	0.086	0.031	0.088	0.009	0.003
Novosphingobium aromaticivorans DSM 12,444	3,561,584	0.115	1.0050536	0.9985292	0.585	0.259	0.156	0.067	0.03	0.018
Staphylococcus aureus subsp. aureus NCTC 8325	2,821,361	0.098	0.9826214	0.9907001	0.882	0.087	0.031	0.086	0.009	0.003
Methanospirillum hungatei JF-1	3,544,738	0.056	1.0192033	1.0034833	0.123	0.449	0.428	0.007	0.025	0.024
Anaplasma phagocytophilum HZ	1,471,282	0.035	0.9808079	1.0365838	0.586	0.209	0.206	0.02	0.007	0.007
Neorickettsia sennetsu str. Miyayama	859,006	0.04	0.9916922	0.9911701	0.581	0.253	0.166	0.023	0.01	0.007
Ehrlichia chaffeensis str. Arkansas	1,176,248	0.132	1.0096918	1.0344833	0.891	0.075	0.034	0.118	0.01	0.004
Jannaschia sp. CCS1	4,317,977	0.083	0.9931693	1.0216969	0.535	0.318	0.147	0.044	0.026	0.012
Francisella tularensis subsp. holarctica	1,895,994	0.116	0.9921483	1.0236353	0.809	0.112	0.078	0.094	0.013	0.009
Chlamydomonas reinhardtii Fe/C-56	1,166,239	0.055	0.9970399	1.0046053	0.592	0.284	0.124	0.033	0.016	0.007
Desulfotomobacterium hafniense Y51	5,727,534	0.035	0.9842052	0.924343	0.072	0.651	0.276	0.003	0.023	0.01
Rhodospirillum rubrum T118	4,712,337	0.075	0.9935519	1.0008069	0.376	0.435	0.19	0.028	0.033	0.014
Saccharophagus degradans 2-40	5,057,531	0.04	0.9999904	0.9972373	0.127	0.666	0.207	0.005	0.026	0.008
Rhodospseudomonas palustris BisB18	5,513,844	0.13	1.003369	0.9953987	0.503	0.338	0.159	0.066	0.044	0.021
Lactobacillus salivarius UCC118	1,827,111	0.1	1.0104687	1.0085265	0.856	0.084	0.06	0.086	0.008	0.006
Rickettsia bellii RML369-C	1,522,076	0.117	0.9829871	0.9701819	0.85	0.112	0.038	0.1	0.013	0.004
Escherichia coli UT189	5,065,741	0.036	0.9985067	0.9959768	0.003	0.49	0.507	0	0.018	0.018
Methylobacillus flagellatus KT	2,971,517	0.047	0.9955457	0.9973244	0.202	0.547	0.251	0.009	0.026	0.012
Polaromonas sp. JS666	5,200,264	0.089	1.0069797	1.0009196	0.511	0.316	0.173	0.045	0.028	0.015
Burkholderia xenovorans LB400	3,363,523	0.106	1.0003951	1.0028221	0.453	0.436	0.111	0.048	0.046	0.012
Shewanella denitrificans OS217	4,545,906	0.033	1.0039421	0.9948169	0.206	0.506	0.288	0.007	0.017	0.01
Methanococcoides burtonii DSM 6242	2,575,032	0.049	0.9946912	1.0170547	0.503	0.3	0.197	0.025	0.015	0.01
Rhodospseudomonas palustris BisB5	4,892,717	0.132	0.9963791	1.0066158	0.485	0.368	0.147	0.064	0.049	0.019
Chromohalobacter salexigens DSM 3043	3,696,649	0.095	0.9982862	1.0119913	0.594	0.243	0.163	0.057	0.023	0.016
Nitrobacter hamburgensis X14	4,406,967	0.09	1.0022467	1.0008254	0.443	0.418	0.139	0.04	0.038	0.013
Psychrobacter cryohalolentis K5	3,059,876	0.049	0.9958633	1.0037719	0.355	0.301	0.345	0.017	0.015	0.017
Ralstonia metallidurans CH34	3,928,089	0.101	0.9839778	0.9986002	0.55	0.271	0.179	0.056	0.027	0.018
Baumanniia cicadellinicola str. Hc (Homalodisca coagulata)	686,194	0.098	0.9982203	0.9989216	0.846	0.065	0.09	0.083	0.006	0.009
Candidatus Koribacter versatilis Ellin345	5,650,368	0.056	1.0019954	1.0024595	0.367	0.486	0.147	0.02	0.027	0.008
Lawsonia intracellularis PHE/MN1-00	1,457,619	0.102	1.0061726	1.0111887	0.809	0.121	0.07	0.082	0.012	0.007
Streptococcus pyogenes MGAS9429	1,836,467	0.06	1.007757	1.01957	0.642	0.247	0.111	0.038	0.015	0.007
Streptococcus pyogenes MGAS10270	1,928,252	0.06	1.001563	1.0075994	0.648	0.242	0.11	0.039	0.015	0.007
Streptococcus pyogenes MGAS2096	1,860,355	0.058	0.9973735	1.0030805	0.637	0.254	0.109	0.037	0.015	0.006
Streptococcus pyogenes MGAS10750	1,937,111	0.061	1.012916	1.0276074	0.654	0.241	0.105	0.04	0.015	0.006
Deinococcus geothermalis DSM 11,300	2,467,205	0.104	0.9950482	0.9968906	0.783	0.114	0.103	0.081	0.012	0.011
Pseudomonas entomophila L48	5,888,780	0.104	0.9984443	1.0034886	0.563	0.168	0.269	0.059	0.017	0.028
Ruegeria sp. TM1040	3,200,938	0.072	0.9993374	1.0004987	0.435	0.392	0.173	0.032	0.028	0.013
Sphingopyxis alaskensis RB2256	3,345,170	0.141	1.0101795	0.9976169	0.501	0.366	0.134	0.07	0.051	0.019
Lactobacillus delbrueckii subsp. bulgaricus ATCC 11,842	1,864,998	0.047	0.9936666	0.9983429	0.001	0.489	0.51	0	0.023	0.024
Burkholderia cenocepacia AU 1054	3,294,563	0.168	1.0094797	0.9911173	0.502	0.384	0.115	0.084	0.065	0.019
Helicobacter pylori HPAG1	1,596,366	0.102	0.9975289	0.996796	0.339	0.442	0.219	0.035	0.045	0.022
Myxococcus xanthus DK 1622	9,139,763	0.139	1.0053698	0.9965547	0.759	0.115	0.126	0.106	0.016	0.017
Mycobacterium sp. MCS	5,705,448	0.15	0.9995024	0.9991954	0.671	0.195	0.134	0.101	0.029	0.02
Rubrobacter xylanophilus DSM 9941	3,225,748	0.149	1.0021593	0.9941782	0.835	0.08	0.085	0.125	0.012	0.013
Yersinia pestis Nepal516	4,534,590	0.027	1.0042279	0.986421	0.062	0.489	0.449	0.002	0.013	0.012
Yersinia pestis Antiqua	4,702,289	0.027	1.0010863	0.9985441	0.056	0.49	0.454	0.002	0.013	0.012
Roseobacter denitrificans OCh 114	4,133,097	0.073	1.0008178	1.0012187	0.32	0.469	0.211	0.023	0.034	0.015
Haloquadratum walsbyi DSM 16,790	3,132,494	0.031	1.0032137	0.9895886	0.043	0.653	0.304	0.001	0.02	0.009
Pseudoalteromonas atlantica T6c	5,187,005	0.03	0.9984517	0.9972599	0.282	0.497	0.22	0.008	0.015	0.007
Helicobacter acinonychis str. Sheeba	1,553,927	0.104	1.0173919	1.0029679	0.39	0.396	0.214	0.041	0.041	0.022
Francisella tularensis subsp. tularensis FSC198	1,892,616	0.115	1.0092969	1.0268899	0.808	0.112	0.08	0.093	0.013	0.009
Escherichia coli 536	4,938,920	0.036	1.0012659	0.9934944	0.002	0.487	0.511	0	0.017	0.018
Mesorhizobium sp. BNC1	4,412,446	0.077	0.996169	0.9945484	0.465	0.398	0.137	0.036	0.031	0.011
Cytophaga hutchinsonii ATCC 33,406	4,433,218	0.074	1.0054503	1.001241	0.488	0.193	0.319	0.036	0.014	0.024
Shigella flexneri 5 str. 8401	4,574,284	0.036	1.0029709	0.995741	0.007	0.48	0.513	0	0.017	0.018
Alcanivorax borkumensis SK2	3,120,143	0.031	0.9996352	1.0020647	0.205	0.484	0.311	0.006	0.015	0.01
Clostridium perfringens ATCC 13,124	3,256,683	0.172	0.9693446	0.9624855	0.812	0.15	0.038	0.14	0.026	0.007
Clostridium perfringens SM101	2,897,393	0.171	0.957863	0.9329466	0.827	0.137	0.035	0.142	0.023	0.006
Rhodococcus jostii RHA1	7,804,765	0.139	0.9924062	1.0046869	0.652	0.246	0.102	0.09	0.034	0.014

(continued on next page)

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
<i>Borrelia afzelii</i> PKo	905,394	0.174	0.9893248	0.9819607	0.807	0.178	0.016	0.14	0.031	0.003
<i>Frankia alni</i> ACN14a	7,497,934	0.196	1.0028725	1.0027535	0.794	0.109	0.097	0.156	0.021	0.019
<i>Haemophilus somnus</i> 129PT	2,007,700	0.069	0.9974321	0.9989048	0.695	0.245	0.06	0.048	0.017	0.004
<i>Trichodesmium erythraeum</i> IMS101	7,750,108	0.093	0.999513	1.00393	0.79	0.159	0.051	0.074	0.015	0.005
<i>Ralstonia eutropha</i> H16	4,052,032	0.13	0.9920429	1.0014819	0.611	0.213	0.176	0.08	0.028	0.023
<i>Synechococcus</i> sp. CC9311	2,606,748	0.038	0.9964054	0.9963261	0.046	0.741	0.213	0.002	0.028	0.008
<i>Shewanella</i> sp. MR-4	4,706,287	0.025	1.0000035	1.0023299	0.051	0.551	0.398	0.001	0.014	0.01
<i>Shewanella</i> sp. MR-7	4,792,610	0.025	1.0075478	0.9961125	0.053	0.553	0.395	0.001	0.014	0.01
<i>Alkalilimnicola ehrlichii</i> MLHE-1	3,275,944	0.114	1.0076593	1.0008412	0.794	0.076	0.13	0.091	0.009	0.015
<i>Granulibacter thesedensis</i> CGDNIH1	2,708,355	0.07	1.00389	0.9975717	0.341	0.368	0.29	0.024	0.026	0.02
<i>Nitrosomonas eutropha</i> C91	2,661,057	0.033	1.0008405	1.0032251	0.02	0.551	0.429	0.001	0.018	0.014
<i>Shewanella frigidimarina</i> NCIMB 400	4,845,257	0.045	0.998969	1.0028798	0.456	0.322	0.223	0.021	0.014	0.01
<i>Syntrophomonas wolfei</i> subsp. <i>wolfei</i> str. Goettingen	2,936,195	0.043	1.0016246	0.9842642	0.176	0.528	0.296	0.008	0.023	0.013
<i>Maricaulis maris</i> MCS10	3,368,780	0.095	1.0028135	0.9996554	0.497	0.296	0.207	0.047	0.028	0.02
<i>Hyphomonas neptunium</i> ATCC 15,444	3,705,021	0.084	0.9937563	0.9964066	0.495	0.337	0.168	0.041	0.028	0.014
<i>Francisella tularensis</i> subsp. <i>holarctica</i> OSU18	1,895,727	0.116	0.9921543	1.0237827	0.809	0.112	0.078	0.094	0.013	0.009
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	5,057,142	0.096	1.0072926	0.976363	0.375	0.415	0.209	0.036	0.04	0.02
<i>Burkholderia ambifaria</i> AMMD	3,556,545	0.167	1.0044522	1.0005316	0.503	0.386	0.111	0.084	0.064	0.018
<i>Rhodopseudomonas palustris</i> BisA53	5,505,494	0.123	0.9979233	1.0006548	0.494	0.351	0.155	0.061	0.043	0.019
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	6,537,648	0.123	1.0008589	0.9996417	0.473	2.513	-1.986	0.078	0.016	0.029
<i>Lactobacillus brevis</i> ATCC 367	2,291,220	0.029	0.9945968	0.9439525	0.151	0.403	0.446	0.004	0.012	0.013
<i>Leptospira borgpetersenii</i> serovar <i>Hardjo-ovis</i> L550	3,614,446	0.148	1.0015305	1.0033829	0.188	-0.004	0.817	0.028	0.053	0.008
<i>Leptospira borgpetersenii</i> serovar <i>Hardjo-ovis</i> JB197	3,576,473	0.089	0.9978802	0.9905028	0.313	0.595	0.092	0.028	0.053	0.008
<i>Candidatus Carsonella ruddii</i> PV	159,662	0.383	1.0038056	0.9588919	0.92	0.063	0.017	0.352	0.024	0.006
<i>Buchnera aphidicola</i> str. Cc (<i>Cinara cedri</i>)	416,380	0.306	1.0066224	1.0071479	0.902	0.055	0.043	0.276	0.017	0.013
<i>Pediococcus pentosaceus</i> ATCC 25,745	1,832,387	0.065	0.9982608	1.0007541	0.714	0.16	0.126	0.047	0.01	0.008
<i>Lactobacillus casei</i> ATCC 334	2,895,264	0.041	0.993948	0.9979027	0.081	0.584	0.335	0.003	0.024	0.014
<i>Lactococcus lactis</i> subsp. <i>cremoris</i> SK11	2,438,589	0.087	0.996917	0.9972455	0.676	0.269	0.055	0.059	0.023	0.005
<i>Oenococcus oeni</i> PSU-1	1,780,517	0.086	1.0060675	1.0098975	0.497	0.37	0.132	0.043	0.032	0.011
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i> ATCC BAA-365	1,856,951	0.048	0.9959156	0.9981525	0.001	0.491	0.509	0	0.023	0.024
<i>Lactobacillus gasserii</i> ATCC 33,323	1,894,360	0.086	1.0166772	1.0205212	0.742	0.167	0.091	0.064	0.014	0.008
<i>Leuconostoc mesenteroides</i> subsp. <i>mesenteroides</i> ATCC 8293	2,038,396	0.065	0.9973447	1.0203057	0.681	0.233	0.086	0.044	0.015	0.006
<i>Streptococcus thermophilus</i> LMD-9	1,856,368	0.054	1.0150133	1.0025174	0.647	0.223	0.13	0.035	0.012	0.007
<i>Streptococcus pneumoniae</i> D39	2,046,115	0.055	1.0028395	1.0070707	0.565	0.308	0.127	0.031	0.017	0.007
<i>Candidatus Solibacter usitatus</i> Ellin6076	9,965,640	0.073	0.9973472	0.9947828	0.564	0.29	0.146	0.041	0.021	0.011
<i>Arthrobacter</i> sp. FB24	4,698,945	0.098	1.0054378	0.9982571	0.714	0.111	0.175	0.07	0.011	0.011
<i>Burkholderia cenocepacia</i> HI2424	3,483,902	0.166	1.0024758	1.0061264	0.498	0.385	0.117	0.083	0.064	0.019
<i>Methanoseta thermophila</i> PT	1,879,471	0.05	0.9996496	0.9881783	0.073	0.616	0.311	0.004	0.031	0.015
<i>Syntrophobacter fumaroxidans</i> MPOB	4,990,251	0.066	1.0115006	1.0017507	0.438	0.416	0.146	0.029	0.027	0.01
<i>Listeria welshimeri</i> serovar <i>6b</i> str. SLCC5334	2,814,130	0.075	1.0080327	0.9951888	0.723	0.201	0.076	0.054	0.015	0.006
<i>Escherichia coli</i> APEC O1	5,082,025	0.036	0.9993486	0.9986058	0.002	0.49	0.507	0	0.018	0.018
<i>Aeromonas hydrophila</i> subsp. <i>hydrophila</i> ATCC 7966	4,744,448	0.074	1.0027996	1.0028161	0.522	0.252	0.226	0.039	0.019	0.017
<i>Gramella forsetii</i> KT0803	3,798,465	0.081	0.9823014	0.9877599	0.648	0.228	0.125	0.052	0.018	0.01
<i>Magnetococcus</i> sp. MC-1	4,719,581	0.041	1.0098443	0.9763376	0.125	0.53	0.345	0.005	0.022	0.014
<i>Shewanella</i> sp. ANA-3	4,972,204	0.025	1.003044	1.0010056	0.044	0.559	0.397	0.001	0.014	0.01
<i>Acidothermus cellulolyticus</i> 11B	2,443,540	0.125	0.9971991	1.0015816	0.672	0.216	0.112	0.084	0.027	0.014
<i>Clostridium novyi</i> NT	2,547,720	0.156	1.0125914	1.0085624	0.855	0.096	0.049	0.133	0.015	0.008
<i>Mycobacterium avium</i> 104	5,475,491	0.149	0.9933461	1.0029082	0.714	0.139	0.147	0.107	0.021	0.022
<i>Mycobacterium smegmatis</i> str. MC2 155	6,988,209	0.139	1.0020544	0.9981747	0.642	0.222	0.136	0.089	0.031	0.019
<i>Campylobacter fetus</i> subsp. <i>fetus</i> 82-40	1,773,615	0.118	1.0297932	1.0243469	0.693	0.165	0.142	0.082	0.019	0.017
<i>Bacillus thuringiensis</i> str. <i>Al Hakam</i>	5,257,091	0.076	1.0008447	0.9983715	0.817	0.093	0.09	0.062	0.007	0.007
<i>Francisella novicida</i> U112	1,910,031	0.113	1.0061815	0.9776343	0.803	0.116	0.081	0.091	0.013	0.009
<i>Pelobacter propionicus</i> DSM 2379	4,008,000	0.05	1.0091618	1.0023944	0.471	0.279	0.25	0.024	0.014	0.013
<i>Candidatus Ruthia magnifica</i> str. Cm (<i>Calyptogenia magnifica</i>)	1,160,782	0.101	1.0009825	1.0090116	0.742	0.175	0.084	0.075	0.018	0.008
<i>Mycobacterium ulcerans</i> Agy99	5,631,606	0.107	1.0087004	0.9925423	0.656	0.166	0.179	0.07	0.018	0.019
<i>Bifidobacterium adolescentis</i> ATCC 15,703	2,089,645	0.071	1.0078912	0.9958461	0.345	0.417	0.237	0.024	0.03	0.017
<i>Chlorobium phaeobacteroides</i> DSM 266	3,133,902	0.041	0.9845577	0.9709462	0.022	0.525	0.454	0.001	0.021	0.019
<i>Paracoccus denitrificans</i> PD1222	2,852,282	0.139	1.0121842	0.9974637	0.59	0.238	0.172	0.082	0.033	0.024
<i>Thermophilum pendens</i> Hrk 5	1,781,889	0.046	0.9931122	0.9868014	0.368	0.381	0.251	0.017	0.018	0.012
<i>Nocardioides</i> sp. JS614	4,985,871	0.194	0.993845	1.0022081	0.722	0.144	0.134	0.14	0.028	0.026
<i>Shewanella amazonensis</i> SB2B	4,306,142	0.038	1.0011637	1.0060981	0.098	0.495	0.407	0.004	0.019	0.015
<i>Pyrobaculum islandicum</i> DSM 4184	1,826,402	0.029	1.001705	1.0036806	0.002	0.258	0.74	0	0.007	0.021
<i>Azoarcus</i> sp. BH72	4,376,040	0.143	1.0038292	0.99521	0.662	0.191	0.146	0.095	0.027	0.021
<i>Mycobacterium</i> sp. KMS	5,737,227	0.15	1.000606	0.9988113	0.671	0.195	0.134	0.1	0.029	0.02
<i>Psychromonas ingrahamii</i> 37	4,559,598	0.052	1.0005045	1.0033549	0.55	0.298	0.153	0.029	0.015	0.008
<i>Borrelia turicatae</i> 91E135	917,330	0.154	0.9840503	1.0020158	0.84	0.141	0.019	0.13	0.022	0.003
<i>Arthrobacter aurescens</i> TC1	4,597,686	0.071	1.0088656	0.999207	0.623	0.184	0.193	0.044	0.013	0.014
<i>Mycobacterium vanbaalenii</i> PYR-1	6,491,865	0.138	1.003291	0.9976668	0.677	0.188	0.135	0.093	0.026	0.019
<i>Marinobacter aquaeolei</i> VT8	4,326,849	0.049	1.0229098	0.9697434	0.314	0.277	0.409	0.015	0.014	0.02
<i>Shewanella</i> sp. W3-18-1	4,708,380	0.028	0.9996009	1.0031857	0.294	0.45	0.256	0.008	0.013	0.007
<i>Desulfotribium vulgare</i> DP4	3,462,887	0.083	0.9968801	0.9977971	0.594	0.179	0.227	0.049	0.015	0.019
<i>Acidovorax citrulli</i> AAC00-1	5,352,772	0.141	0.9977868	1.0042177	0.722	0.162	0.116	0.102	0.023	0.016
<i>Neisseria meningitidis</i> FAM18	2,194,961	0.057	1.0023763	1.0125089	0.014	0.734	0.253	0.001	0.042	0.014
<i>Mycobacterium bovis</i> BCG str. Pasteur 1173P2	4,374,522	0.108	0.999709	0.995051	0.663	0.159	0.178	0.072	0.017	0.019

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Polaromonas naphthalenivorans CJ2	4,410,291	0.105	1.0070116	0.9974162	0.436	0.353	0.211	0.046	0.037	0.022
Acidovorax sp. JS42	4,448,856	0.119	0.9987066	1.0025957	0.644	0.234	0.122	0.077	0.028	0.015
Bartonella bacilliformis KC583	1,445,021	0.068	0.9799886	1.0680613	0.597	0.31	0.093	0.041	0.021	0.006
Burkholderia mallei SAVP1	1,734,922	0.206	1.0106632	0.9760324	0.513	0.42	0.067	0.106	0.087	0.014
Verminephrobacter eiseniae EF01-2	5,566,749	0.122	1.0045777	0.9978213	0.562	0.283	0.155	0.068	0.035	0.019
Campylobacter jejuni subsp. jejuni 81–176	1,616,554	0.159	1.0073507	0.9907783	0.699	0.235	0.065	0.111	0.037	0.01
Halorhodospira halophila SL1	2,678,452	0.124	1.0673981	0.9140553	0.782	0.089	0.13	0.097	0.011	0.016
Yersinia enterocolitica subsp. enterocolitica 8081	4,615,899	0.032	0.9997068	0.9975081	0.067	0.518	0.415	0.002	0.017	0.013
Prochlorococcus marinus str. AS9601	1,669,886	0.132	0.9961465	1.0053448	0.781	0.201	0.018	0.103	0.027	0.002
Prochlorococcus marinus str. MIT 9515	1,704,176	0.136	1.0008092	0.9919569	0.801	0.185	0.014	0.109	0.025	0.002
Hyperthermus butylicus DSM 5456	1,667,163	0.029	1.0106986	1.0020178	0.139	0.553	0.308	0.004	0.016	0.009
Prochlorococcus marinus str. NATL1A	1,864,731	0.092	1.0008496	1.0069602	0.717	0.256	0.026	0.066	0.024	0.002
Prochlorococcus marinus str. MIT 9303	2,682,675	0.033	0.9783919	1.0040439	0.001	0.765	0.234	0	0.025	0.008
Methylilium petroleiphilum PM1	4,044,195	0.169	0.9985927	0.9965783	0.644	0.191	0.165	0.109	0.032	0.028
Burkholderia mallei NCTC 10,229	2,284,095	0.207	1.0033747	1.0021005	0.513	0.421	0.066	0.106	0.087	0.014
Methanocorpusculum labreanum Z	1,804,962	0.041	0.9980648	0.9919623	0	0.568	0.432	0	0.023	0.018
Lactococcus lactis subsp. cremoris MG1363	2,529,478	0.087	0.9929224	0.9895916	0.681	0.266	0.053	0.059	0.023	0.005
Streptococcus sanguinis SK36	2,388,435	0.046	1.0000311	0.9987851	0.277	0.489	0.235	0.013	0.022	0.011
Clostridium thermocellum ATCC 27,405	3,843,301	0.071	0.9967891	0.9945397	0.495	0.247	0.258	0.035	0.018	0.018
Staphylothermus marinus F1	1,570,485	0.078	1.0424782	1.0560511	0.771	0.126	0.103	0.06	0.01	0.008
Rhodobacter sphaeroides ATCC 17,029	3,147,721	0.16	1.0076931	0.991095	0.676	0.219	0.105	0.108	0.035	0.017
Methanoculleus marisnigri JR1	2,478,101	0.08	0.9953352	1.0000064	0.528	0.312	0.159	0.042	0.025	0.013
Shewanella baltica OS155	5,127,376	0.028	1.0006413	0.9973175	0.143	0.524	0.334	0.004	0.015	0.009
Actinobacillus pleuropneumoniae L20	2,274,482	0.066	0.9963457	0.997856	0.335	0.482	0.183	0.022	0.032	0.012
Pyrobaculum calidifontis JCM 11,548	2,009,313	0.042	1.0058989	1.0001655	0.349	0.142	0.509	0.015	0.006	0.022
Burkholderia pseudomallei 668	3,912,947	0.192	0.9937984	0.9948668	0.499	0.431	0.07	0.096	0.083	0.013
Burkholderia pseudomallei 1106a	3,988,455	0.191	0.9902879	0.9930697	0.499	0.431	0.07	0.095	0.082	0.013
Mycobacterium sp. JLS	6,048,425	0.148	1.001487	0.9969798	0.672	0.194	0.134	0.1	0.029	0.02
Burkholderia mallei NCTC 10,247	2,352,693	0.207	1.0097862	0.9949476	0.513	0.421	0.066	0.106	0.087	0.014
Acinetobacter baumannii ATCC 17,978	3,976,747	0.058	1.0022262	1.0091571	0.612	0.281	0.107	0.036	0.016	0.006
Clostridium difficile 630	4,290,252	0.151	1.005555	1.0173331	0.866	0.095	0.039	0.131	0.014	0.006
Prochlorococcus marinus str. MIT 9301	1,641,879	0.132	0.9916927	0.9993512	0.779	0.203	0.018	0.103	0.027	0.002
Shewanella loihica PV-4	4,602,594	0.031	0.9990646	0.9981217	0.127	0.316	0.557	0.004	0.01	0.017
Methanococcus maripaludis C5	1,780,761	0.116	0.9832225	0.9909816	0.734	0.194	0.072	0.085	0.022	0.008
Herminiimonas arsenicoxydans	3,424,307	0.06	1.0009075	0.997764	0.089	0.62	0.292	0.005	0.037	0.017
Saccharopolyspora erythraea NRRL 2338	8,212,805	0.177	0.9994498	0.996271	0.752	0.127	0.122	0.133	0.022	0.022
Desulfotomaculum reducens MI-1	3,608,104	0.046	1.0064223	1.0157641	0.378	0.485	0.137	0.017	0.022	0.006
Burkholderia vietnamiensis G4	1,241,007	0.156	1.0036136	1.0113573	0.51	0.402	0.088	0.08	0.063	0.014
Francisella tularensis subsp. tularensis WY96-3418	1,898,476	0.115	0.998976	0.9958881	0.808	0.112	0.08	0.093	0.013	0.009
Geobacillus thermodenitrificans NG80-2	3,550,319	0.044	1.0042392	0.9912819	0.007	0.782	0.211	0	0.035	0.009
Streptococcus pyogenes str. Manfredo	1,841,271	0.059	0.9983661	1.0055408	0.638	0.251	0.111	0.038	0.015	0.007
Chlorobium phaeovibrioides DSM 265	1,966,858	0.042	1.0291183	0.9953557	0.063	0.544	0.393	0.003	0.023	0.017
Mycobacterium gilvum PYR-GCK	5,619,607	0.142	0.9967127	1.000139	0.667	0.213	0.12	0.095	0.03	0.017
Corynebacterium glutamicum R	3,314,179	0.03	0.9985408	1.004122	0.164	0.575	0.26	0.005	0.017	0.008
Aeromonas salmonicida subsp. salmonicida A449	4,702,402	0.053	0.9957006	1.0014716	0.394	0.348	0.257	0.021	0.019	0.014
Pyrobaculum arsenaticum DSM 13,514	2,121,076	0.025	1.002758	0.993705	0.29	0.21	0.499	0.007	0.005	0.013
Polynucleobacter necessarius subsp. asymbioticus QLW-P1DMWA-1	2,159,490	0.035	1.0051774	0.9909313	0.221	0.571	0.208	0.008	0.02	0.007
Salinispora tropica CNB-440	5,183,331	0.147	1.0003426	1.0041424	0.761	0.09	0.148	0.112	0.013	0.022
Yersinia pestis Pestoides F	4,517,345	0.028	1.0200336	0.9481509	0.068	0.486	0.446	0.002	0.014	0.012
Rhodobacter sphaeroides ATCC 17,025	3,217,726	0.15	0.9890522	1.0130305	0.673	0.209	0.118	0.101	0.031	0.018
Pseudomonas stutzeri A1501	4,567,418	0.104	0.9990913	1.0007335	0.539	0.221	0.24	0.056	0.023	0.025
Enterobacter sp. 638	4,518,712	0.042	1.0033955	0.9967547	0.061	0.518	0.421	0.003	0.022	0.018
Caldicellulosiruptor saccharolyticus DSM 8903	2,970,275	0.103	1.0486006	1.1271162	0.632	0.267	0.101	0.065	0.027	0.01
Shewanella putrefaciens CN-32	4,659,220	0.029	0.9978147	0.9947443	0.31	0.437	0.253	0.009	0.013	0.007
Pseudomonas mendocina ymp	5,072,807	0.113	0.9966583	0.9979685	0.558	0.199	0.243	0.063	0.022	0.027
Metallosphaera sedula DSM 5348	2,191,517	0.034	1.0134474	1.0288719	0.123	0.624	0.253	0.004	0.021	0.009
Flavobacterium johnsoniae UW101	6,096,872	0.108	1.0041295	0.9712018	0.688	0.21	0.103	0.074	0.023	0.011
Streptococcus suis 05ZYH33	2,096,309	0.045	1.0040274	0.9617431	0.514	0.319	0.167	0.023	0.014	0.007
Streptococcus suis 98HAH33	2,095,698	0.045	1.0036857	0.9623569	0.514	0.318	0.167	0.023	0.014	0.007
Bradyrhizobium sp. ORS278	7,456,587	0.132	1.0002894	1.0004272	0.534	0.336	0.13	0.071	0.044	0.017
Dichelobacter nodosus VCS1703A	1,389,350	0.096	0.9892905	0.9748854	0.096	0.797	0.108	0.009	0.076	0.01
Pelotomaculum thermopropionicum SI	3,025,375	0.053	1.01283	1.0073592	0.048	0.549	0.403	0.003	0.029	0.021
Dehalococcoides sp. BAV1	1,341,892	0.046	0.9956617	0.9847585	0.051	0.383	0.567	0.002	0.018	0.026
Vibrio cholerae O395	1,108,250	0.032	0.9973906	0.9713591	0.09	0.6	0.31	0.003	0.019	0.01
uncultured methanogenic archaeon RC-1	3,179,916	0.027	0.9886983	0.998292	0.226	0.203	0.57	0.006	0.006	0.015
Candidatus Vesicomysocius okutanii HA	1,022,154	0.117	1.010209	0.997804	0.854	0.089	0.058	0.1	0.01	0.007
Clavibacter michiganensis subsp. michiganensis NCPPB 382	3,297,891	0.215	0.9879244	1.0021097	0.715	0.211	0.074	0.154	0.045	0.016
Synechococcus sp. WH 7803	2,366,980	0.079	0.9906116	0.9989301	0.387	0.42	0.193	0.03	0.033	0.015
Synechococcus sp. RCC307	2,224,914	0.095	1.0042104	0.9980858	0.358	0.419	0.223	0.034	0.04	0.021
Geobacter uraniireducens RF4	5,136,364	0.031	0.9966574	0.9897913	0.168	0.493	0.339	0.005	0.015	0.01
Acidiphilium cryptum JF-5	3,389,227	0.154	0.9915856	0.9844203	0.62	0.27	0.111	0.096	0.042	0.017
Bradyrhizobium sp. BTA11	8,264,687	0.126	0.9980003	0.996753	0.517	0.347	0.136	0.065	0.044	0.017
Thermotoga petrophila RKU-1	1,823,511	0.07	1.0002546	0.9668288	0.064	0.585	0.352	0.004	0.041	0.025

(continued on next page)

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Staphylococcus aureus subsp. aureus JH9	2,906,700	0.097	1.0031719	1.0186874	0.882	0.087	0.031	0.086	0.008	0.003
Orientia tsutsugamushi str. Boryong	2,127,051	0.13	1.0001707	0.991478	0.861	0.109	0.03	0.112	0.014	0.004
Legionella pneumophila str. Corby	3,576,470	0.063	1.0125095	0.9509946	0.619	0.324	0.057	0.039	0.02	0.004
Clostridium botulinum A str. ATCC 3502	3,886,916	0.17	0.9778612	0.9746871	0.833	0.122	0.045	0.141	0.021	0.008
Mycoplasma agalactiae PG2	877,438	0.152	0.9740023	0.9457392	0.806	0.178	0.016	0.123	0.027	0.002
Brucella ovis ATCC 25,840	1,164,220	0.071	0.995164	0.9946858	0.213	0.566	0.221	0.015	0.04	0.016
Sphingomonas wittichii RW1	5,382,261	0.17	0.9950267	1.0001971	0.588	0.266	0.146	0.1	0.045	0.025
Spingomonas putida F1	5,959,964	0.086	1.002277	0.9962809	0.478	0.224	0.298	0.041	0.019	0.026
Lactobacillus reuteri DSM 20,016	1,999,618	0.055	1.0024212	1.0209348	0.66	0.204	0.136	0.036	0.011	0.007
Methanobrevibacter smithii ATCC 35,061	1,853,160	0.135	1.0418485	0.9273902	0.79	0.14	0.069	0.107	0.019	0.009
Roseiflexus sp. RS-1	5,801,598	0.076	0.9995674	0.9984343	0.414	0.388	0.198	0.031	0.029	0.015
Psychrobacter sp. PRwf-1	2,978,976	0.044	0.9918055	0.9932427	0.171	0.387	0.442	0.008	0.017	0.02
Mycobacterium tuberculosis H37Ra	4,419,977	0.108	0.9998773	0.9960947	0.663	0.159	0.178	0.072	0.017	0.019
Mycobacterium tuberculosis F11	4,424,435	0.108	0.9998854	0.9960358	0.663	0.159	0.178	0.072	0.017	0.019
Haemophilus influenzae PittEE	1,813,033	0.072	0.9989163	1.0063825	0.577	0.341	0.082	0.042	0.025	0.006
Haemophilus influenzae PittGG	1,887,192	0.072	1.0001113	0.9799165	0.579	0.34	0.081	0.042	0.025	0.006
Flavobacterium psychrophilum JIP02/86	2,861,988	0.116	0.9697072	0.9770989	0.775	0.171	0.055	0.09	0.02	0.006
Bacteroides vulgatus ATCC 8482	5,163,189	0.042	1.0034132	0.9910571	0.418	0.235	0.347	0.018	0.01	0.015
Parabacteroides distasonis ATCC 8503	4,811,379	0.031	1.0128228	1.0075294	0.231	0.377	0.392	0.007	0.012	0.012
Thermosiphon melanesiensis BI429	1,915,238	0.134	0.9840054	0.9781396	0.761	0.164	0.075	0.102	0.022	0.01
Clostridium beijerinckii NCIMB 8052	6,000,632	0.139	1.0162871	1.0204792	0.866	0.101	0.033	0.121	0.014	0.005
Staphylococcus aureus subsp. aureus JH1	2,906,507	0.097	1.003163	1.018675	0.882	0.087	0.031	0.086	0.008	0.003
Alkaliphilus metalliredigens QYMF	4,929,566	0.069	1.0016928	1.0119376	0.73	0.241	0.029	0.051	0.017	0.002
Methanococcus vannielii SB	1,720,048	0.132	1.0073038	1.0080564	0.783	0.164	0.052	0.103	0.022	0.007
Methanococcus aeolicus Nankai-3	1,569,500	0.152	0.9920395	1.0012906	0.777	0.172	0.051	0.118	0.026	0.008
Sinorhizobium medicae WSM419	3,781,904	0.089	0.99513	0.9958828	0.433	0.405	0.162	0.039	0.036	0.014
Methanococcus maripaludis C7	1,772,694	0.113	1.0086433	1.001788	0.728	0.195	0.077	0.082	0.022	0.009
Staphylococcus aureus subsp. aureus str. Newman	2,878,897	0.098	0.9955897	1.0078781	0.883	0.086	0.031	0.086	0.008	0.003
Klebsiella pneumoniae subsp. pneumoniae MGH 78,578	5,315,120	0.059	1.0020599	1.0015012	0.274	0.309	0.417	0.016	0.018	0.025
Marinomonas sp. MWYL1	5,100,344	0.036	1.0024714	0.9882759	0.441	0.38	0.179	0.016	0.014	0.006
Actinobacillus succinogenes 130Z	2,319,663	0.06	1.0046961	1.0058445	0.125	0.689	0.185	0.007	0.041	0.011
Pseudomonas aeruginosa PA7	6,588,339	0.125	1.0025666	0.9984034	0.637	0.132	0.231	0.08	0.017	0.029
Janthinobacterium sp. Marseille	4,110,251	0.053	0.9954811	0.9993549	0.097	0.523	0.38	0.005	0.028	0.02
Nitratiruptor sp. SB155-2	1,877,931	0.075	0.9979186	0.9952643	0.413	0.289	0.297	0.031	0.022	0.022
Sulfurovum sp. NBC37-1	2,562,277	0.04	0.993184	1.0002476	0.273	0.284	0.443	0.011	0.011	0.018
Kineococcus radiotolerans SRS30216	4,761,183	0.22	1.0051649	0.9922918	0.819	0.093	0.088	0.18	0.021	0.019
Shewanella baltica OS185	5,229,686	0.028	1.0039068	1.0008161	0.14	0.522	0.339	0.004	0.015	0.01
Ochrobactrum anthropi ATCC 49,188	2,887,297	0.061	0.9778218	1.040936	0.177	0.564	0.259	0.011	0.034	0.016
Bacillus cytotoxicus NVH 391–98	4,087,024	0.076	0.9959656	0.9902719	0.765	0.141	0.094	0.058	0.011	0.007
Anaeromyxobacter sp. Fw109-5	5,277,990	0.226	0.9985518	0.9956803	0.734	0.184	0.082	0.166	0.042	0.019
Clostridium botulinum A str. ATCC 19,397	3,863,450	0.17	0.9689181	0.957711	0.834	0.121	0.045	0.142	0.021	0.008
Clostridium botulinum A str. Hall	3,760,560	0.171	0.975256	0.9690441	0.834	0.121	0.045	0.142	0.021	0.008
Clostridium botulinum F str. Langeland	3,995,387	0.168	0.9705653	0.961136	0.836	0.12	0.045	0.141	0.02	0.008
Clostridium kluyveri DSM 555	3,964,618	0.124	0.9852684	0.9655366	0.769	0.148	0.083	0.095	0.018	0.01
Campylobacter jejuni subsp. doylei 269.97	1,845,106	0.158	0.996372	0.9662393	0.706	0.233	0.061	0.112	0.037	0.01
Yersinia pseudotuberculosis IP 31,758	4,723,306	0.028	0.9970727	0.9910349	0.063	0.487	0.45	0.002	0.014	0.013
Candidatus Methanoregula boonei 6A8	2,542,943	0.044	1.0062282	1.0021695	0.133	0.366	0.502	0.006	0.016	0.022
Campylobacter hominis ATCC BAA-381	1,711,273	0.153	1.0007931	1.0063732	0.644	0.295	0.061	0.099	0.045	0.009
Campylobacter curvus 525.92	1,971,264	0.074	0.9851751	0.9943732	0.116	0.54	0.344	0.009	0.04	0.026
Fervidobacterium nodosum Rt17-B1	1,948,941	0.088	0.9731121	0.9774139	0.754	0.137	0.109	0.066	0.012	0.01
Parvibaculum lavamentivorans DS-1	3,914,745	0.098	0.9680823	1.0234557	0.455	0.411	0.134	0.044	0.04	0.013
Xanthobacter autotrophicus Py2	5,308,934	0.129	0.9853451	1.0249016	0.696	0.174	0.129	0.09	0.022	0.017
Bacillus amyloliquefaciens FZB42	3,918,589	0.059	0.9973289	1.0018728	0.06	0.499	0.441	0.004	0.03	0.026
Coxiella burnetii Dugway 5 J108-111	2,158,758	0.048	0.9882273	0.9885907	0.345	0.612	0.044	0.017	0.029	0.002
Francisella tularensis subsp. holarctica FTNF002-00	1,890,909	0.116	0.992199	1.0245217	0.809	0.113	0.078	0.094	0.013	0.009
Roseiflexus castenholzii DSM 13,941	5,723,298	0.077	0.997972	1.000501	0.433	0.426	0.14	0.033	0.033	0.011
Ignicoccus hospitalis KIN4/I	1,297,538	0.053	0.9960267	0.9997244	0.231	0.373	0.397	0.012	0.02	0.021
Cronobacter sakazakii ATCC BAA-894	4,368,373	0.065	0.9929034	1.0008519	0.203	0.421	0.376	0.013	0.028	0.025
Staphylococcus aureus subsp. aureus Mu3	2,880,168	0.098	0.9942605	1.0061054	0.883	0.086	0.031	0.086	0.008	0.003
Vibrio Harveyi ATCC BAA-1116	3,765,351	0.028	0.9983108	0.9977749	0.204	0.447	0.349	0.006	0.013	0.01
Streptococcus gordonii str. Challis substr. CH1	2,196,662	0.052	1.0062433	1.0106317	0.499	0.361	0.141	0.026	0.019	0.007
Citrobacter koseri ATCC BAA-895	4,720,462	0.047	1.0069717	0.9677813	0.093	0.442	0.465	0.004	0.021	0.022
Escherichia coli HS	4,643,538	0.037	1.0030355	0.9958712	0.005	0.496	0.499	0	0.018	0.018
Escherichia coli E24377A	4,979,619	0.036	1.0041236	0.9957862	0.003	0.489	0.508	0	0.018	0.018
Campylobacter concisus 13,826	2,052,007	0.093	1.0007564	0.9880828	0.349	0.388	0.263	0.032	0.036	0.025
Thermotoga lettingae TMO	2,135,342	0.073	1.0121772	1.0029716	0.511	0.316	0.173	0.037	0.023	0.013
Shewanella sediminis HAW-EB3	5,517,674	0.02	0.9956019	0.9975853	0.228	0.276	0.497	0.004	0.005	0.01
Serratia proteamaculans 568	5,448,853	0.049	1.0047957	0.9977572	0.151	0.381	0.468	0.007	0.019	0.023
Campylobacter jejuni subsp. jejuni 81,116	1,628,115	0.16	1.0050248	0.987803	0.701	0.235	0.064	0.112	0.038	0.01
Prochlorococcus marinus str. MIT 9215	1,738,790	0.134	0.9904089	0.9961116	0.784	0.199	0.017	0.105	0.027	0.002
Bacillus pumilus SAFR-032	3,704,465	0.051	0.9970052	1.0000378	0.429	0.427	0.144	0.022	0.022	0.007
Arcobacter butzleri RM4018	2,341,251	0.195	1.0061073	0.9944404	0.811	0.137	0.052	0.158	0.027	0.01
Rickettsia canadensis str. McKiel	1,159,772	0.121	0.994306	1.0005611	0.877	0.073	0.051	0.106	0.009	0.006
Rickettsia akari str. Hartford	1,231,060	0.108	0.9988385	1.0172297	0.854	0.085	0.06	0.092	0.009	0.007

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Rickettsia rickettsii str. 'Sheila Smith'	1,257,710	0.108	0.996338	1.0154282	0.836	0.099	0.065	0.091	0.011	0.007
Rickettsia bellii OSU 85–389	1,528,980	0.117	1.0042162	1.0106476	0.851	0.112	0.038	0.1	0.013	0.004
Rickettsia massiliae MTU5	1,360,898	0.108	1.0068118	1.0230035	0.832	0.101	0.067	0.09	0.011	0.007
Shewanella pealeana ATCC 700,345	5,174,581	0.028	0.9983908	1.0084517	0.297	0.356	0.346	0.008	0.01	0.01
Frankia sp. EAN1pec	8,982,042	0.169	0.9964772	1.0033838	0.788	0.112	0.1	0.133	0.019	0.017
Alkaliphilus oremlandii OhILAs	3,123,558	0.076	1.0123466	1.0268015	0.731	0.182	0.086	0.055	0.014	0.007
Acaryochloris marina MBIC11017	6,503,724	0.023	1.0030513	0.9935339	0.094	0.664	0.242	0.002	0.015	0.006
Azorhizobium caulinodans ORS 571	5,369,772	0.131	0.9993711	0.99576	0.673	0.206	0.121	0.088	0.027	0.016
Desulfococcus oleovorans Hxd3	3,944,167	0.065	1.0040074	1.0031576	0.17	0.414	0.417	0.011	0.027	0.027
Dinoroseobacter shibae DFL 12	3,789,584	0.112	1.0043219	1.0019405	0.674	0.202	0.124	0.075	0.023	0.014
Salinispora arenicola CNS-205	5,786,361	0.151	0.9949311	1.0080865	0.748	0.103	0.149	0.113	0.016	0.022
Caldvirga maquilingsensis IC-167	2,077,567	0.075	1.0029172	0.9979314	0.185	0.339	0.477	0.014	0.025	0.036
Herpetosiphon aurantiacus ATCC 23,779	6,346,587	0.066	0.9939334	1.0051529	0.002	0.564	0.434	0	0.037	0.029
Methanococcus maripaludis C6	1,744,193	0.112	0.988983	0.9955466	0.724	0.199	0.077	0.081	0.022	0.009
Prochlorococcus marinus str. MIT 9211	1,688,963	0.07	1.0104426	1.0026926	0.596	0.358	0.046	0.042	0.025	0.003
Shewanella baltica OS195	5,347,283	0.028	1.0004674	1.003335	0.143	0.52	0.337	0.004	0.015	0.009
Clostridium phytofermentans ISDg	4,847,594	0.076	1.0037011	1.0228723	0.823	0.081	0.097	0.063	0.006	0.007
Delftia acidovorans SPH-1	6,767,514	0.13	0.9946284	1.0082889	0.615	0.246	0.14	0.08	0.032	0.018
Petrotoxa mobilis SJ95	2,169,548	0.1	1.0085881	0.9916948	0.742	0.191	0.067	0.074	0.019	0.007
Salmonella enterica subsp. arizonae serovar 62:z4,z23:-	4,600,800	0.038	0.9955315	0.997761	0.014	0.501	0.485	0.001	0.019	0.018
Staphylococcus aureus subsp. aureus USA300_TCH1516	2,872,915	0.099	0.9884646	0.9990571	0.883	0.086	0.031	0.088	0.009	0.003
Lactobacillus helveticus DPC 4571	2,080,931	0.073	1.0052465	1.036367	0.667	0.212	0.121	0.049	0.016	0.009
Nitrosopumilus maritimus SCM1	1,645,259	0.12	0.9809081	0.9845373	0.613	0.307	0.08	0.074	0.037	0.01
Burkholderia multivorans ATCC 17,616	2,472,928	0.185	1.0036509	1.0047736	0.467	0.43	0.102	0.086	0.08	0.019
Salmonella enterica subsp. enterica serovar Paratyphi B str. SPB7	4,858,887	0.041	1.001124	0.9978473	0.031	0.495	0.474	0.001	0.02	0.019
Brucella canis ATCC 23,365	2,105,969	0.072	1.0010016	0.9932233	0.208	0.565	0.227	0.015	0.041	0.016
Coxiella burnetii RSA 331	2,016,427	0.048	1.002304	0.9977642	0.318	0.632	0.049	0.015	0.03	0.002
Candidatus Sulcia muelleri GWSS	245,530	0.26	1.0013558	1.0345996	0.892	0.065	0.042	0.232	0.017	0.011
Neisseria meningitidis 053,442	2,153,416	0.057	1.0004788	1.0083215	0.015	0.736	0.249	0.001	0.042	0.014
Gluconacetobacter diazotrophicus PAI 5	3,944,163	0.131	1.0078919	0.9927649	0.608	0.199	0.194	0.079	0.026	0.025
Yersinia pestis Angola	4,504,254	0.027	1.0116282	0.9791281	0.063	0.489	0.449	0.002	0.013	0.012
Bartonella tribocorum CIP 105,476	2,619,061	0.07	1.0357095	0.9222498	0.524	0.34	0.135	0.037	0.024	0.009
Sorangium cellulosum 'So ce 56'	13,033,779	0.189	1.0060987	1.0009464	0.721	0.217	0.063	0.136	0.041	0.012
Acholeplasma laidlawii PG-8A	1,496,992	0.112	1.0021889	0.9975843	0.862	0.056	0.083	0.096	0.006	0.009
Renibacterium salmoninarum ATCC 33,209	3,155,250	0.045	1.0033641	0.9849154	0.255	0.405	0.34	0.011	0.018	0.015
Brucella suis ATCC 23,445	1,923,763	0.071	0.9985488	1.0018948	0.206	0.568	0.226	0.015	0.04	0.016
Bordetella petrii DSM 12,804	5,287,950	0.118	1.0033727	0.994269	0.595	0.221	0.184	0.07	0.026	0.022
Methylobacterium extorquens PA1	5,471,154	0.143	1.0083792	0.9966161	0.684	0.208	0.109	0.098	0.03	0.016
Chloroflexus aurantiacus J-10-fl	5,258,541	0.042	0.9938948	0.9985192	0.307	0.298	0.396	0.013	0.013	0.017
Bacillus weihenstephanensis KBAB4	5,262,775	0.075	0.993577	0.9870568	0.816	0.098	0.085	0.061	0.007	0.006
Rickettsia rickettsii str. Iowa	1,268,175	0.109	0.9972932	1.0149056	0.836	0.099	0.065	0.091	0.011	0.007
Actinobacillus pleuropneumoniae serovar 3 str. JL03	2,242,062	0.066	0.9948583	0.995308	0.339	0.478	0.183	0.022	0.031	0.012
Chlamydia trachomatis L2b/UCH-1/proctitis	1,038,869	0.05	1.0046933	1.0006429	0.437	0.393	0.17	0.022	0.02	0.008
Chlamydia trachomatis 434/Bu	1,038,842	0.05	1.0047365	1.0006617	0.437	0.393	0.17	0.022	0.02	0.008
Microcystis aeruginosa NIES-843	5,842,795	0.041	1.0016664	0.9985378	0.418	0.44	0.142	0.017	0.018	0.006
Thermoanaerobacter sp. X514	2,457,259	0.1	1.0441701	1.1347393	0.718	0.204	0.078	0.072	0.02	0.008
Thermoanaerobacter pseudethanolicus ATCC 33,223	2,362,816	0.099	1.009163	1.0205434	0.71	0.208	0.083	0.07	0.021	0.008
Pseudomonas putida GB-1	6,078,430	0.088	1.001471	0.9985913	0.471	0.224	0.304	0.042	0.02	0.027
Shewanella halifaxensis HAW-EB4	5,226,917	0.028	1.0051238	1.0013703	0.304	0.339	0.357	0.008	0.009	0.01
Francisella philomiragia subsp. philomiragia ATCC 25,017	2,045,775	0.111	1.0062687	1.0171056	0.807	0.121	0.072	0.09	0.013	0.008
Helicobacterium modesticaldum Icel1	3,075,407	0.047	1.016431	1.0114551	0.302	0.398	0.3	0.014	0.019	0.014
Caulobacter sp. K31	5,477,872	0.139	1.0041572	0.9966205	0.644	0.136	0.22	0.09	0.019	0.031
Halobacterium salinarum R1	2,000,962	0.148	1.0012539	0.9968468	0.649	0.28	0.071	0.096	0.041	0.01
Finexgoldia magna ATCC 29,328	1,797,577	0.115	0.9791886	0.9750321	0.805	0.144	0.051	0.092	0.017	0.006
Streptococcus pneumoniae Hungary19A-6	2,245,615	0.055	1.014996	1.0171056	0.574	0.304	0.122	0.031	0.017	0.007
Lysinibacillus sphaericus C3-41	4,639,821	0.061	1.0136206	1.0513394	0.78	0.16	0.06	0.047	0.01	0.004
Mycobacterium abscessus	5,067,172	0.089	1.0024776	0.9940497	0.654	0.195	0.151	0.058	0.017	0.014
Acinetobacter baumannii SDF	3,421,954	0.057	0.9995735	1.0092857	0.598	0.288	0.114	0.034	0.016	0.006
Clavibacter michiganensis subsp. sepedonicus	3,258,645	0.212	1.0003783	0.9928338	0.717	0.205	0.077	0.152	0.044	0.016
Acinetobacter baumannii AYE	3,936,291	0.056	1.0032441	0.9944759	0.59	0.296	0.114	0.033	0.016	0.006
Candidatus Desulfurudis audaxviator MP104C	2,349,476	0.061	1.0199358	0.9979025	0.563	0.197	0.24	0.034	0.012	0.015
Yersinia pseudotuberculosis YPIII	4,689,441	0.028	1.0020375	0.9975079	0.063	0.493	0.444	0.002	0.014	0.012
Escherichia coli ATCC 8739	4,746,218	0.037	0.9974729	1.0040462	0.006	0.495	0.499	0	0.018	0.019
Leuconostoc citreum KM20	1,796,284	0.062	0.9878671	1.0082324	0.572	0.298	0.13	0.035	0.018	0.008
Escherichia coli str. K-12 substr. DH10B	4,686,137	0.037	1.0007162	0.998254	0.005	0.5	0.495	0	0.018	0.018
Synechococcus sp. PCC 7002	3,008,047	0.033	0.9995014	0.9994336	0.001	0.798	0.201	0	0.027	0.007
Candidatus Korarchaeum cryptofilum OPF8	1,590,757	0.063	1.0138587	1.0311125	0.006	0.779	0.215	0	0.049	0.014
Thermotoga sp. RQ2	1,877,693	0.07	0.9877542	0.9716913	0.061	0.584	0.354	0.004	0.041	0.025
Escherichia coli SMS-3-5	5,068,389	0.036	1.0020452	0.9968712	0.002	0.496	0.502	0	0.018	0.018
Pseudomonas putida W619	5,774,330	0.081	0.9945059	1.00052	0.467	0.234	0.298	0.038	0.019	0.024
Ureaplasma parvum serovar 3 str. ATCC 27,815	751,679	0.207	1.0029362	1.0312448	0.876	0.087	0.037	0.181	0.018	0.008
Methylobacterium radiotolerans JCM 2831	6,077,833	0.184	0.9954333	1.0039736	0.748	0.151	0.101	0.138	0.028	0.019
Shewanella woodyi ATCC 51,908	5,935,403	0.03	0.9949141	0.9939899	0.382	0.295	0.323	0.011	0.009	0.01

(continued on next page)

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Burkholderia cenocepacia MC0-3	3,532,883	0.165	1.0037697	1.0029784	0.498	0.385	0.117	0.082	0.064	0.019
Methylobacterium sp. 4-46	7,659,055	0.178	1.0043237	0.999412	0.781	0.143	0.076	0.139	0.025	0.014
Xylella fastidiosa M12	2,475,130	0.024	0.9920141	1.0167536	0.045	0.712	0.243	0.001	0.017	0.006
Clostridium botulinum B1 str. Okra	3,958,233	0.168	0.9739237	0.9670667	0.836	0.119	0.045	0.14	0.02	0.007
Haemophilus somnus 2336	2,263,857	0.068	0.9878722	1.0607569	0.684	0.255	0.06	0.047	0.017	0.004
Clostridium botulinum A3 str. Loch Maree	3,992,906	0.168	0.9777193	0.9744535	0.837	0.119	0.044	0.14	0.02	0.007
Leptothrix cholodnii SP-6	4,909,403	0.176	1.0062147	0.9948579	0.601	0.199	0.2	0.106	0.035	0.035
Thermoproteus neutrophilus V24Sta	1,769,823	0.055	1.0141854	1.0128402	0.517	0.119	0.364	0.029	0.007	0.02
Cupriavidus taiwanensis	3,416,911	0.146	0.9888456	1.0014826	0.62	0.202	0.177	0.09	0.029	0.026
Polynucleobacter necessarius subsp. necessarius STIR1	1,560,469	0.036	1.0022813	1.0023519	0.157	0.624	0.219	0.006	0.023	0.008
Candidatus Phytoplasma australiense	879,959	0.199	0.973227	1.0059517	0.767	0.215	0.018	0.153	0.043	0.004
Corynebacterium urealyticum DSM 7109	2,369,219	0.079	1.0195067	0.9929844	0.744	0.129	0.127	0.059	0.01	0.01
Cyanotheca sp. ATCC 51,142	4,934,271	0.062	1.004137	1.001091	0.693	0.247	0.061	0.043	0.015	0.004
Burkholderia ambifaria MC40-6	3,443,583	0.168	1.0057364	1.004311	0.501	0.387	0.112	0.084	0.065	0.019
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'	4,063,606	0.052	1.0017754	0.9945715	0.692	0.202	0.107	0.036	0.01	0.006
Exiguobacterium sibiricum 255-15	3,034,136	0.051	1.0013904	1.0077435	0.03	0.682	0.289	0.002	0.035	0.015
Opiritatus terrae PB90-1	5,957,605	0.12	1.000142	1.013991	0.577	0.333	0.091	0.069	0.04	0.011
Streptomyces griseus subsp. griseus NBRC 13,350	8,545,929	0.179	1.0087423	0.9974521	0.824	0.089	0.087	0.148	0.016	0.016
Xylella fastidiosa M23	2,535,690	0.024	0.9950895	1.0210316	0.039	0.718	0.243	0.001	0.017	0.006
Beijerinckia indica subsp. indica ATCC 9039	4,170,153	0.066	0.9972697	0.9984722	0.22	0.609	0.171	0.014	0.04	0.011
Streptococcus pneumoniae CGSP14	2,209,198	0.056	1.0076575	1.0098326	0.581	0.299	0.12	0.032	0.017	0.007
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'	3,599,677	0.084	1.0032012	0.9923441	0.428	0.462	0.11	0.036	0.039	0.009
Lactobacillus reuteri JCM 1112	2,039,414	0.055	1.0018967	1.0194099	0.661	0.203	0.137	0.036	0.011	0.007
Lactobacillus fermentum IFO 3956	2,098,685	0.033	1.0242813	1.0075138	0.02	0.483	0.496	0.001	0.016	0.017
Acinetobacter baumannii ACICU	3,904,116	0.058	0.9951139	1.0003793	0.604	0.286	0.11	0.035	0.017	0.006
Mycobacterium marinum M	6,636,827	0.109	1.0074386	0.9937001	0.665	0.161	0.174	0.073	0.018	0.019
Kocuria rhizophila DC2201	2,697,540	0.164	0.9961201	1.0017457	0.816	0.099	0.085	0.133	0.016	0.014
Burkholderia phyatum STM815	3,479,187	0.12	0.998647	0.9808887	0.414	0.479	0.108	0.05	0.057	0.013
Nostoc punctiforme PCC 73,102	8,234,322	0.04	1.0004111	0.9989758	0.54	0.305	0.154	0.021	0.012	0.006
Yersinia pseudotuberculosis PB1/+	4,695,619	0.028	0.9957917	1.0098248	0.063	0.491	0.446	0.002	0.014	0.012
Elusimicrobium minutum Pei191	1,643,562	0.098	1.0105688	1.0135904	0.299	0.539	0.161	0.029	0.053	0.016
Bordetella avium 197 N	3,732,255	0.079	1.0039973	0.9961602	0.495	0.326	0.178	0.039	0.026	0.014
Akkermansia muciniphila ATCC BAA-835	2,664,102	0.071	1.0068351	1.007348	0.135	0.395	0.47	0.01	0.028	0.034
Shigella boydii CDC 3083-94	4,615,997	0.036	0.9951109	1.006507	0.014	0.458	0.527	0.001	0.017	0.019
Borrelia hermsii DAH	922,307	0.147	0.985276	0.9999343	0.824	0.156	0.021	0.121	0.023	0.003
Clostridium botulinum B str. Eklund 17B	3,800,327	0.171	1.0072608	1.0150288	0.885	0.086	0.029	0.151	0.015	0.005
Francisella tularensis subsp. mediasiatica FSC147	1,893,886	0.115	1.0085818	1.0221327	0.809	0.112	0.079	0.093	0.013	0.009
Burkholderia phytofirmans PsJN	4,467,537	0.108	0.9937379	1.002021	0.429	0.458	0.113	0.046	0.049	0.012
Ralstonia pickettii 12 J	3,942,557	0.101	0.9946264	0.9982942	0.521	0.328	0.151	0.053	0.033	0.015
Xanthomonas campestris pv. campestris str. B100	5,079,002	0.121	0.992451	1.0047899	0.55	0.282	0.168	0.066	0.034	0.02
Erwinia tasmaniensis Et1/99	3,883,467	0.045	0.9962438	1.0015981	0.09	0.379	0.531	0.004	0.017	0.024
Helicobacter pylori Shi470	1,608,548	0.104	0.9877006	0.9812307	0.345	0.438	0.217	0.036	0.045	0.023
Xanthomonas oryzae pv. oryzae PXO99A	5,240,075	0.108	1.0047481	0.9934611	0.505	0.317	0.178	0.054	0.034	0.019
Natronaerobius thermophilus JW/NM-WN-LF	3,165,557	0.073	1.0054901	1.0070679	0.749	0.213	0.039	0.055	0.016	0.003
Clostridium botulinum E3 str. Alaska E43	3,659,644	0.173	1.0085106	1.0189772	0.888	0.083	0.028	0.153	0.014	0.005
Methylobacterium populi BJ001	5,800,441	0.158	1.0031936	0.9960198	0.703	0.192	0.105	0.111	0.03	0.017
Porphyromonas gingivalis ATCC 33,277	2,354,886	0.027	1.0041464	0.9946195	0.029	0.394	0.577	0.001	0.01	0.015
Sulfurihydrogenibium sp. YO3AOP1	1,838,442	0.13	0.9930239	0.9790004	0.735	0.171	0.094	0.095	0.022	0.012
Treponema pallidum subsp. pallidum SS14	1,139,457	0.034	0.9942708	1.0144051	0.066	0.362	0.572	0.002	0.012	0.019
Bruceella abortus S19	2,122,487	0.072	1.0023805	0.9939257	0.207	0.567	0.226	0.015	0.041	0.016
Orientia tsutsugamushi str. Ikeda	2,008,987	0.131	0.9975774	1.0207072	0.857	0.113	0.03	0.113	0.015	0.004
Methylobacterium inferorum V4	2,287,145	0.06	0.997312	0.9853483	0.098	0.757	0.145	0.006	0.046	0.009
Burkholderia multivorans ATCC 17,616	919,805	0.151	0.9915479	1.0037855	0.485	0.424	0.091	0.073	0.064	0.014
Chlorobium limicola DSM 245	2,763,181	0.045	0.9962931	0.9856869	0.011	0.453	0.535	0.001	0.021	0.024
Geobacter lovleyi SZ	3,917,761	0.051	0.9997643	0.996247	0.129	0.338	0.534	0.007	0.017	0.027
Bifidobacterium longum DJO10A	2,375,792	0.065	0.9944766	1.0061068	0.457	0.304	0.239	0.03	0.02	0.016
Candidatus Amoebophilus asiaticus 5a2	1,884,364	0.091	1.0030406	1.0129918	0.72	0.223	0.057	0.066	0.02	0.005
Chlorobium phaeobacteroides BS1	2,736,403	0.036	0.9930367	1.0101826	0.01	0.47	0.521	0	0.017	0.019
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Ames)'	3,603,977	0.084	1.0030888	0.9922807	0.428	0.461	0.11	0.036	0.039	0.009
Actinobacillus pleuropneumoniae serovar 7 str. AP76	2,331,981	0.065	0.999501	1.0015391	0.346	0.474	0.18	0.022	0.031	0.012
Stenotrophomonas maltophilia K279a	4,851,126	0.137	0.9961572	1.0025338	0.572	0.186	0.241	0.078	0.026	0.033
Wolbachia endosymbiont of Culex quinquefasciatus Pel	1,482,455	0.093	0.9900445	0.9875187	0.787	0.159	0.054	0.073	0.015	0.005
Rhizobium etli CIAT 652	4,513,324	0.102	1.0030311	0.9756391	0.388	0.409	0.203	0.04	0.042	0.021
Cellvibrio japonicus Ueda107	4,576,573	0.038	1.0055938	0.9916307	0.03	0.558	0.412	0.001	0.021	0.016
Lactobacillus casei BL23	3,079,196	0.041	1.0023036	1.0047532	0.095	0.581	0.324	0.004	0.024	0.013
Burkholderia cenocepacia J2315	3,870,082	0.164	0.9953438	0.9973261	0.5	0.386	0.114	0.082	0.063	0.019
Rhodopseudomonas palustris TIE-1	5,744,041	0.128	0.9977618	1.003588	0.506	0.321	0.173	0.065	0.041	0.022
Mycoplasma arthritidis 158L3-1	820,453	0.145	1.0413956	1.0138843	0.761	0.192	0.048	0.11	0.028	0.007
Chloroherpeton thalassium ATCC 35,110	3,293,456	0.069	0.9971504	1.0001332	0.102	0.805	0.093	0.007	0.056	0.006
Chlorobaculum parvum NCIB 8327	2,289,249	0.053	1.0045938	1.0022166	0.182	0.508	0.31	0.01	0.027	0.017
Neisseria gonorrhoeae NCCP11945	2,232,025	0.063	0.9753745	1.0201859	0.027	0.721	0.252	0.002	0.046	0.016
Candidatus Phytoplasma mali	601,943	0.28	1.0249283	1.0784384	0.899	0.088	0.013	0.251	0.025	0.004
Prosthecochloris aestuarii DSM 271	2,512,923	0.037	1.0072653	0.9886201	0.001	0.529	0.47	0	0.02	0.017
Pelodictyon phaeoclathratiforme BU-1	3,018,238	0.038	1.0134619	0.986097	0.029	0.567	0.404	0.001	0.021	0.015

Table 1 (continued)

Organism	Length	Id (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Stenotrophomonas maltophilia R551-3	4,573,969	0.14	0.9955854	0.9996248	0.559	0.19	0.251	0.078	0.027	0.035
Streptococcus pneumoniae G54	2,078,953	0.055	1.0055701	0.9983897	0.567	0.307	0.126	0.031	0.017	0.007
Salmonella enterica subsp. enterica serovar Newport str. SL254	4,827,641	0.042	1.0011601	1.0007475	0.034	0.495	0.471	0.001	0.021	0.02
Salmonella enterica subsp. enterica serovar Heidelberg str. SL476	4,888,768	0.041	1.0029815	1.000227	0.031	0.499	0.47	0.001	0.02	0.019
Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633	4,709,075	0.041	1.0048892	1.0008134	0.033	0.495	0.472	0.001	0.02	0.019
Hydrogenobaculum sp. Y04AAS1	1,559,514	0.119	0.9826337	1.0121368	0.565	0.146	0.289	0.067	0.017	0.034
Streptococcus equi subsp. zooepidemicus MGCS10565	2,024,171	0.058	1.0005879	1.0109246	0.333	0.423	0.245	0.019	0.025	0.014
Alteromonas macleodii 'Deep ecotype'	4,412,282	0.029	1.0076141	1.0014614	0.265	0.525	0.21	0.008	0.015	0.006
Phenylobacterium zincinum HLK1	3,996,255	0.182	1.0082218	1.0046932	0.747	0.101	0.152	0.136	0.018	0.028
Anaeromyxobacter sp. K	5,061,632	0.239	0.991122	0.9982509	0.778	0.124	0.099	0.186	0.03	0.024
Geobacter bemidjiensis Bem	4,615,150	0.053	1.0080015	1.004982	0.577	0.226	0.197	0.031	0.012	0.01
Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601	4,581,797	0.041	0.9988751	0.9927791	0.033	0.499	0.468	0.001	0.021	0.019
Salmonella enterica subsp. enterica serovar Agona str. SL483	4,798,660	0.042	1.0041358	0.9960745	0.03	0.5	0.47	0.001	0.021	0.02
Vibrio fischeri MJ11	1,418,848	0.065	1.0160726	0.9854318	0.736	0.156	0.109	0.047	0.01	0.007
Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853	4,842,908	0.041	1.0039761	0.9993961	0.033	0.496	0.471	0.001	0.021	0.019
Acidithiobacillus ferrooxidans ATCC 53,993	2,885,038	0.051	0.9960704	1.010919	0.446	0.313	0.241	0.023	0.016	0.012
Borrelia duttonii Ly	931,674	0.174	0.9871404	0.9888725	0.862	0.121	0.017	0.15	0.021	0.003
Borrelia recurrentis A1	930,981	0.175	0.990496	0.9805872	0.863	0.12	0.017	0.151	0.021	0.003
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91	4,658,697	0.042	1.0002113	1.0057813	0.034	0.499	0.467	0.001	0.021	0.019
Klebsiella pneumoniae 342	5,641,239	0.058	0.9974308	0.9962146	0.266	0.306	0.428	0.015	0.018	0.025
Salmonella enterica subsp. enterica serovar Enteritidis str. P125109	4,685,848	0.042	1.0021284	0.9971544	0.032	0.502	0.466	0.001	0.021	0.019
Coprothermobacter proteolyticus DSM 5265	1,424,912	0.032	1.0108619	1.0302243	0.247	0.486	0.268	0.008	0.016	0.009
Thermodesulfobacterium yellowstonii DSM 11,347	2,003,803	0.121	0.9984209	0.9873702	0.609	0.305	0.086	0.074	0.037	0.01
Dictyoglomus thermophilum H-6-12	1,959,987	0.126	1.0327612	1.0567491	0.618	0.297	0.085	0.078	0.038	0.011
Aliivibrio salmonicida LFI1238	1,206,461	0.056	0.9880845	1.0119101	0.722	0.174	0.104	0.04	0.01	0.006
Helicobacter pylori G27	1,652,982	0.103	0.9872312	0.9858286	0.35	0.434	0.216	0.036	0.044	0.022
Escherichia coli O157:H7 str. EC4115	5,572,075	0.036	1.0002349	0.9960559	0.002	0.477	0.521	0	0.017	0.019
Gluconacetobacter diazotrophicus PAI 5	3,887,492	0.13	1.0067894	1.0056973	0.606	0.199	0.196	0.079	0.026	0.025
Rhizobium leguminosarum bv. trifolii WSM2304	4,537,948	0.102	1.0086688	0.9740607	0.381	0.406	0.213	0.039	0.041	0.022
Ureaplasma urealyticum serovar 10 str. ATCC 33,699	874,478	0.207	0.9843051	1.0279276	0.854	0.091	0.055	0.177	0.019	0.011
Streptococcus pyogenes NZ131	1,815,785	0.06	0.9962865	1.011405	0.638	0.251	0.111	0.038	0.015	0.007
Oligotropha carboxidovorans OM5	3,745,629	0.101	1.0056253	0.9993627	0.445	0.434	0.121	0.045	0.044	0.012
Escherichia coli SE11	4,887,515	0.037	0.9995956	0.993564	0.005	0.493	0.502	0	0.018	0.019
Rhodospirillum centenum SW	4,355,548	0.164	0.9935386	0.9982304	0.758	0.094	0.147	0.124	0.015	0.024
Helicobacter pylori P12	1,673,813	0.103	0.9872392	0.9903179	0.356	0.43	0.215	0.036	0.044	0.022
Coxiella burnetii CbuC_Q212	2,008,870	0.048	1.0040135	0.9715604	0.329	0.625	0.047	0.016	0.03	0.002
Coxiella burnetii CbuK_Q154	2,063,100	0.048	0.9998308	0.978369	0.325	0.627	0.048	0.016	0.03	0.002
Thermococcus onnurineus NA1	1,847,607	0.028	1.0173503	1.0140145	0.018	0.571	0.411	0.001	0.016	0.012
Candidatus Azobacteroides pseudotrichonymphae genomovar. CFP2	1,114,206	0.106	0.9957283	0.9694305	0.833	0.105	0.061	0.089	0.011	0.007
Shewanella piezotolerans WP3	5,396,476	0.034	1.003755	1.0035613	0.387	0.361	0.252	0.013	0.012	0.008
Anoxybacillus flavithermus WK1	2,846,746	0.059	1.0068969	1.0145305	0.333	0.545	0.122	0.02	0.032	0.007
Acinetobacter baumannii AB0057	4,050,513	0.056	0.9977707	1.0052395	0.601	0.289	0.111	0.034	0.016	0.006
Bifidobacterium longum subsp. infantis ATCC 15,697	2,832,748	0.062	1.0037432	1.008031	0.459	0.336	0.205	0.028	0.021	0.013
Acinetobacter baumannii AB307-0294	3,760,981	0.058	1.0031673	0.9953989	0.602	0.286	0.112	0.035	0.017	0.007
Escherichia coli O127:H6 str. E2348/69	4,965,553	0.036	1.0021794	1.001865	0.003	0.493	0.505	0	0.018	0.018
Thermosiphon africanus TCF52B	2,016,657	0.14	0.972029	1.0131087	0.779	0.164	0.057	0.109	0.023	0.008
Bacillus cereus AH187	5,269,030	0.075	0.9968755	0.9925518	0.813	0.097	0.09	0.061	0.007	0.007
Listeria monocytogenes HCC23	2,976,212	0.063	0.9982868	1.0114254	0.648	0.267	0.086	0.041	0.017	0.005
Dictyoglomus turgidum DSM 6724	1,855,560	0.129	1.0287949	1.0411637	0.588	0.324	0.088	0.076	0.042	0.011
Thauera sp. MZ1T	4,496,212	0.15	0.9936007	1.0044723	0.667	0.207	0.126	0.1	0.031	0.019
Shewanella baltica OS223	5,145,902	0.028	1.0016028	1.0042132	0.14	0.518	0.343	0.004	0.015	0.01
Methylocella silvestris BL2	4,305,430	0.125	1.0007528	0.9979477	0.399	0.481	0.12	0.05	0.06	0.015
Bacillus cereus B4264	5,419,036	0.077	0.9969726	0.9957889	0.823	0.093	0.085	0.063	0.007	0.007
Cyanothece sp. PCC 8801	4,679,413	0.052	0.99858	0.9985886	0.587	0.308	0.105	0.03	0.016	0.005
Borrelia burgdorferi ZS7	906,707	0.174	0.9873567	0.9946354	0.789	0.193	0.018	0.137	0.034	0.003
Cyanothece sp. PCC 7424	5,942,652	0.06	0.9903774	1.0018427	0.626	0.294	0.08	0.038	0.018	0.005
Escherichia fergusonii ATCC 35,469	4,588,711	0.034	1.0068523	0.9987728	0	0.488	0.512	0	0.017	0.018
Escherichia coli IAI1	4,700,560	0.037	1.0023158	0.9946072	0.005	0.495	0.5	0	0.018	0.018
Escherichia coli S88	1,632,268	0.036	0.999689	0.9991693	0.004	0.486	0.51	0	0.018	0.019
Vibrio splendidus LGP32	5,075,519	0.032	0.998859	0.9944684	0.367	0.329	0.304	0.012	0.01	0.01
Escherichia coli ED1a	5,209,548	0.036	1.0018236	0.9901506	0.005	0.477	0.519	0	0.017	0.019
Escherichia coli 55,989	5,154,862	0.036	1.0025637	0.9960419	0.004	0.481	0.516	0	0.017	0.019
Escherichia coli IAI39	5,132,068	0.036	1.0018527	0.9900596	0.003	0.483	0.513	0	0.017	0.019
Escherichia coli UMN026	5,202,090	0.036	1.0032184	0.9981483	0.004	0.491	0.505	0	0.018	0.018
Methylobacterium chloromethanicum CM4	5,777,908	0.143	1.0063543	0.9963483	0.687	0.205	0.108	0.098	0.029	0.015
Acidithiobacillus ferrooxidans ATCC 23,270	2,982,397	0.051	0.994021	1.0169309	0.437	0.319	0.244	0.022	0.016	0.012
Desulfurococcus kamchatkensis 1221n	1,365,223	0.026	0.9667267	0.95907	0.247	0.496	0.257	0.007	0.013	0.007

(continued on next page)

Table 1 (continued)

Organism	Length	ld (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
Desulfatibacillum alkenivorans AK-01	6,517,073	0.056	0.9979562	0.9904228	0.104	0.595	0.301	0.006	0.033	0.017
Desulfovibrio vulgaris str. 'Miyazaki F'	4,040,304	0.124	1.0063476	0.9964687	0.698	0.161	0.141	0.086	0.02	0.017
Pseudomonas aeruginosa LESB58	6,601,757	0.124	0.9961324	1.002585	0.631	0.131	0.238	0.078	0.016	0.03
Bacillus cereus G9842	5,387,334	0.077	0.9976883	0.9938997	0.825	0.092	0.083	0.064	0.007	0.006
Bacillus cereus AH820	5,302,683	0.076	0.9968924	0.991567	0.816	0.093	0.09	0.062	0.007	0.007
Desulfotobacterium hafniense DCB-2	5,279,134	0.036	1.0048066	1.0101146	0.049	0.675	0.276	0.002	0.024	0.01
Chloroflexus aggregans DSM 9485	4,684,931	0.044	0.9989159	0.9999351	0.273	0.325	0.402	0.012	0.014	0.018
Methanosphaerula palustris E1-9c	2,922,917	0.05	0.9877183	1.0033902	0.166	0.485	0.349	0.008	0.024	0.017
Buchnera aphidicola str. 5A (Acyrtosiphon pisum)	642,122	0.187	1.0123634	1.0113924	0.903	0.059	0.038	0.169	0.011	0.007
Buchnera aphidicola str. Tuc7 (Acyrtosiphon pisum)	641,895	0.187	1.0124673	1.0107464	0.903	0.058	0.039	0.169	0.011	0.007
Bifidobacterium animalis subsp. lactis AD011	1,933,695	0.076	0.995643	1.0128356	0.42	0.421	0.159	0.032	0.032	0.012
Clostridium kluyveri NBRC 12,016	3,896,121	0.125	0.9888637	0.9723739	0.769	0.147	0.083	0.096	0.018	0.01
Haemophilus parasuis SH0165	2,269,156	0.052	1.0015494	1.0143654	0.558	0.337	0.106	0.029	0.018	0.006
Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27,774	2,873,437	0.069	1.0036721	0.9987076	0.275	0.422	0.303	0.019	0.029	0.021
Cyanotheca sp. PCC 7425	5,374,574	0.033	0.9995039	0.994329	0.006	0.575	0.419	0	0.019	0.014
Arthrobacter chlorophenolicus A6	4,395,537	0.109	1.0015356	0.9981222	0.714	0.095	0.191	0.078	0.01	0.021
Anaeromyxobacter dehalogenans 2CP-1	5,029,329	0.237	0.9898209	0.9996042	0.776	0.125	0.099	0.184	0.03	0.023
Methylobacterium nodulans ORS 2060	7,772,460	0.144	0.9960864	0.9968926	0.73	0.183	0.087	0.105	0.026	0.013
Mycobacterium leprae Br4923	3,268,071	0.04	0.9924363	1.0122523	0.44	0.234	0.326	0.018	0.009	0.013
Clostridium cellulolyticum H10	4,068,724	0.071	0.9961118	0.9970747	0.649	0.155	0.196	0.046	0.011	0.014
Halofermothrix orenii H 168	2,578,146	0.097	1.0044858	0.9937251	0.442	0.331	0.227	0.043	0.032	0.022
Streptococcus pneumoniae ATCC 700,669	2,221,315	0.055	0.9971191	0.9922281	0.582	0.299	0.119	0.032	0.017	0.007
Thioalkalivibrio sp. HL-EbG7	3,464,554	0.096	1.0003401	0.9974957	0.696	0.157	0.148	0.066	0.015	0.014
Caulobacter crescentum NA1000	4,042,929	0.138	1.0070154	0.9969074	0.628	0.153	0.218	0.087	0.021	0.03
Thermomicrobium roseum DSM 5159	2,003,006	0.1	0.9965783	1.0045314	0.542	0.3	0.158	0.054	0.03	0.016
Rhodobacter sphaeroides KD131	3,152,792	0.161	1.0008627	0.9922446	0.678	0.217	0.104	0.109	0.035	0.017
Bacillus cereus Q1	5,214,195	0.075	0.9970223	0.9897236	0.813	0.096	0.09	0.061	0.007	0.007
Thermotoga neapolitana DSM 4359	1,884,562	0.073	0.9978319	0.9895897	0.037	0.581	0.382	0.003	0.042	0.028
Geobacter sp. FRC-32	4,304,501	0.036	0.9889932	1.0255382	0.099	0.524	0.377	0.004	0.019	0.014
Agrobacterium radiobacter K84	4,005,130	0.088	1.0106793	0.9821038	0.354	0.422	0.224	0.031	0.037	0.02
Agrobacterium vitis S4	3,726,375	0.068	1.0008775	0.9966164	0.25	0.464	0.286	0.017	0.031	0.019
Diaphorobacter sp. TPSY	3,796,573	0.127	1.0083322	1.0021044	0.654	0.226	0.12	0.083	0.029	0.015
Macrocooccus caseolyticus JCSC5402	2,102,324	0.064	1.0087963	1.0166558	0.784	0.094	0.122	0.05	0.006	0.008
Streptococcus uberis 0140 J	1,852,352	0.076	0.9994992	1.0016934	0.687	0.245	0.068	0.052	0.019	0.005
Anaplasma marginale str. Florida	1,202,435	0.016	1.0034069	0.9800832	0.003	0.627	0.369	0	0.01	0.006
Halorubrum lacusprofundi ATCC 49,239	2,735,295	0.153	0.990664	1.0000111	0.538	0.41	0.051	0.082	0.063	0.008
Chloroflexus sp. Y-400-fl	5,268,950	0.042	0.9942132	0.9979739	0.306	0.298	0.396	0.013	0.013	0.017
Anaerocellum thermophilum DSM 6725	2,919,718	0.103	0.9954153	1.0224094	0.624	0.267	0.109	0.064	0.028	0.011
Campylobacter lari RM2100	1,525,460	0.173	1.0038328	1.0027361	0.705	0.215	0.079	0.122	0.037	0.014
Desulfobacterium autotrophicum HRM2	5,589,073	0.053	0.9985782	0.9962684	0.007	0.703	0.29	0	0.037	0.015
Nautilia profundicola AmH	1,676,444	0.11	0.9957658	1.0074583	0.726	0.171	0.103	0.08	0.019	0.011
Staphylococcus carnosus subsp. carnosus TM300	2,566,424	0.085	0.9877615	0.9757926	0.812	0.124	0.064	0.069	0.011	0.005
Salmonella enterica subsp. enterica serovar Paratyphi C strain RKS4594	4,833,080	0.041	0.9962984	1.025679	0.034	0.492	0.473	0.001	0.02	0.02
Mycobacterium bovis BCG str. Tokyo 172	4,371,711	0.108	0.9996805	0.9956392	0.663	0.159	0.177	0.072	0.017	0.019
Brachyspira hyodysenteriae WA1	3,000,694	0.186	1.0045077	1.0164924	0.847	0.094	0.059	0.158	0.017	0.011
Wolbachia sp. wRi	1,445,873	0.084	0.9835741	0.9617328	0.774	0.17	0.057	0.065	0.014	0.005
Sulfurihydrogenibium azorense Az-Fu1	1,640,877	0.131	1.0033279	1.0078117	0.67	0.201	0.129	0.088	0.026	0.017
Persephonella marina EX-H1	1,930,284	0.112	0.9756838	1.0483685	0.431	0.239	0.33	0.048	0.027	0.037
Brucella melitensis ATCC 23,457	2,125,701	0.071	1.0022486	0.9938456	0.207	0.567	0.226	0.015	0.041	0.016
Streptococcus pneumoniae JJA	2,120,234	0.054	1.0032444	0.9974896	0.564	0.308	0.128	0.031	0.017	0.007
Streptococcus pneumoniae P1031	2,111,882	0.054	1.0113314	1.0130296	0.563	0.31	0.127	0.031	0.017	0.007
Streptococcus pneumoniae 70,585	2,184,682	0.054	1.0053912	1.0074252	0.566	0.308	0.126	0.031	0.017	0.007
Streptococcus pneumoniae Taiwan19F-14	2,112,148	0.054	1.0042599	1.0111291	0.562	0.31	0.127	0.03	0.017	0.007
Streptococcus equi subsp. zooepidemicus	2,149,868	0.057	0.9913846	1.0033026	0.362	0.411	0.227	0.021	0.024	0.013
Streptococcus equi subsp. equi 4047	2,253,793	0.057	0.9971643	1.0025438	0.389	0.394	0.216	0.022	0.022	0.012
Bacillus cereus O3BB102	5,269,628	0.076	0.9997903	0.9971818	0.816	0.094	0.09	0.062	0.007	0.007
Acidobacterium capsulatum ATCC 51,196	4,127,356	0.067	1.0062993	0.990229	0.478	0.407	0.115	0.032	0.027	0.008
Listeria monocytogenes Clip80459	2,912,690	0.064	1.0064485	0.9859769	0.651	0.26	0.089	0.042	0.017	0.006
Gemmatimonas aurantiaca T-27	4,636,964	0.106	1.0089805	0.9966525	0.561	0.308	0.13	0.06	0.033	0.014
Rhodococcus erythropolis PR4	6,516,310	0.093	1.0011215	0.9993237	0.475	0.368	0.157	0.044	0.034	0.015
Brevibacillus brevis NBRC 100,599	6,296,436	0.027	0.9982252	0.9990967	0.079	0.575	0.345	0.002	0.016	0.009
Rhodococcus opacus B4	7,913,450	0.144	0.9952656	1.0054521	0.658	0.242	0.101	0.094	0.035	0.014
Deinococcus deserti VCD115	2,819,842	0.085	1.0016172	0.9987299	0.613	0.16	0.227	0.052	0.014	0.019
Laribacter hongkongensis HLHK9	3,169,329	0.097	0.9998343	1.0003077	0.459	0.228	0.313	0.044	0.022	0.03
Azotobacter vinelandii DJ	5,365,318	0.112	0.9922589	1.0056567	0.647	0.15	0.204	0.072	0.017	0.023
Clostridium botulinum A2 str. Kyoto	4,155,278	0.169	0.9753911	0.968021	0.838	0.118	0.044	0.142	0.02	0.007
Vibrio cholerae M66-2	2,892,523	0.032	0.9881825	1.0078628	0.043	0.581	0.376	0.001	0.018	0.012
Bacillus anthracis str. CDC 684	5,230,115	0.078	0.9741147	0.8994818	0.816	0.094	0.09	0.063	0.007	0.007
Vibrio cholerae O395 chromosome I	3,024,078	0.031	0.9897639	0.9979009	0.046	0.584	0.371	0.001	0.018	0.012
Rhizobium sp. NGR234	3,925,702	0.106	1.0000081	0.9941522	0.466	0.365	0.169	0.05	0.039	0.018
Sulfolobus islandicus M.14.25	2,608,832	0.081	0.9983041	1.00597	0.8	0.143	0.057	0.065	0.012	0.005
Sulfolobus islandicus L.S.2.15	2,736,272	0.081	1.0006049	1.0220385	0.8	0.144	0.055	0.065	0.012	0.004

Table 1 (continued)

Organism	Length	Id (bits)	A/T	G/C	rd1	rd2	rd3	d1	d2	d3
<i>Corynebacterium aurimucosum</i> ATCC 700,975	2,790,189	0.05	0.9867218	1.0002553	0.658	0.18	0.161	0.033	0.009	0.008
<i>Sulfolobus islandicus</i> Y.G.57.14	2,702,058	0.079	1.0016108	1.0192116	0.79	0.153	0.058	0.063	0.012	0.005
<i>Sulfolobus islandicus</i> Y.N.15.51	2,812,165	0.08	0.9922486	0.9828287	0.791	0.151	0.058	0.063	0.012	0.005
<i>Sulfolobus islandicus</i> M.16.27	2,692,402	0.082	1.0004802	1.0097394	0.802	0.143	0.055	0.066	0.012	0.004
<i>Clostridium botulinum</i> Ba4 str. 657	3,977,794	0.169	0.9809699	0.977547	0.836	0.119	0.045	0.142	0.02	0.008
<i>Bacillus anthracis</i> str. A0248	5,227,419	0.077	0.9956929	0.9884153	0.816	0.094	0.09	0.063	0.007	0.007
<i>Pseudomonas fluorescens</i> SBW25	6,722,539	0.07	0.9924043	1.000753	0.461	0.273	0.266	0.032	0.019	0.018
<i>Vibrio cholerae</i> MJ-1236	3,149,584	0.031	1.0113612	0.9968339	0.058	0.578	0.364	0.002	0.018	0.011
<i>Beutenbergia cavernae</i> DSM 12,333	4,669,183	0.23	1.0044842	0.9971434	0.697	0.213	0.09	0.16	0.049	0.021
<i>Exiguobacterium</i> sp. AT1b	2,999,895	0.041	0.9979233	1.0010021	0.017	0.84	0.143	0.001	0.034	0.006
<i>Chlamydia trachomatis</i> Jali20	1,044,352	0.05	1.0045486	1.0013414	0.439	0.391	0.17	0.022	0.02	0.008
<i>Chlamydia trachomatis</i>	1,044,282	0.05	1.0046536	1.001415	0.439	0.391	0.17	0.022	0.02	0.009
<i>Tolomonas auensis</i> DSM 9187	3,471,292	0.042	0.9991968	0.9945493	0.007	0.379	0.615	0	0.016	0.026
<i>Corynebacterium kroppenstedtii</i> DSM 44,385	2,446,804	0.034	0.9845487	1.0011273	0.478	0.419	0.102	0.016	0.014	0.003
<i>Burkholderia glumae</i> BGR1	3,906,529	0.161	1.0002459	1.0025123	0.6	0.285	0.115	0.097	0.046	0.019
<i>Sulfolobus islandicus</i> M.16.4	2,586,647	0.082	0.9982885	1.0053005	0.805	0.14	0.055	0.066	0.011	0.005
<i>Rickettsia peacockii</i> str. Rustic	1,288,492	0.107	0.9913306	0.9825492	0.837	0.099	0.065	0.09	0.011	0.007
<i>Klebsiella pneumoniae</i> NTUH-K2044	5,248,520	0.06	1.0005341	1.0015214	0.283	0.305	0.413	0.017	0.018	0.025
<i>Candidatus Hamiltonella defensa</i> 5AT (<i>Acyrtosiphon pisum</i>)	2,110,331	0.057	1.0018538	0.9942462	0.475	0.361	0.164	0.027	0.021	0.009
<i>Escherichia coli</i> BW2952	4,578,159	0.037	1.0030667	0.9980684	0.005	0.499	0.496	0	0.018	0.018
<i>Eubacterium eligens</i> ATCC 27,750	2,144,190	0.082	0.994753	0.9957278	0.535	0.188	0.277	0.044	0.015	0.023
<i>Edwardsiella ictaluri</i> 93–146	3,812,315	0.053	1.0024506	1.0020838	0.303	0.257	0.439	0.016	0.014	0.023
<i>Eubacterium rectale</i> ATCC 33,656	3,449,685	0.053	0.989158	0.9772537	0.396	0.251	0.353	0.021	0.013	0.019
<i>Kosmotoga olearia</i> TBF 19.5.1	2,302,126	0.051	0.9802909	1.0111804	0.408	0.351	0.242	0.021	0.018	0.012
<i>Variovorax paradoxus</i> S110	5,626,353	0.151	0.9972108	1.0012759	0.607	0.237	0.156	0.092	0.036	0.024
<i>Geobacillus</i> sp. WCH70	3,464,618	0.053	0.9982213	0.9980546	0.279	0.598	0.123	0.015	0.032	0.007
<i>Desulfovibrio magneticus</i> RS-1	5,248,049	0.087	0.9994651	0.9997375	0.547	0.228	0.224	0.048	0.02	0.019
<i>Micrococcus luteus</i> NCTC 2665	2,501,097	0.196	1.0044126	0.9961057	0.81	0.106	0.084	0.158	0.021	0.017
<i>Thermococcus gammatolerans</i> EJ3	2,045,438	0.037	0.9811681	0.988462	0.1	0.535	0.365	0.004	0.02	0.014
<i>Mycoplasma conjunctivae</i>	846,214	0.172	1.0200133	1.0119672	0.8	0.159	0.042	0.137	0.027	0.007
<i>Methylobacterium extorquens</i> AM1	5,511,322	0.15	1.0098597	0.996052	0.69	0.201	0.109	0.104	0.03	0.016
<i>Bifidobacterium animalis</i> subsp. lactis BI-04	1,938,709	0.076	0.9967163	0.9979826	0.422	0.419	0.158	0.032	0.032	0.012
<i>Bifidobacterium animalis</i> subsp. lactis DSM 10,140	1,938,483	0.076	0.99669	0.9979687	0.422	0.419	0.158	0.032	0.032	0.012
<i>Bartonella grahamii</i> as4aup	2,341,328	0.071	1.0290461	0.9887091	0.585	0.301	0.113	0.042	0.021	0.008
<i>Rhizobium leguminosarum</i> bv. trifolii WSM1325	4,767,043	0.096	1.0091622	0.973902	0.372	0.417	0.211	0.036	0.04	0.02
<i>Ralstonia pickettii</i> 12D	3,647,724	0.103	0.9915309	0.9934715	0.523	0.329	0.148	0.054	0.034	0.015
<i>Dickeya dadantii</i> Ech703	4,679,450	0.048	0.9952639	0.9960018	0.151	0.452	0.398	0.007	0.022	0.019
<i>Thermococcus sibiricus</i> MM 739	1,845,800	0.062	0.9945083	0.9988818	0.447	0.43	0.124	0.028	0.027	0.008
<i>Streptococcus dysgalactiae</i> subsp. equisimilis GGS_124	2,106,340	0.054	0.9939978	0.9895393	0.587	0.295	0.118	0.032	0.016	0.006
<i>Escherichia coli</i> BL21	4,557,041	0.037	1.0020424	0.9981925	0.005	0.499	0.496	0	0.018	0.018
<i>Dickeya zeae</i> Ech1591	4,813,854	0.045	0.9995353	1.0096769	0.13	0.425	0.445	0.006	0.019	0.02
<i>Aggregatibacter aphrophilus</i> NJ8700	2,313,035	0.054	1.0065474	1.0064273	0.323	0.557	0.12	0.017	0.03	0.007
<i>Pectobacterium carotovorum</i> subsp. carotovorum PC1	4,862,913	0.036	0.9992326	0.9933098	0.03	0.492	0.478	0.001	0.018	0.017
<i>Geobacter</i> sp. M21	4,745,806	0.054	0.9844139	1.0024677	0.585	0.23	0.185	0.032	0.013	0.01
<i>Streptococcus suis</i> SC84	2,095,898	0.045	1.0039887	0.9625	0.513	0.319	0.168	0.023	0.014	0.008
<i>Streptococcus suis</i> P1/7	2,007,491	0.044	1.0078634	0.9666953	0.499	0.327	0.175	0.022	0.014	0.008
<i>Streptococcus suis</i> BM407	2,146,229	0.045	1.0077219	1.0062962	0.516	0.319	0.165	0.023	0.014	0.007
<i>Mycobacterium tuberculosis</i> KZN 1435	4,398,250	0.108	0.99868	1.0069159	0.663	0.159	0.178	0.072	0.017	0.019
<i>Escherichia coli</i> BL21(DE3)	4,570,938	0.037	0.997745	1.001669	0.006	0.5	0.495	0	0.018	0.018
<i>Candidatus Hodgkinia cicadicola</i> Dsem	143,795	0.125	0.74593	1.1106416	0.223	0.555	0.222	0.028	0.07	0.028
<i>Photorhabdus asymbiotica</i>	5,064,808	0.039	0.9920192	0.9941565	0.454	0.299	0.247	0.018	0.012	0.01
<i>Escherichia coli</i> B str. REL606	4,629,812	0.037	1.0020134	0.9971366	0.005	0.498	0.497	0	0.018	0.018
<i>Methylothermobacter mobilis</i> JLW8	2,547,570	0.041	1.0036458	0.9959682	0.143	0.6	0.257	0.006	0.024	0.01
<i>Methylovorus</i> sp. SIP3-4	2,995,511	0.052	0.9900874	1.0014104	0.131	0.587	0.282	0.007	0.031	0.015
<i>Escherichia coli</i> BL21(DE3)	4,557,508	0.037	1.0019324	0.9980693	0.005	0.499	0.496	0	0.018	0.018
<i>Helicobacter pylori</i> B38	1,576,758	0.102	0.9867581	0.9709897	0.335	0.443	0.222	0.034	0.045	0.023
<i>Hirschia baltica</i> ATCC 49,814	3,455,622	0.037	1.0036246	0.9925014	0.18	0.658	0.163	0.007	0.024	0.006
<i>Lactobacillus plantarum</i> JDM1	3,197,759	0.036	1.0026867	0.9978733	0.229	0.339	0.432	0.008	0.012	0.016
<i>Candidatus Liberibacter asiaticus</i> str. psy62	1,227,204	0.072	1.0712538	0.8851626	0.762	0.152	0.086	0.055	0.011	0.006
<i>Methylobacterium extorquens</i> DM4	5,943,768	0.141	1.0067594	0.9951896	0.685	0.207	0.108	0.097	0.029	0.015
<i>Teredinibacter turnerae</i> T7901	5,193,164	0.027	1.0034239	0.9999359	0.008	0.622	0.37	0	0.017	0.01
<i>Escherichia coli</i> O157:H7 str. TW14359	5,528,136	0.036	1.000705	0.9978462	0.002	0.478	0.519	0	0.017	0.018
<i>Neorickettsia risticii</i> str. Illinois	879,977	0.039	0.9934466	0.9981733	0.566	0.263	0.171	0.022	0.01	0.007
<i>Neisseria meningitidis</i> alpha14	2,145,295	0.058	0.9974756	1.0085466	0.019	0.727	0.254	0.001	0.042	0.015
<i>Flavobacteriaceae bacterium</i> 3519–10	2,768,102	0.058	1.0031746	1.0010785	0.268	0.443	0.288	0.015	0.026	0.017
Uncultured Termite group 1 bacterium phylotype Rs-D17	1,125,857	0.103	1.0291123	1.0485962	0.622	0.207	0.17	0.064	0.021	0.018

Table 2
Information density values for Eukaryotic chromosomes.

Scientific Name	Length(bp)	%AT	A/T	G/C	ld (bits)	RD1	RD2	RD3	d1	d2	d3
Arabidopsis thaliana-1	29,881,111	64.03	1.001	1.006	0.075	0.766	0.173	0.061	0.058	0.013	0.005
Arabidopsis thaliana-2	19,643,390	64.14	1	0.994	0.076	0.772	0.168	0.061	0.059	0.013	0.005
Arabidopsis thaliana-3	23,463,327	63.66	1.005	1	0.072	0.753	0.183	0.063	0.055	0.013	0.005
Arabidopsis thaliana-4	17,549,525	63.99	1.006	0.997	0.075	0.764	0.173	0.063	0.057	0.013	0.005
Arabidopsis thaliana-5	26,689,407	64.07	0.997	1.005	0.076	0.762	0.175	0.063	0.058	0.013	0.005
Caenorhabditis elegans-1	14,752,005	64.02	1	1	0.098	0.588	0.332	0.08	0.058	0.033	0.008
Caenorhabditis elegans-2	16,617,341	63.69	1.004	0.997	0.094	0.585	0.335	0.08	0.055	0.031	0.008
Caenorhabditis elegans-3	11,604,498	64.1	1.004	1.006	0.101	0.578	0.343	0.079	0.058	0.034	0.008
Caenorhabditis elegans-4	14,447,291	65.27	0.995	0.991	0.106	0.644	0.293	0.063	0.068	0.031	0.007
Caenorhabditis elegans-5	20,519,990	64.51	0.999	0.996	0.098	0.628	0.301	0.071	0.062	0.030	0.007
Caenorhabditis elegans-X	17,287,036	64.8	1.003	0.999	0.096	0.67	0.274	0.056	0.064	0.026	0.005
Drosophila melanogaster 2r	20,302,755	56.53	1.001	1	0.034	0.36	0.464	0.176	0.012	0.016	0.006
Drosophila melanogaster 2 l	22,217,931	58.07	1.001	1.001	0.043	0.443	0.414	0.143	0.019	0.018	0.006
Drosophila melanogaster 3r	27,890,790	57.08	1.003	0.997	0.039	0.374	0.461	0.165	0.015	0.018	0.006
Drosophila melanogaster 3 l	23,352,213	57.08	0.998	0.998	0.042	0.429	0.341	0.23	0.018	0.014	0.010
Drosophila melanogaster 4	1,237,870	64.65	0.994	0.997	0.077	0.827	0.108	0.065	0.063	0.008	0.005
Drosophila melanogaster x	21,780,003	57.35	0.998	1.001	0.042	0.377	0.438	0.184	0.016	0.018	0.008
Encephalitozoon cuniculi-1	209,982	50.37	0.997	1.007	0.044	0.004	0.664	0.332	0.000	0.029	0.015
Encephalitozoon cuniculi-2	194,439	53.07	1.014	1.026	0.044	0.063	0.604	0.333	0.003	0.027	0.015
Encephalitozoon cuniculi-3	218,328	52.77	1.027	1.007	0.045	0.051	0.646	0.303	0.002	0.029	0.014
Encephalitozoon cuniculi-4	211,018	53.27	1.029	1.098	0.048	0.082	0.596	0.323	0.004	0.028	0.015
Encephalitozoon cuniculi-5	220,294	52.71	0.986	0.981	0.048	0.045	0.629	0.326	0.002	0.030	0.016
Encephalitozoon cuniculi-6	226,573	52.82	1.029	1.083	0.048	0.061	0.638	0.301	0.003	0.030	0.014
Encephalitozoon cuniculi-7	238,147	52.21	1.006	1.047	0.045	0.035	0.653	0.312	0.002	0.030	0.014
Encephalitozoon cuniculi-8	262,796	52.66	0.994	1.007	0.046	0.044	0.637	0.319	0.002	0.030	0.015
Encephalitozoon cuniculi-9	267,509	52.72	0.985	0.938	0.045	0.056	0.627	0.317	0.002	0.028	0.014
Encephalitozoon cuniculi-10	250,202	53.3	0.993	0.979	0.046	0.07	0.625	0.306	0.003	0.029	0.014
Encephalitozoon cuniculi-11	197,426	53.37	0.99	0.977	0.046	0.071	0.609	0.32	0.003	0.028	0.015
Guillardia theta-1	196,216	74.25	0.997	0.99	0.204	0.87	0.113	0.018	0.177	0.023	0.004
Guillardia theta-2	180,915	73.29	0.991	0.993	0.189	0.862	0.121	0.017	0.163	0.023	0.003
Guillardia theta-3	174,133	73.05	0.999	1.013	0.185	0.859	0.121	0.02	0.159	0.022	0.004
Leishmania –1	268,984	36.92	1.026	0.951	0.086	0.587	0.232	0.181	0.050	0.020	0.015
Homo sapiens 1	282,193,664	52.66	0.99	1	0.067	0.031	0.757	0.212	0.002	0.050	0.089
Homo sapiens 2	253,256,583	59.76	0.997	1.001	0.081	0.341	0.536	0.123	0.027	0.043	0.010
Homo sapiens 3	227,524,578	60.29	0.999	1	0.084	0.367	0.521	0.112	0.030	0.043	0.009
Homo sapiens 4	202,328,347	61.74	1	1	0.089	0.452	0.447	0.101	0.040	0.039	0.009
Homo sapiens 5	203,085,532	60.56	0.997	1.001	0.084	0.378	0.48	0.141	0.032	0.040	0.012
Homo sapiens 6	182,415,242	60.11	1.001	1	0.082	0.363	0.518	0.119	0.030	0.042	0.010
Homo sapiens 7	166,623,906	59.28	0.998	0.999	0.079	0.318	0.545	0.138	0.025	0.042	0.010
Homo sapiens 8	152,776,421	59.86	1.001	1	0.081	0.348	0.531	0.121	0.028	0.043	0.010
Homo sapiens 9	142,271,444	58.67	1	0.999	0.077	0.284	0.577	0.138	0.022	0.044	0.011
Homo sapiens 10	145,589,288	58.38	0.998	0.999	0.076	0.267	0.593	0.139	0.020	0.045	0.011
Homo sapiens 11	150,783,553	49.58	0.999	0.999	0.077	0.268	0.597	0.134	0.021	0.046	0.010
Homo sapiens 12	144,282,489	59.22	0.999	1.002	0.079	0.314	0.555	0.131	0.025	0.044	0.010
Homo sapiens 13	119,744,898	61.46	0.997	0.998	0.087	0.438	0.451	0.111	0.038	0.039	0.009
Homo sapiens 14	106,953,321	59.11	0.992	1.003	0.079	0.307	0.561	0.132	0.024	0.044	0.010
Homo sapiens 15	101,380,521	57.8	1.001	0.998	0.075	0.235	0.619	0.146	0.018	0.046	0.011
Homo sapiens 16	104,298,331	55.24	0.994	1.003	0.068	0.117	0.7	0.184	0.008	0.048	0.013
Homo sapiens 17	89,504,553	54.44	0.997	0.998	0.068	0.084	0.716	0.2	0.006	0.049	0.014
Homo sapiens 18	86,677,548	60.2	0.999	1.001	0.275	0.11	0.153	0.737	0.030	0.042	0.203
Homo sapiens 19	74,962,845	51.65	0.997	1.002	0.069	0.012	0.761	0.227	0.001	0.052	0.016
Homo sapiens 20	66,668,005	55.88	0.987	1.003	0.078	0.128	0.746	0.126	0.010	0.058	0.010
Homo sapiens 21	44,907,570	59.12	1.006	0.998	0.077	0.314	0.536	0.15	0.024	0.041	0.012
Homo sapiens 22	234,226	52.13	1.005	0.999	0.069	0.019	0.772	0.209	0.001	0.053	0.014
Homo sapiens X	162,599,930	60.54	0.996	1.001	0.084	0.387	0.509	0.104	0.033	0.043	0.009
Homo sapiens Y	51,513,584	60.94	0.987	1.007	0.089	0.39	0.509	0.101	0.035	0.045	0.009
Mus musculus-1	4,276,346	56.65	0.994	1	0.076	0.168	0.698	0.134	0.013	0.053	0.010
Mus musculus-2	3,436,263	54.93	1.006	0.998	0.071	0.099	0.744	0.156	0.007	0.053	0.011
Mus musculus-3	1,629,407	57.55	0.975	0.998	0.076	0.217	0.662	0.121	0.017	0.051	0.009
Mus musculus-4	2,246,120	55.98	0.955	1.008	0.073	0.144	0.717	0.139	0.011	0.053	0.010
Mus musculus-5	5,070,949	55.18	1.024	1.001	0.072	0.109	0.717	0.174	0.008	0.051	0.013
Mus musculus-6	4,961,195	57.61	0.997	1.002	0.076	0.22	0.645	0.135	0.017	0.049	0.010
Mus musculus-7	3,844,875	55.59	1.002	1.011	0.075	0.121	0.748	0.131	0.009	0.056	0.010
Mus musculus-8	1,268,655	56.5	0.972	0.996	0.076	0.163	0.716	0.121	0.012	0.054	0.009
Mus musculus-9	1,408,650	55.31	0.964	1.027	0.069	0.121	0.712	0.167	0.008	0.049	0.011
Mus musculus-10	2,794,344	53.93	0.965	1.014	0.069	0.067	0.767	0.166	0.005	0.053	0.011
Mus musculus-11	11,286,516	55.01	0.998	0.999	0.072	0.101	0.738	0.161	0.007	0.053	0.012
Mus musculus-12	1,447,283	58.94	1.022	1.004	0.083	0.28	0.612	0.109	0.023	0.051	0.009
Mus musculus-13	3,832,228	57.69	1.035	0.991	0.076	0.228	0.646	0.126	0.017	0.049	0.010
Mus musculus-14	2,795,615	58.79	0.968	1.008	0.083	0.272	0.621	0.107	0.023	0.051	0.009
Mus musculus-15	3,250,526	52.37	0.98	1.007	0.072	0.023	0.821	0.156	0.002	0.059	0.011
Mus musculus-16	1,758,822	53.68	0.985	1.011	0.073	0.054	0.801	0.145	0.004	0.059	0.011
Mus musculus-17	4,254,052	53.61	0.982	0.994	0.07	0.054	0.78	0.166	0.004	0.055	0.012

Table 2 (continued)

Scientific Name	Length(bp)	%AT	A/T	G/C	Id (bits)	RD1	RD2	RD3	d1	d2	d3
Mus musculus-18	2,281,345	58.7	0.98	1.012	0.074	0.298	0.583	0.119	0.022	0.043	0.009
Mus musculus-19	505,529	52.59	1.004	0.984	0.069	0.028	0.782	0.19	0.002	0.054	0.013
Mus musculus-X	6,176,721	58.58	1.029	0.994	0.081	0.264	0.626	0.11	0.021	0.051	0.009
Mus musculus-Y	369,507	61.41	0.901	1.046	0.093	0.422	0.494	0.084	0.039	0.046	0.008
Pan troglodytes 1	229,575,298	57.28	0.998	0.999	0.077	0.273	0.593	0.134	x	x	x
Pan troglodytes 2	203,813,066	59.06	0.998	1.001	0.085	0.375	0.517	0.109	x	x	x
Pan troglodytes 3	209,662,276	59.72	0.998	1.001	0.091	0.462	0.44	0.097	x	x	x
Pan troglodytes 4	188,378,868	59	0.998	1.001	0.085	0.386	0.505	0.109	x	x	x
Pan troglodytes 5	175,429,504	59.18	0.999	1.001	0.084	0.385	0.503	0.112	x	x	x
Pan troglodytes 6	161,576,975	57.98	0.999	0.999	0.08	0.341	0.531	0.128	x	x	x
Pan troglodytes 7	149,542,033	58.7	1.001	1.001	0.082	0.358	0.525	0.117	x	x	x
Pan troglodytes 8	138,322,177	57.17	0.996	1.001	0.077	0.275	0.59	0.135	x	x	x
Pan troglodytes 9	136,640,551	57.24	0.999	1.001	0.077	0.274	0.595	0.131	x	x	x
Pan troglodytes 10	135,301,796	56.15	0.998	1.001	0.078	0.296	0.571	0.132	x	x	x
Pan troglodytes 11	123,086,034	57	1.001	0.998	0.078	0.3	0.57	0.13	x	x	x
Pan troglodytes 12	117,159,028	57.46	0.998	0.999	0.079	0.307	0.565	0.128	x	x	x
Pan troglodytes 13	134,309,081	59.17	0.996	1.002	0.084	0.385	0.504	0.111	x	x	x
Pan troglodytes 14	97,804,244	58.29	0.995	0.998	0.086	0.435	0.455	0.111	x	x	x
Pan troglodytes 15	106,954,593	56.82	0.992	1.002	0.079	0.317	0.556	0.127	x	x	x
Pan troglodytes 16	101,535,987	55.43	1.001	0.997	0.076	0.261	0.604	0.136	x	x	x
Pan troglodytes 17	73,346,066	59.16	0.998	0.999	0.083	0.38	0.505	0.115	x	x	x
Pan troglodytes 18	83,875,239	53.98	0.992	1.003	0.069	0.123	0.7	0.177	x	x	x
Pan troglodytes 19	82,489,036	52.08	1.005	0.996	0.068	0.058	0.736	0.206	x	x	x
Pan troglodytes 20	61,571,712	51.47	0.996	1.002	0.067	0.017	0.739	0.245	x	x	x
Pan troglodytes 21	65,473,740	54.99	0.986	1.002	0.072	0.151	0.695	0.154	x	x	x
Pan troglodytes 22	47,338,174	55.88	1.006	0.999	0.077	0.318	0.536	0.146	x	x	x
Pan troglodytes 23	50,034,486	51.28	1.006	0.998	0.069	0.022	0.775	0.203	x	x	x
Pan troglodytes X	108,849,721	60.73	0.997	1	0.085	0.394	0.507	0.099	x	x	x
Pan troglodytes Y	6,391,777	60.4	0.988	1.003	0.086	0.364	0.527	0.109	x	x	x
Plasmodium falciparum 1	643,292	79.45	1.004	1.024	0.305	0.877	0.021	0.102	0.268	0.007	0.031
Plasmodium falciparum 2	947,102	80.25	0.996	0.99	0.354	0.023	0.103	0.324	0.283	0.007	0.033
Plasmodium falciparum 3	1,060,087	80.12	0.992	1.007	0.32	0.878	0.023	0.099	0.281	0.007	0.032
Plasmodium falciparum 4	1,204,112	79.32	0.978	0.94	0.306	0.865	0.029	0.105	0.265	0.009	0.032
Plasmodium falciparum 5	1,343,552	80.66	1.026	1.03	0.335	0.87	0.024	0.105	0.292	0.008	0.035
Plasmodium falciparum 6	1,378,756	80.25	0.989	0.968	0.325	0.874	0.026	0.1	0.284	0.008	0.032
Plasmodium falciparum 7	1,351,552	80.14	0.973	0.971	0.318	0.877	0.026	0.097	0.279	0.008	0.031
Plasmodium falciparum 8	1,325,595	80.98	0.997	0.984	0.328	0.865	0.027	0.108	0.284	0.009	0.035
Plasmodium falciparum 9	1,541,723	80.98	1.005	0.976	0.341	0.874	0.022	0.104	0.298	0.008	0.035
Plasmodium falciparum 10	1,694,445	80.3	0.987	1.008	0.325	0.875	0.021	0.104	0.284	0.007	0.034
Plasmodium falciparum 11	2,035,250	81.03	1.006	1.02	0.343	0.873	0.023	0.104	0.299	0.008	0.036
Plasmodium falciparum 12	2,271,477	80.68	1.011	1.006	0.335	0.873	0.025	0.102	0.292	0.008	0.034
Plasmodium falciparum 13	2,732,359	80.75	1.008	1.018	0.337	0.878	0.024	0.098	0.295	0.008	0.033
Plasmodium falciparum 14	3,291,006	81.55	0.994	1.014	0.354	0.877	0.023	0.1	0.310	0.008	0.035
Rattus norvegicus 1	815,236	52.29	0.998	1.001	0.074	0.213	0.653	0.135	0.016	0.048	0.010
Rattus norvegicus 2	3,685,213	54.23	0.994	0.999	0.08	0.333	0.549	0.117	0.027	0.044	0.009
Rattus norvegicus 3	3,333,240	52.73	0.999	1.001	0.074	0.229	0.634	0.137	0.016	0.046	0.010
Rattus norvegicus 4	246,532	53.5	1.002	0.999	0.077	0.274	0.599	0.127	0.020	0.045	0.009
Rattus norvegicus 5	974,599	53.34	1.005	0.998	0.076	0.258	0.611	0.131	0.019	0.046	0.090
Rattus norvegicus 6	1,871,365	53.24	0.997	0.999	0.075	0.261	0.608	0.131	0.020	0.046	0.010
Rattus norvegicus 7	1,200,819	52.66	1.004	0.999	0.072	0.233	0.628	0.14	0.017	0.045	0.010
Rattus norvegicus 8	3,636,970	52.7	0.996	0.999	0.073	0.212	0.648	0.14	0.016	0.047	0.010
Rattus norvegicus 9	978,581	53.35	0.998	1.001	0.075	0.268	0.598	0.133	0.020	0.045	0.010
Rattus norvegicus 10	221,701	50.21	1.006	0.998	0.069	0.102	0.735	0.163	0.007	0.051	0.011
Rattus norvegicus 11	224,915	53.97	0.993	0.999	0.077	0.314	0.558	0.128	0.024	0.043	0.010
Rattus norvegicus 12	729,011	47.39	1.008	0.998	0.13	0.022	0.39	0.588	0.003	0.051	0.077
Rattus norvegicus 13	1,952,718	53.23	1.007	0.998	0.075	0.248	0.613	0.14	0.018	0.046	0.010
Rattus norvegicus 14	711,680	53.21	1	0.998	0.075	0.286	0.582	0.132	0.021	0.043	0.010
Rattus norvegicus 15	634,942	53.5	0.999	1	0.076	0.301	0.572	0.127	0.023	0.044	0.010
Rattus norvegicus 16	9,330,449	53.5	0.988	1.001	0.074	0.265	0.599	0.136	0.020	0.044	0.010
Rattus norvegicus 17	26,616,412	51.89	0.995	1.001	0.073	0.232	0.631	0.137	0.017	0.046	0.010
Rattus norvegicus 18	5,123,804	53.08	0.995	0.999	0.074	0.262	0.602	0.135	0.019	0.044	0.010
Rattus norvegicus 19	11,097,011	50.79	0.998	1.001	0.069	0.15	0.692	0.159	0.010	0.048	0.011
Rattus norvegicus 20	926,543	50.9	1.003	1.001	0.151	0.085	0.085	0.83	0.013	0.013	0.126
Rattus norvegicus X	1,685,449	55.71	1.001	1	0.087	0.388	0.515	0.096	0.034	0.045	0.008
Rattus norvegicus un	535,868	49.29	1.004	0.998	0.078	0.299	0.576	0.125	0.023	0.045	0.010
Schizosaccharomyces pombe-1	5,566,797	63.93	1.001	0.998	0.071	0.804	0.155	0.041	x	x	x
Schizosaccharomyces pombe-2	4,467,299	61.01	0.996	0.999	0.071	0.808	0.151	0.041	x	x	x
Schizosaccharomyces pombe-3	2,455,984	63.77	0.996	1.007	0.07	0.793	0.164	0.044	x	x	x
Saccharomyces cerevisiae-1	230,203	60.73	0.998	1.025	0.048	0.7	0.208	0.092	x	x	x
Saccharomyces cerevisiae-2	813,139	61.66	0.992	0.981	0.054	0.731	0.194	0.075	x	x	x
Saccharomyces cerevisiae-3	316,613	61.47	1.029	0.956	0.052	0.744	0.176	0.08	x	x	x
Saccharomyces cerevisiae-4	1,531,929	62.09	1.005	1.007	0.057	0.751	0.182	0.067	x	x	x
Saccharomyces cerevisiae-5	576,869	61.49	0.991	1.023	0.051	0.749	0.18	0.072	x	x	x
Saccharomyces cerevisiae-6	270,148	61.27	1.004	1.005	0.051	0.725	0.195	0.08	x	x	x

(continued on next page)

Table 2 (continued)

Scientific Name	Length(bp)	%AT	A/T	G/C	Id (bits)	RD1	RD2	RD3	d1	d2	d3
Saccharomyces cerevisiae-7	1,090,937	61.94	1.003	0.999	0.056	0.745	0.181	0.073	x	x	x
Saccharomyces cerevisiae-8	562,639	61.51	1.012	0.985	0.052	0.735	0.188	0.077	x	x	x
Saccharomyces cerevisiae-9	439,885	61.1	0.999	1.002	0.05	0.717	0.202	0.081	x	x	x
Saccharomyces cerevisiae-10	745,444	61.63	1.012	1.011	0.053	0.746	0.185	0.069	x	x	x
Saccharomyces cerevisiae-11	666,445	61.93	0.997	0.987	0.056	0.744	0.188	0.068	x	x	x
Saccharomyces cerevisiae-12	1,078,173	61.52	0.994	0.997	0.052	0.738	0.191	0.07	x	x	x
Saccharomyces cerevisiae-13	924,430	61.8	1.005	0.998	0.055	0.741	0.185	0.073	x	x	x
Saccharomyces cerevisiae-14	784,328	61.36	1.008	0.998	0.052	0.73	0.189	0.081	x	x	x
Saccharomyces cerevisiae-15	1,091,284	61.48	1.012	0.992	0.055	0.749	0.181	0.07	x	x	x
Saccharomyces cerevisiae-16	948,061	61.94	1.002	1.001	0.055	0.748	0.186	0.066	x	x	x
Cyanidioschyzon merolae strain 10D chromosome 1	422,616	AP006483	0.988902	0.997922	0.035	0.195	0.538	0.268	0.007	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 2	457,013	AP006484	1.009431	1.005338	0.034	0.207	0.531	0.263	0.007	0.018	0.009
Cyanidioschyzon merolae strain 10D chromosome 3	481,791	AP006485	0.997295	0.989286	0.033	0.153	0.55	0.297	0.005	0.018	0.01
Cyanidioschyzon merolae strain 10D chromosome 4	513,455	AP006486	1.019418	0.999585	0.033	0.173	0.556	0.272	0.006	0.018	0.009
Cyanidioschyzon merolae strain 10D chromosome 5	528,682	AP006487	0.999734	1.004042	0.033	0.176	0.566	0.258	0.006	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 6	536,163	AP006488	1.008933	0.998934	0.033	0.157	0.553	0.291	0.005	0.018	0.009
Cyanidioschyzon merolae strain 10D chromosome 7	584,452	AP006489	1.007657	1.005786	0.036	0.204	0.52	0.276	0.007	0.019	0.01
Cyanidioschyzon merolae strain 10D chromosome 8	739,753	AP006490	0.996757	1.011331	0.034	0.196	0.529	0.275	0.007	0.018	0.009
Cyanidioschyzon merolae strain 10D chromosome 9	810,151	AP006491	1.005448	1.004048	0.037	0.248	0.514	0.238	0.009	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 10	839,707	AP006492	0.989838	1.002259	0.036	0.212	0.536	0.252	0.008	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 11	852,849	AP006493	1.010852	0.996585	0.033	0.177	0.554	0.269	0.006	0.018	0.009
Cyanidioschyzon merolae strain 10D chromosome 12	859,119	AP006494	1.000823	0.999163	0.036	0.201	0.539	0.26	0.007	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 13	866,983	AP006495	0.990726	0.996449	0.034	0.193	0.531	0.276	0.007	0.018	0.009
Cyanidioschyzon merolae strain 10D chromosome 14	852,727	AP006496	1.003511	1.007732	0.039	0.264	0.501	0.236	0.01	0.02	0.009
Cyanidioschyzon merolae strain 10D chromosome 15	902,900	AP006497	0.998537	1.003305	0.035	0.199	0.533	0.267	0.007	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 16	908,485	AP006498	1.007288	0.998935	0.035	0.202	0.53	0.268	0.007	0.019	0.01
Cyanidioschyzon merolae strain 10D chromosome 17	1,232,258	AP006499	0.997062	1.001592	0.037	0.222	0.526	0.252	0.008	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 18	1,253,087	AP006500	1.012051	0.990479	0.038	0.24	0.504	0.256	0.009	0.019	0.01
Cyanidioschyzon merolae strain 10D chromosome 19	1,282,939	AP006501	0.998943	0.998534	0.036	0.215	0.523	0.262	0.008	0.019	0.009
Cyanidioschyzon merolae strain 10D chromosome 20	1,621,617	AP006502	1.003626	0.999722	0.036	0.197	0.542	0.261	0.007	0.019	0.009
Ashbya gossypii ATCC chromosome 1	691,920	NC_005782	1.01259	1.003906	0.012	0.091	0.545	0.365	0.001	0.006	0.004
Ashbya gossypii ATCC chromosome 2	867,696	NC_005783	0.994982	1.021529	0.011	0.054	0.55	0.397	0.001	0.006	0.004
Ashbya gossypii ATCC chromosome 3	907,057	NC_005784	0.995271	1.001765	0.013	0.196	0.5	0.304	0.003	0.007	0.004
Ashbya gossypii ATCC chromosome 4	1,466,912	NC_005785	1.003963	0.998957	0.012	0.125	0.516	0.359	0.001	0.006	0.004
Ashbya gossypii ATCC chromosome 5	1,519,138	NC_005786	1.000503	1.004315	0.011	0.051	0.57	0.379	0.001	0.006	0.004
Ashbya gossypii ATCC chromosome 6	1,813,154	NC_005787	1.004685	1.001206	0.011	0.091	0.546	0.363	0.001	0.006	0.004
Ashbya gossypii ATCC chromosome 7	1,476,507	NC_005788	0.997991	1.007202	0.012	0.099	0.536	0.365	0.001	0.006	0.004
Candida glabrata Chromosome A	485,192	NC_005967	1.017629	0.99058	0.043	0.703	0.207	0.089	0.03	0.009	0.004
Candida glabrata Chromosome B	502,101	NC_005968	0.99061	0.98321	0.049	0.732	0.192	0.075	0.036	0.009	0.004
Candida glabrata Chromosome C	558,804	NC_006026	1.002562	1.014808	0.043	0.713	0.2	0.086	0.031	0.009	0.004
Candida glabrata Chromosome D	651,701	NC_006027	1.011651	1.002263	0.048	0.719	0.206	0.074	0.035	0.01	0.004
Candida glabrata Chromosome E	687,501	NC_006028	0.993806	1.009159	0.049	0.72	0.202	0.078	0.035	0.01	0.004
Candida glabrata Chromosome F	927,101	NC_006029	0.998797	0.99858	0.055	0.756	0.183	0.062	0.041	0.01	0.003
Candida glabrata Chromosome G	992,211	NC_006030	0.990895	1.00046	0.052	0.735	0.196	0.069	0.038	0.01	0.004
Candida glabrata Chromosome H	1,050,361	NC_006031	0.996487	0.99833	0.055	0.754	0.181	0.065	0.042	0.01	0.004
Candida glabrata Chromosome I	1,089,401	NC_006032	1.001018	0.994455	0.051	0.753	0.179	0.068	0.039	0.009	0.003
Candida glabrata Chromosome J	1,192,501	NC_006033	0.995579	0.994516	0.05	0.726	0.2	0.074	0.036	0.01	0.004
Candida glabrata Chromosome K	1,302,002	NC_006034	0.993393	0.981993	0.054	0.75	0.187	0.063	0.04	0.01	0.003
Candida glabrata Chromosome L	1,440,588	NC_006035	1.007403	1.007102	0.053	0.743	0.186	0.07	0.04	0.01	0.004
Candida glabrata Chromosome M	1,400,893	NC_006036	0.994619	1.000045	0.052	0.738	0.193	0.069	0.038	0.01	0.004
Kluyveromyces lactis NRRL Y-1140 Chromosome A	1,062,590	NC_006037	0.988934	0.998577	0.051	0.674	0.211	0.115	0.034	0.011	0.006
Kluyveromyces lactis NRRL Y-1140 Chromosome B	1,320,834	NC_006038	0.994937	1.004558	0.052	0.67	0.221	0.109	0.035	0.012	0.006
Kluyveromyces lactis NRRL Y-1140 Chromosome C	1,753,957	NC_006039	1.012193	0.992446	0.057	0.692	0.213	0.095	0.039	0.012	0.005
Kluyveromyces lactis NRRL Y-1140 Chromosome D	1,715,506	NC_006040	1.007924	1.000075	0.052	0.683	0.214	0.104	0.036	0.011	0.005
Kluyveromyces lactis NRRL Y-1140 Chromosome E	2,234,072	NC_006041	0.991351	0.993255	0.053	0.678	0.215	0.106	0.036	0.011	0.006
Kluyveromyces lactis NRRL Y-1140 Chromosome F	2,602,197	NC_006042	1.000983	0.997551	0.056	0.69	0.213	0.097	0.038	0.012	0.005
Debaryomyces hansenii CBS767 Chromosome A	1,249,565	NC_006043	1.018219	1.015674	0.064	0.815	0.118	0.068	0.052	0.008	0.004
Debaryomyces hansenii Chromosome B	1,349,926	NC_006044	1.008018	0.996714	0.063	0.819	0.114	0.068	0.052	0.007	0.004
Debaryomyces hansenii Chromosome C	1,592,360	NC_006045	1.002515	1.002723	0.066	0.818	0.114	0.068	0.054	0.008	0.004
Debaryomyces hansenii Chromosome D	1,602,771	NC_006046	0.990971	1.002998	0.066	0.82	0.112	0.068	0.054	0.007	0.004
Debaryomyces hansenii Chromosome E	2,037,969	NC_006047	0.991918	0.993979	0.071	0.84	0.107	0.053	0.06	0.008	0.004
Debaryomyces hansenii Chromosome F	2,336,804	NC_006048	1.010479	1.011831	0.066	0.823	0.114	0.064	0.054	0.008	0.004
Debaryomyces hansenii Chromosome G	2,051,428	NC_006049	0.983594	0.988159	0.069	0.824	0.112	0.063	0.057	0.008	0.004
Yarrowia lipolytica_CLIB122 Chromosome A	2,303,261	NC_006067	1.001724	0.996923	0.018	0.027	0.625	0.348	0	0.011	0.006
Yarrowia lipolytica_CLIB122 Chromosome B	3,066,374	NC_006068	1.006719	0.996352	0.018	0.019	0.635	0.346	0	0.011	0.006
Yarrowia lipolytica_CLIB122 Chromosome C	3,272,609	NC_006069	0.999203	0.999671	0.018	0.029	0.65	0.321	0.001	0.012	0.006
Yarrowia lipolytica_CLIB122 Chromosome D	3,633,272	NC_006070	0.99858	0.996336	0.018	0.007	0.649	0.344	0	0.012	0.006
Yarrowia lipolytica_CLIB122 Chromosome E	4,224,103	NC_006071	1.000077	0.997063	0.018	0.007	0.638	0.356	0	0.012	0.007
Yarrowia lipolytica_CLIB122 Chromosome F	4,003,362	NC_006072	0.999409	1.004889	0.018	0.012	0.631	0.357	0	0.011	0.006
Cryptococcus neoformans var. neoformans JEC21 Chromosome 1	2,300,533	NC_006670	0.998888	1.000025	0.015	0.037	0.664	0.299	0.001	0.01	0.005
Cryptococcus neoformans var. neoformans JEC21 Chromosome 2	1,085,720	NC_006679	0.993195	1.001661	0.016	0.024	0.666	0.31	0	0.011	0.005
Cryptococcus neoformans var. neoformans JEC21	1,019,846	NC_006680	1.001737	1.000448	0.016	0.035	0.645	0.32	0.001	0.011	0.005

Table 2 (continued)

Scientific Name	Length(bp)	%AT	A/T	G/C	Id (bits)	RD1	RD2	RD3	d1	d2	d3
Chromosome 3											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 3	906,719	NC_006681	1.002951	1.002422	0.016	0.046	0.647	0.308	0.001	0.01	0.005
Chromosome 4											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 4	787,999	NC_006682	0.996732	1.015856	0.015	0.038	0.641	0.321	0.001	0.01	0.005
Chromosome 5											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 5	762,694	NC_006683	1.000788	1.005849	0.016	0.031	0.646	0.323	0.001	0.01	0.005
Chromosome 6											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 6	1,632,307	NC_006684	1.011343	1.000116	0.016	0.045	0.654	0.3	0.001	0.011	0.005
Chromosome 7											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 7	2,105,742	NC_006685	0.99856	1.001434	0.015	0.049	0.667	0.283	0.001	0.01	0.004
Chromosome 8											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 8	1,783,081	NC_006686	1.007766	1.003564	0.015	0.043	0.677	0.279	0.001	0.01	0.004
Chromosome 9											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 9	1,507,550	NC_006687	0.99718	1.009387	0.016	0.028	0.667	0.305	0	0.011	0.005
Chromosome 10											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 10	1,438,950	NC_006691	1.001251	1.004042	0.016	0.046	0.648	0.306	0.001	0.01	0.005
Chromosome 11											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 11	1,347,793	NC_006692	0.985676	0.99094	0.016	0.04	0.658	0.302	0.001	0.011	0.005
Chromosome 12											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 12	1,194,300	NC_006693	0.994512	0.997912	0.016	0.038	0.66	0.301	0.001	0.01	0.005
Chromosome 13											
Cryptococcus neoformans var. neoformans JEC21 Chromosome 13	1,178,688	NC_006694	0.99804	0.996886	0.016	0.038	0.654	0.308	0.001	0.011	0.005
Chromosome 14											
Pichia stipitis CBS 6054 Chromosome 1	2,740,984	NC_009068	0.998934	1.011557	0.046	0.492	0.304	0.204	0.023	0.014	0.009
Pichia stipitis CBS 6054 Chromosome 2	1,841,851	NC_009042	1.000067	0.994533	0.045	0.49	0.299	0.211	0.022	0.013	0.009
Pichia stipitis CBS 6054 Chromosome 3	1,803,401	NC_009043	1.004566	0.999635	0.046	0.506	0.298	0.196	0.023	0.014	0.009
Pichia stipitis CBS 6054 Chromosome 4	1,725,948	NC_009044	1.002762	0.992702	0.046	0.497	0.301	0.202	0.023	0.014	0.009
Pichia stipitis CBS 6054 Chromosome 5	1,724,953	NC_009045	1.00001	0.998555	0.046	0.5	0.298	0.202	0.023	0.014	0.009
Pichia stipitis CBS 6054 Chromosome 6	1,114,415	NC_009046	1.003458	1.008575	0.047	0.501	0.301	0.198	0.024	0.014	0.009
Pichia stipitis CBS 6054 Chromosome 7	979,380	NC_009047	1.011391	1.004315	0.045	0.511	0.297	0.193	0.023	0.013	0.009
Pichia stipitis CBS 6054 Chromosome 8	3,510,247	NC_009048	1.002246	1.005197	0.045	0.494	0.311	0.195	0.022	0.014	0.009

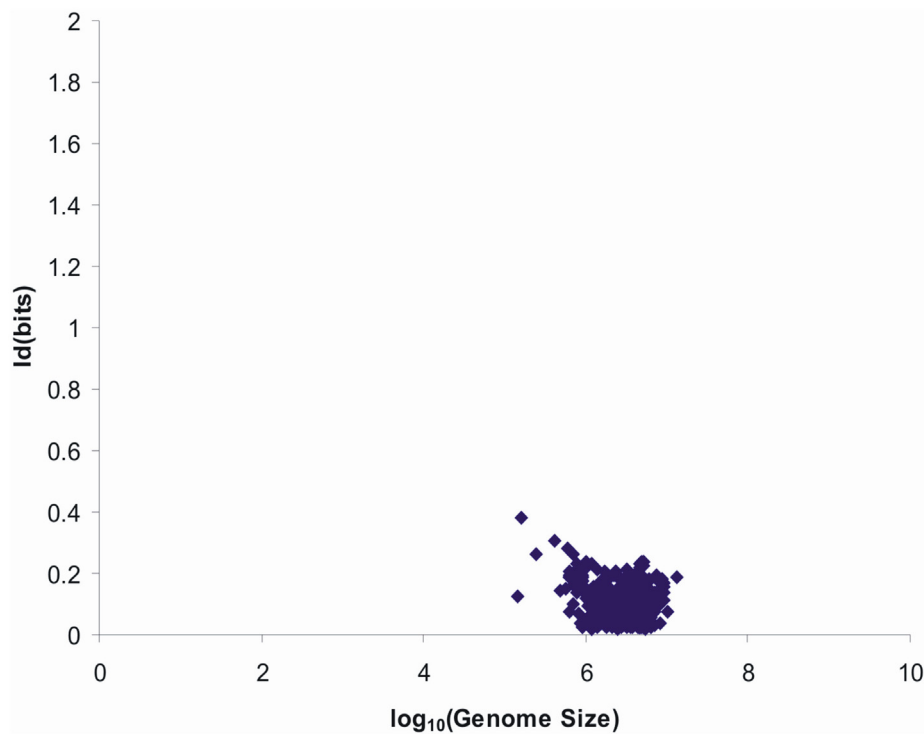


Fig. 1a. Information density distribution of Prokaryotes. Information density distribution with genome size (prokaryotes). It is found that the information density does not depend on genome size (Correlation coeff: -0.211).

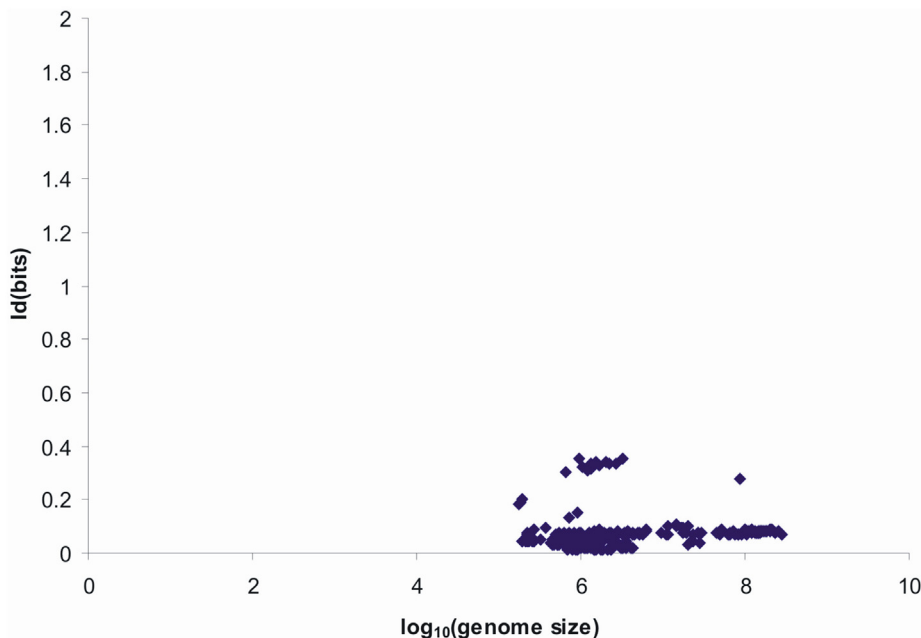


Fig. 1b. Information density distribution of Eukaryotes. Information density distribution with genome size (eukaryotes). It is found that the information density does not depend on genome size (Correlation coeff: 0.047).

of genome lengths varies over four orders of magnitude. But the Id values together with the low mean values (0.088, 0.077) suggests that both prokaryotic and eukaryotic genomes do not deviate significantly from an equiprobable and random occurrence of the letters.

3.4. Nucleotide distribution until markov third order is independent of genome size

Information content of genomes analysed is found to be independent of the size of the genome (Fig. 1a and Fig. 1b, Correlation coefficient prokaryotes: -0.211, eukaryotes: 0.047). Information density values are similar for members within certain taxonomic groups such as enterobacteriaceae, archaea, bacillaceae, vibriaceae, corynebacteria, mycobacteriaceae, streptococcaceae and chlamydiaceae (Table 1). Interestingly, the two arms of chromosome 3 and 4 in *A. thaliana* maintain similar information density. This is also seen to a limited extent for all the chromosomes in

H. sapiens (Fig. 2) and for the chromosomes of *Rattus norvegicus*. When the composition of the genomes deviates to a large extent from the equiprobable (50% AT or GC), the effect of the Markov process of order one and two on the information content is minimal (Fig. 3).

3.5. Shannon and compositional redundancy relation

There is an inverse correlation of the markovian contribution to the relative information content or Shannon redundancy arising from di and tri nucleotide arrangements (RD2 + RD3) with | %AT-50 |. This is shown by correlation values -0.902 and -0.83 for prokaryotes and eukaryotes. This is also reflected in the relative contribution of compositional redundancy (RD1) being inversely correlated to the relative contribution of the Shannon redundancy arising from a first and second order Markov process, as measured by RD2 and RD3 (RD1 with RD2: -0.831 and -0.91 prokaryotes and eukaryotes; RD1 with RD3: -0.76 and -0.70 for prokaryotes

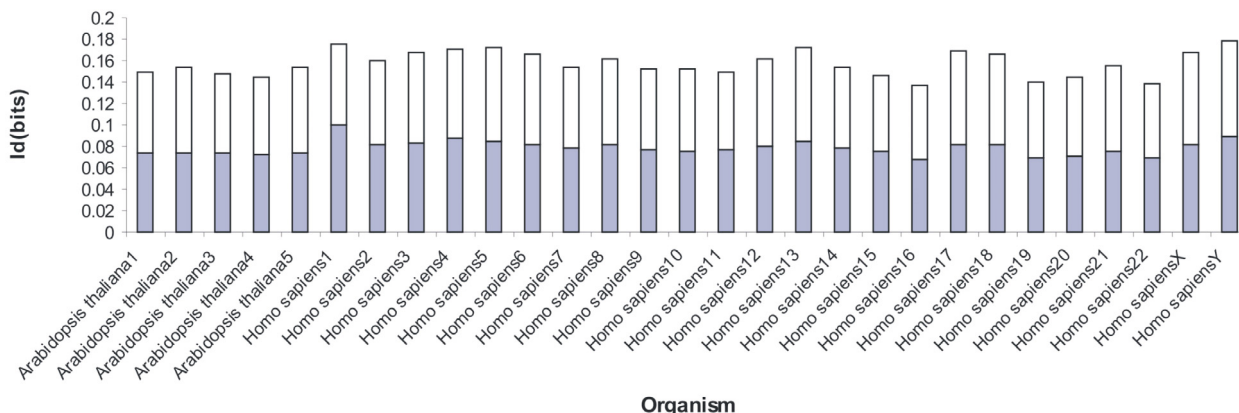


Fig. 2. Information density of Eukaryotic chromosomes. Information density values for two arms of chromosomes. p and q arms do not show much variation. All chromosomes of a particular genome show similar information density values.

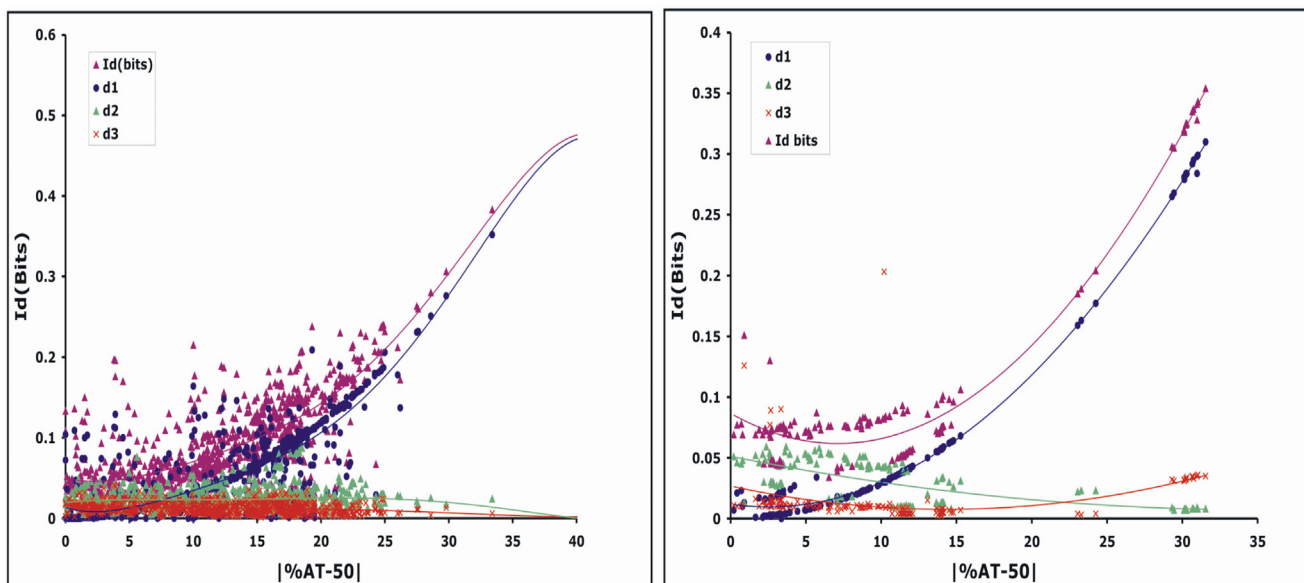


Fig. 3. Variation of %AT-50 with information density. The %AT content of the genomes was obtained from the site <http://www.cbs.dtu.dk>. The thinner line corresponds to the D1 values for the respective genomes. The inverse correlation of (RD2 + RD3) with %AT-50 and the trend of D1 with %AT-50 illustrates the balance between scalar (variation of nucleotides composition) and vector (variation in the order of occurrence of nucleotides) strategies to combat error (Key: \blacklozenge -Id \bullet -D1 \blacktriangle -D2 \times -D3).

and eukaryotes respectively). Thus, genome sequences having a high compositional redundancy show a low Shannon redundancy arising from doublet and triplet frequency distributions and vice versa. RD2 and RD3 show a positive correlation with each other (0.29 for prokaryotes and 0.40 eukaryotes). Most of the prokaryote sequences show a higher value for compositional redundancy (RD1) as compared to Shannon redundancy (RD2, RD3) (Tab.1). The contribution of RD1 is greater than RD2 and RD3 in general for eukaryotes as obviously noticed in chromosomes of *Guillardia theta*, *A. thaliana*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *C.elegans*. However an exception arise in case of *Plasmodium falciparum* (Chromosome 2) deviates with RD2 and RD3 contributing higher than RD1 (Tab.2). Also in most chromosomes associated with *homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila* and *chimpanzee*, RD2 contributes higher than RD1, while the rest of the other chromosomes in these genomes show a mixed trend of RD2 and RD3 contributing higher than RD1.

3.6. Eukaryotic chromosomes are highly variant and affine within organism

Information content of the complete genomes of organisms can be used to cluster or group these organisms. Both prokaryotes and eukaryotes follow the clustering pattern, when grouped with information density values., when grouped with information density values. The chromosomes are found to group according to the taxa at the species level. In case of prokaryotes (Supplementary Fig. 1), the grouping is very complex and obscure, while the eukaryotes (Fig. 4) in general tend to show clustering at the species level. For prokaryotes, atleast five groups are formed. Since there is no organism wise clustering in the prokaryotes, we replaced organism with the next Taxonomical hierarchy. There is a mix of actinobacteria, archae, chlamydiae, cyanobacteria, firmicutes and proteobacteria. Chromosomes tend to widespread in terms of information content. For eukaryotes, this clustering also suggests a closer affinity amongst the chromosomes of the same species. The genomes of different groups of organisms (eukaryotes and prokaryotes) probably exhibit different strategies to achieve homeostasis of Id.

4. Discussion

This study is an attempt to apply information theoretic concepts to a variety of whole genome sequences. The genomes have been considered as a string of letters similar to any sentence in the English language with fewer alphabet elements, similar to investigating the evolution of literary compositions. The frequencies of letters and their combinations affect the vocabulary, syntax, grammar and error correction capabilities of the genome.

Chargaff proposed that, to a close approximation, the first parity rule also holds for individual single strands of a double stranded DNA duplex. This rule, called the second parity rule, has been shown to hold for several organisms, including parts of *E. coli* genome, Vaccinia virus, Herpes Simplex virus and yeast. This work provides further confirmation of the above rule for all the genomes studied. The equivalence of **intra**-strand %A=%T and %G=%C introduces a degree of compositional redundancy (D1). The first parity rule of Chargaff relating to the well-known **inter**-strand equivalence results in both strands of dsDNA genomes carrying the same information density and being subject to the same messaging strategies during evolution.

It can be seen here that genomes have a narrow range of information density compared to the range they have as a language built from four letters. Further, Id values are quite low for all genomes. An equiprobable and random sequence provides the maximum potential for diverse message. Only large decreases from the state of maximum message variety, given by high Id values, will provide for innately high error combating possibilities while, of course, reducing the message variety. Hence, we suggest, that the maintenance of information density values, or the homeostasis, close to zero is a common aspect of all genomes to provide high message variety at the expense of retrievable information. This has probably resulted in the requirement and development of error correction through other molecular mechanisms in organisms to counter mutation, deletion, and insertion of DNA.

The deviation seen in the *P. falciparum* (where Id values are around 0.3) needs to be further examined whether this arises from the high A + T content and the frequency of repeats. However, in terms of messaging strategy it suggests, that being a parasite

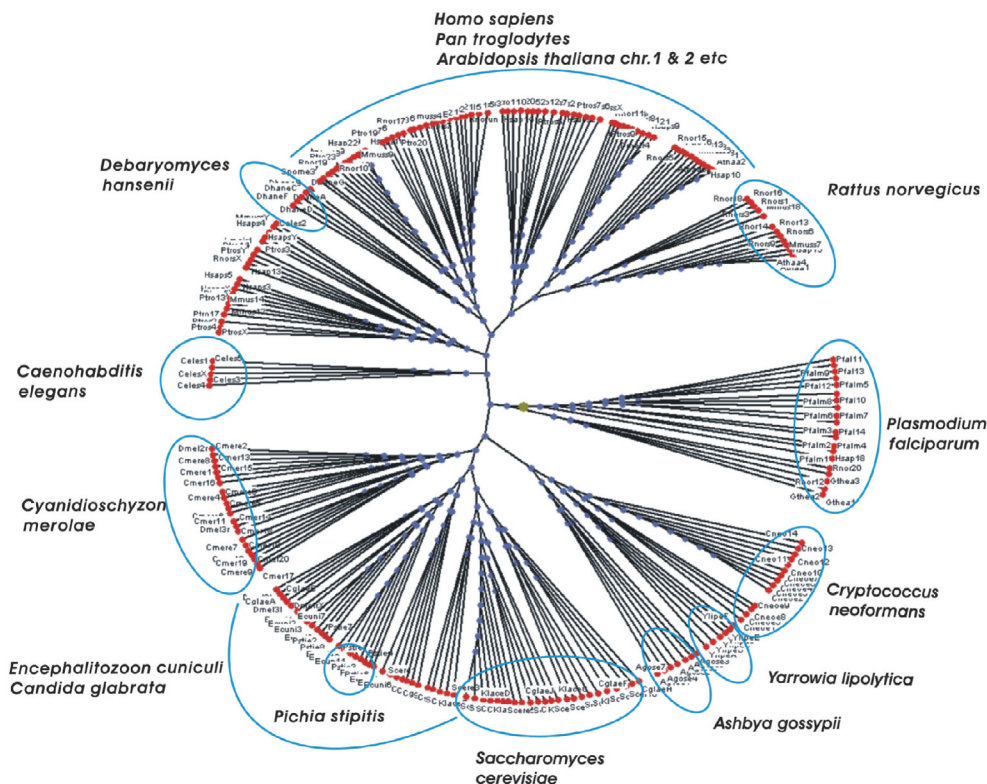


Fig. 4. Clustering of selected eukaryotic genomes based on Id values. Id values calculated was used to cluster based on UPGMA method.

which has to go through two organisms with different error correction capabilities, it would be good strategy to provide for better inherent redundancy and message retrieval framework at the expense of message variety.

In addition to the existing methods of classification of organisms, such as sequence homology, morphological similarity, fossil records (era of living) etc, information density values can also be used for the clustering of the organisms. In the case of eukaryotes, the grouping seen at the genus levels suggests that although the junk DNA or non-coding regions are not same for chromosomes of same organism, they maintain a homeostasis in terms of information content.

The genomes of different groups of organisms (eukaryotes and prokaryotes) probably exhibit different strategies to achieve homeostasis of Id. The inverse correlation of RD1 with RD2 and RD3 suggests that fine-tuning of compositional redundancy and Shannon redundancy might be one of the major ways of achieving the homeostasis. In other words, as the genome deviates from an equiprobable distribution of its letters (thus, resulting in a higher compositional redundancy), the inversely correlated contribution from the n-tuplets ($n \geq 2$) distribution (Shannon’s redundancy) help maintain a low information density (Fig. 3). The inverse correlation of (RD2 + RD3) with |%AT-50| and the trend of D_1 with |% AT-50| also illustrate the balance between scalar (variation of nucleotides composition) and vector (variations in the order of occurrence of nucleotides) strategies to combat error. Thus, compositional redundancy and Shannon redundancy, which arises from syntax, act in opposition to each other and probably helps maintain Id values within limits.

Interestingly, constant information density values and the relative contributions of the scalar and vector contributions to the redundancy are maintained amongst the different chromosomes of a particular organism and amongst the two arms of the chromosome also, in certain cases. This raises the question of what is the

smallest unit of a chromosome that maintains a constant information density. Statistical similarity between the various chromosomes of yeast has been demonstrated previously and it has been suggested that this would be the case with all the genomes (Li et al., 1998). Two hypotheses, ‘single common origin’ or ‘duplication/polyploidization of a limited set of chromosome’ were suggested to explain the uniformity seen in the various chromosomes of an organism (Bertalanffy, 1975). However, duplication/polyploidization may be ruled out because it will result in a few rather than one constant value of information density. It is possible that polyploidization of a few related sequences of common origin, which is a mix of two hypotheses, could explain the constancy amongst the chromosomes. The similarity amongst the various chromosomes of an organism could be a result of functional constraints imposed by the need to use common cellular machinery.

5. Conclusions

We have showed that Shannon’s information measure does provide a useful tool for understanding whole genomes. Information theory provides useful insights when looking at genomes as language strings which are evolving (Searls, 2002). Possibly, the presence of a number of proof-reading mechanisms at various levels in the living systems (both within organisms and in evolution in the form of natural selection) allows biological language strings to maintain higher potential information at the expense of retrievable information thereby providing the possibility of higher message variety. Due to the inherent hierarchical nature of biological systems, Shannon’s information measure might not provide an adequate means of characterizing all aspects of living systems. However, as shown here, it can provide insights into the clustering of genomes and evolution of messaging strategies of the genomes.

Funding

The current work was funded by Taif University Researchers Supporting Project number (TURSP – 2020/59), Taif university, Taif, Saudi Arabia.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors extend their appreciation to the Deanship of Scientific Research at Taif University for funding this work through Taif University Researchers Supporting Project number (TURSP - 2020/59), Taif University, Taif, Saudi Arabia. Earlier part of the information theory based genome analysis was performed at the Bioinformatics facility at MKU by DBT, Government of India.

References

- Bertalanffy, V.L., 1975. Perspectives on general system theory. George Braziller, New York.
- Brendel, V., Busse, H.G., 1984. Genome structure described by formal languages. *Nucleic Acids Res.* 12, 2561–2568.
- Brooks, D.R., Wiley, E.O., 1986. Evolution as entropy: Toward a Unified Theory of Biology. University of Chicago Press, Chicago.
- Chang, C.H., Hsieh, L.C., Chen, T.Y., Chen, H.D., Luo, L., Lee, H.C., 2005. Shannon information in complete genomes. *J. Bioinform. Comput. Biol.* 3, 587–608.
- Chomsky, N., 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge.
- Clote, P., Backofen, R., 2000. Computational Molecular Biology. John Wiley & Sons, San Francisco.
- Dewey, G., 1923. Relative frequency of English Speech Sounds. Harvard University Press, Cambridge.
- Doig, A.J., 1997. Improving the efficiency of the genetic code by varying the codon length—the perfect genetic code. *J. Theor. Biol.* 188, 355–360.
- Felsenstein, J., 1989. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164–166.
- Forsdyke, D.R., 1999. Two levels of information in DNA: relationship of Romanes' "intrinsic" variability of the reproductive system, and Bateson's "residue" to the species-dependent component of the base composition, (C+G)%. *J. Theor. Biol.* 201, 47–61.
- Gatlin, L., 1972. Information theory of living systems. Columbia University Press, New York.
- Hariri, A., Weber, B., Olmsted, J., 1990. On the validity of Shannon-information calculations for molecular biological sequences. *J. Theor. Biol.* 147, 235–254.
- Li, W., Stolovitzky, G., Bernaola-Galvan, P., Oliver, J.L., 1998. Compositional heterogeneity within, and uniformity between, DNA sequences of yeast chromosomes. *Genome Res.* 8, 916–928.
- Pesole, G., Attimonelli, M., Saccone, C., 1994. Linguistic approaches to the analysis of sequence information. *Trends Biotechnol.* 12, 401–408.
- Reddy, A.D., Prasad, B.V.L.S., Mitra, C.K., 2006a. Functional classification of transcription factor binding sites: Information content as a metric. *Journal of Integrative Bioinformatics*, 3, 0020.
- Reddy, D.A., Prasad, B.V., Mitra, C.K., 2006b. Comparative analysis of core promoter region: information content from mono and dinucleotide substitution matrices. *Comput. Biol. Chem.* 30, 58–62.
- Schneider, T.D., 1997. Information content of individual genetic sequences. *J. Theor. Biol.* 189, 427–441.
- Searls, D.B., 2002. The language of genes. *Nature*, 420, 211–217.
- Shannon, C.E., 1951. Prediction and entropy of printed English. *Bell. Sys. Tech. J.* 30, 50–64.
- Shannon, C.E., Weaver, W., 1949. The mathematical theory of communication. University of Illinois Press, Illinois.
- Sibbald, P.R., Banerjee, S., Maze, J., 1989. Calculating higher order DNA sequence information measures. *J. Theor. Biol.* 136, 475–483.
- Sneath, P.H., Snokal, R.R., 1973. Numerical Taxonomy. Freeman W.H., San Francisco.
- Staden, R., 1984. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res.* 12, 505–519.
- Tanaka, M., 1980. An application of information theory to biological evolution. *J. Theor. Biol.* 85, 789–806.
- Wallace, R., Wallace, R.G., 1998. Information Theory, Scaling Laws and the Thermodynamics of Evolution. *J. Theor. Biol.* 192, 545–559.
- Yockey, H.P., 1974. An application of information theory to the Central Dogma and the Sequence Hypothesis. *J. Theor. Biol.* 46, 369–406.
- Zeeberg, B., 2002. Shannon information theoretic computation of synonymous codon usage biases in coding regions of human and mouse genomes. *Genome Res.* 12, 944–955.
- Zipf, G.K., 1935. The Psycho-biology of language. MIT Press, New York.
- Zipf, G.K., 1942. Children's Speech. *Science* 96, 344–345.

Further Reading

- Schmidt, J.P., 1998. An information theoretic view of gapped and other alignments. *Pac. Symp. Biocomput.*, 561–572