

Research Article

A Whole Genome Pairwise Comparative and Functional Analysis of *Geobacter sulfurreducens* PCA

Ashok Selvaraj, Subazini Thankaswamy Kosalai, Rajadurai Chinnasamy Perumal, Subhashini Pitchai, and Gopal Ramesh Kumar

Bioinformatics Lab, AU-KBC Research Centre, M.I.T Campus of Anna University, Chromepet, Chennai, Tamil Nadu 600044, India

Correspondence should be addressed to Gopal Ramesh Kumar; grameshpub@au-kbc.org

Received 16 May 2013; Accepted 25 June 2013

Academic Editors: F. Barbault, S. Chavali, G. Colonna, and W. B. Fischer

Copyright © 2013 Ashok Selvaraj et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Geobacter species are involved in electricity production, bioremediations, and various environmental friendly activities. Whole genome comparative analyses of *Geobacter sulfurreducens* PCA, *Geobacter bemidjiensis* Bem, *Geobacter* sp. FRC-32, *Geobacter lovleyi* SZ, *Geobacter* sp. M21, *Geobacter metallireducens* GS-15, *Geobacter uraniireducens* Rf4 have been made to find out similarities and dissimilarities among them. For whole genome comparison of *Geobacter* species, an in-house tool, Geobacter Comparative Genomics Tool (GCGT) has been developed using BLASTALL program, and these whole genome analyses yielded conserved genes and they are used for functional prediction. The conserved genes identified are about 2184 genes, and these genes are classified into 14 groups based on the pathway information. Functions for 74 hypothetical proteins have been predicted based on the conserved genes. The predicted functions include pilus type proteins, flagellar proteins, ABC transporters, and other proteins which are involved in electron transfer. A phylogenetic tree from 16S rRNA of seven *Geobacter* species showed that *G. sulfurreducens* PCA is closely related to *G. metallireducens* GS-15 and *G. lovleyi* SZ. For evolutionary study, acetate kinase protein is used, which showed closeness to *Pelobacter propionicus*, *Pelobacter carbinolicus*, and Deferribacteraceae family bacterial species. These results will be useful to enhance electricity production by using biotechnological approaches.

1. Introduction

Geobacter species have been placed in the Geobacteraceae family which falls under the δ -Proteobacteria class. They are found in abundance where Fe(III) reduction is important, and these reductions play an important role in bioremediation [1]. These species have the ability to transfer electrons directly to electrodes or metals without any electron mediators and as a result they produce electricity or precipitate soluble metals [2, 3]. *Geobacter* species are involved in a variety of environmental friendly actions like bioremediation, electricity production, and so forth [2, 4]. *Geobacter* species till now have shown higher current production compared to other organisms, and these species are found to be having nanowires, which allows it to reach distant electron acceptors, and in turn helps in making a microbial fuel cell easier [5]. *Geobacter* species which are a strictly anaerobic organism also proved to be having the ability to grow in low oxygen

concentrations, and this has effects on this organism's growth in the subsurface environment [6].

As there is an increase in genome sequencing and thus the availability of whole genome data of many microorganisms, comparative genome analysis provides valuable information and these comparative analyses help in annotating the protein function and understanding evolutionary relationships. Comparative genome analyses have been achieved through various ways like pairwise or multiple genome comparisons or through local or global alignment methods. Computational application of these algorithms plays an important role in comparative analysis, and comparison of genomes provides information about conserved and unique genes across these genomes [7]. By identifying homologous genes through pairwise comparison with the closely related species, the hypothetical proteins can be functionally annotated [8, 9]. Phylogenetic analysis between the organisms provides a profile through which the relationship between these organisms

can be identified. Since these organisms are in a way evolved by adapting to their environmental conditions, by analyzing these phylogenetic profiles, the evolution of these organisms can be inferred [10]. An evolutionary based comparative study has been carried out by Butler et al. which showed the relationship based on evolution between Geobacteraceae species [11].

The whole genome sequence data of *Geobacter* spp. are provided in public databases, and comparative genome analysis provides insights into their metabolism. This analysis provides information about the conserved gene across the *Geobacter* species and the metabolism involved and their evolution, and we also predicted the function of hypothetical proteins in *G. sulfurreducens* PCA. Here, we compared the genomes of 7 *Geobacter* species specific to *G. sulfurreducens* PCA for which the whole genome data are available in the NCBI database [12]. A pairwise comparison tool was built specifically for these 7 species, where the output interface provides information about common and unique genes. The 7 *Geobacter* genomes compared in this work are *G. sulfurreducens* PCA, *G. bemidjiensis* Bem, *G. sp. FRC-32*, *G. lovleyi* SZ, *G. sp. M21*, *G. metallireducens* GS-15, and *G. uraniireducens* Rf4. To know the relationship between these bacteria, a phylogenetic tree was built from the 16S rRNA of these genomes. And to analyze the evolutionary relationship, a tree was constructed for acetate kinase protein, a protein involved in acetate metabolism which is also conserved in all these species. Acetate kinase enzyme is involved in phosphorylation of acetate and ATP with the formation of acetyl phosphate and ADP. This process is linked to central carbon metabolism which is very important in energy production of an organism.

2. Materials and Methods

2.1. Genome Data. The genome data of the 7 *Geobacter* species were obtained from the NCBI genome database in 2012 [12]. The genome information like genome length, gene content, GC%, and coding genes for each species was also obtained from the NCBI.

2.2. Geobacter Comparative Genomics Tool. *Geobacter* Comparative Genomics Tool (GCGT) has been built for pairwise comparative analysis of *Geobacter* genomes. Currently, this tool consists of all the 7 *Geobacter* species. This tool is part of an in-house tool, MCGT tool, which has 1205 organisms in total. GCGT is written in Perl script and runs on an Apache web server. GCGT takes total protein sequences, in fasta format, from two different organisms and aligns all sequences based on a local alignment program. The FORMATDB and BLASTALL programs are used to automate the task of sequence comparison for a given *E*-value. The corresponding BLAST [13] output is parsed by a Perl program that extracts information like accession number, protein name, score, *E*-value, identity, similarity, and individual alignment patterns for each homologous gene. The output is given in the form of tables showing homologous and nonhomologous sequences between the compared genomes. <http://mcgt.bioinfo.au-kbc.org.in/GCGT/>.

2.3. Functional Genomics and Gene Reannotation. Annotation of hypothetical proteins of *G. sulfurreducens* was carried out based on the results obtained. Each *Geobacter* species was run against *G. sulfurreducens* at an *E*-value of 0.0001 and, based on the results obtained, the function of hypothetical protein was predicted. The core genome across the *Geobacter* species was identified by running a Macro program which identified the common reference sequence IDs of *G. sulfurreducens* species among the homologs obtained from comparing with the other 6 *Geobacter* species in the GCGT tool. For the identified core genome genes, the pathway information was obtained using UniProt [14]. And also comparisons of protein coding genes, GC content, and various aspects of genome of the *Geobacter* were carried out.

2.4. Phylogenetic Analysis. A Phylogenetic analysis was done for 16S rRNA of the 7 *Geobacter* species to identify the relationship between these bacteria. The 16S rRNA protein sequence of 7 *Geobacter* species was obtained from the genome data downloaded from NCBI, and multiple sequence alignment of these proteins was carried out using ClustalX [15]. Phylogenetic tree was constructed using PHYLIP [16] package. Neighbor-joining method was used to construct phylogenetic trees, and a bootstrap analysis (1000 data resamplings) was used to determine levels of branch points obtained in neighbor-joining analysis. A consensus tree was developed using CONSENSE program, and the tree was viewed using Archaeopteryx [17] visualization software. To study the evolutionary relationship, a tree was constructed for acetate kinase protein, which is involved in acetate metabolism in *Geobacter* and also conserved across the 7 *Geobacter* genomes.

3. Results and Discussion

3.1. Pairwise Comparison. The pairwise comparison is one of the important methods of comparing microbial genomes. This method was used in identifying the list of genes which are homologous between the compared organisms. The inter-pairwise comparison between the seven *Geobacter* species genomes carried out at an *E*-value of 1×10^{-5} provided the homologous and specific genes between these genomes (Table 1). A high similarity between *Geobacter* species of *G. sulfurreducens* PCA and *G. metallireducens* GS-15 was found based on pairwise comparison. The similarity was in the range of 80%, and the least similarity was found between *G. sulfurreducens* PCA and *G. lovleyi* SZ which was in the range of ~60%. The other *Geobacter* genomes showed similarity with *G. sulfurreducens* PCA in the range of 65–70%. These data were used for the functional annotation of hypothetical genes in *G. sulfurreducens* PCA which is used as a reference organism in the comparative analysis.

The protein sequences which descend from a common ancestry will have the same function [18]. Based on the conserved genes across the *Geobacter* genomes, the functions of 74 hypothetical genes have been predicted (Table S1A in Supplementary Material found online at <http://dx.doi.org/10.1155/2013/850179>). The predicted functions include NAD dependent dehydratase (NP_951137.1)

TABLE 1: Comparative analysis of *Geobacter* species with *Geobacter sulfurreducens* PCA. This comparative analysis shows the homologous and specific genes with respect to *Geobacter sulfurreducens* PCA against other *Geobacter* species, and this comparison is carried out at an E -value of 1×10^{-5} .

Comparison organism	Reference organism (<i>Geobacter sulfurreducens</i> PCA)		
	Homologous to comparison organism	Specific to reference organism	Specific to comparison organism
<i>Geobacter metallireducens</i> GS-15	2785	642	670
<i>Geobacter uraniireducens</i> Rf4	2757	670	1176
<i>Geobacter bemidjensis</i> Bem	2709	718	1058
<i>Geobacter</i> sp. FRC-32	2675	752	927
<i>Geobacter lovleyi</i> SZ	2565	862	970
<i>Geobacter</i> sp. M21	2727	700	1088

enzyme which is involved in the cellular metabolic process, membrane protein which plays an important role in the maintenance of concentrations of ions, type-IV pilus assembly PilZ that helps in transport (NP_952245.1), which transports with the help of a transporter or pore within a cell or between cells, and FAD dependent oxidoreductase (NP_952174.1), an enzyme involved in oxidation-reduction process.

3.2. Core Genome. The genes identified from the genomes are analyzed for their conservation across the seven *Geobacter* genomes. This concept is used for studying the genomic relationship among the bacteria [19]. Here, we have identified the core genome of seven *Geobacter* genomes which was about 2184 genes (Table S1B). These core genome proteins are classified according to their functional information (Table 2). This core gene contains essential genes which code for transcription, translation, and various metabolic proteins. The core genes contain high number of membrane proteins, electron transport proteins, and efflux proteins which are involved in the electron transfer mechanism indicating the high rate of electron transfer in these organisms.

3.3. Genome Length and GC Content. The 7 *Geobacter* genomes show a considerable size variation. These variations in the size of bacterial genomes have been studied by Ussery and Hallin [20]. The genome of *G. uraniireducens* Rf4 is the largest with a size of 5,136,364nt and with 4358 protein coding genes. This *G. uraniireducens* Rf4 exhibits a slow growth rate and it also has long flagella [21]. The smallest is *G. sulfurreducens* PCA which has a genome length of 3,814,139nt and 3428 protein coding genes. The GC content is very important for a small organism and even small change in a base composition affects the coding regions [22]. The GC content of *Geobacter* organisms is in the range of 25%–75% (Table 3).

3.4. Phylogenetic Analysis. Phylogenetic tree was constructed using the 16S rRNA of *Geobacter* genomes to study the relationship among these bacteria. The 16S rRNA is used for taxonomic purposes and identifying new classes of bacteria [23]. Butler et al. have constructed phylogeny tree based on

TABLE 2: Core genome of *Geobacter* spp. classified based on their functional classes. This classification shows a high number of electron transport proteins, ATPase, and signaling and membrane proteins to be conserved across *Geobacter* species. These electron transfer proteins and membrane proteins play a significant role in electricity production.

S. no.	Functional classes	No. of proteins
1	Cell membrane proteins	26
2	ATPase proteins	26
3	Efflux proteins	19
4	Lipid metabolism	23
5	Signaling pathway	44
6	ABC transporter substrate binding proteins	32
7	Flagellar proteins	24
8	Virulence, detoxification	4
9	Phage proteins	3
10	Transcriptional regulator proteins	69
11	RNA and ribosomal proteins	48
12	DNA binding proteins	34
13	Electron transport/cytochrome type	55
14	Cell growth proteins	1503
15	Hypothetical proteins	275

697 ortholog proteins among six *Geobacter* and a *Pelobacter* species [11]. Here, the tree based on 16S rRNA provided the relationship between these 7 organisms (Figure 1). We found that *G. sulfurreducens* PCA and *G. metallireducens* GS-15 are very closely related. The species *G. uraniireducens* Rf4, *G. bemidjensis* Bem, and *G. lovleyi* SZ are separated by a clade.

A tree was also constructed based on the acetate kinase protein NP_953752.1 which is involved in acetate metabolism (Figure 2). Acetate is a primary electron donor for *G. sulfurreducens* PCA, and the anode electrode acts as an electron acceptor [24]. When *G. sulfurreducens* is grown with acetate as a substrate, the enzyme acetate kinase phosphorylates

TABLE 3: Genome data comparison of different *Geobacter* species. Comparing genome characteristics like protein coding genes, GC content, and coding percentage shows the relationship between these bacteria.

Name	RefSeq	Length	No. of genes	Protein coding	Structural RNAs	GC content	% coding
<i>Geobacter sulfurreducens</i> PCA	NC_002939	3,814,139 nt	3505	3428	57	60%	89%
<i>Geobacter metallireducens</i> GS-15	NC_007517	3,997,420 nt	3620	3520	59	59%	90%
<i>Geobacter uraniireducens</i> Rf4	NC_009483	5,136,364 nt	4490	4358	58	54%	86%
<i>Geobacter</i> sp. FRC-32	NC_011979	4,304,501 nt	3901	3798	62	53%	89%
<i>Geobacter bemidjiensis</i> Bem	NC_011146	4,615,150 nt	4129	4057	71	60%	87%
<i>Geobacter lovleyi</i> SZ	NC_010814	3,917,761 nt	3697	3606	54	54%	91%
<i>Geobacter</i> sp. M21	NC_012918	4,745,806 nt	4227	4080	75	60%	86%

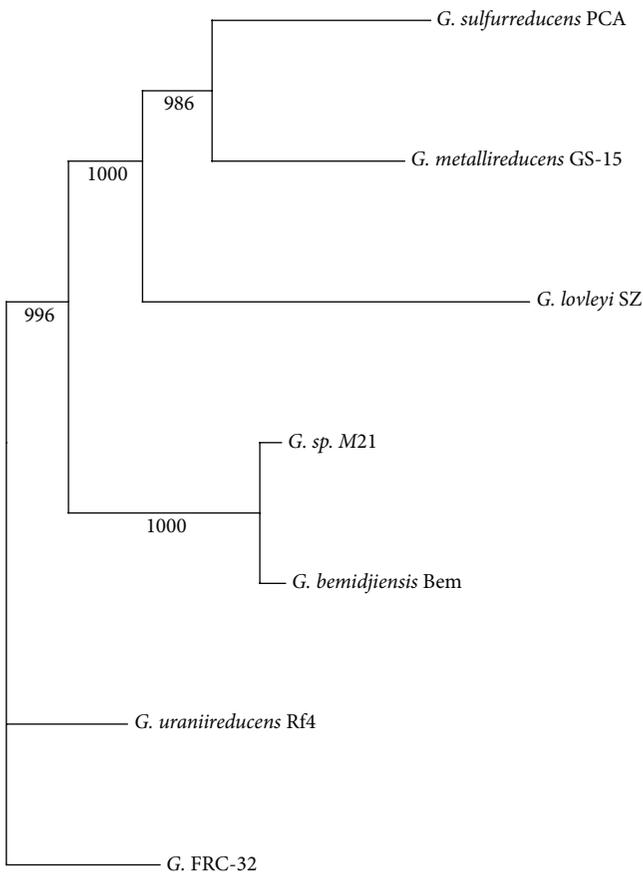


FIGURE 1: Phylogenetic tree constructed using 16S rRNA of *Geobacter* species. This shows the relationship between *Geobacter* species. *G. sulfurreducens* PCA is found to be closely related to *G. metallireducens*.

acetate to acetyl phosphate which is then acted upon by phosphate acetyltransferase and converts it into acetyl-CoA. This acetyl-CoA enters TCA cycle where it is oxidized through the TCA cycle producing NADH, NADPH, ATP, and reduced ferredoxin. Acetate kinase phosphorylation is an important step in energy production. Thus acetate kinase first enzyme which acts on acetate is taken for evolutionary studies. Apart from its closeness to *Geobacter* species, acetate

kinase protein is found to be closely related to *Pelobacter propionicus*, *Pelobacter carbinolicus*, and Deferribacteraceae family bacteria like *Calditerrivibrio nitroreducens*, *Denitrovibrio acetiphilus*, *Flexistipes sinusarabici*, and *Desulfovibrio desulfuricans*. The closeness to these bacteria indicates its evolvability, and these bacteria have developed different strategies to oxidize organic compounds. *Pelobacter* is a nonmotile bacterium indicating its loss in its motility towards evolution or in other ways. *Geobacter* and *Pelobacter* evolved differently from their ancestors. Apart from *Pelobacter*, these organisms use organic substance for anaerobic respiration. *Desulfovibrio desulfuricans*, a known bacterium found to be involved in bioremediation, is found to form a close linkage with Spirochaetaceae family bacteria *Treponema primitia* ZAS-2. This bacterium is found to possess a large flagellar motor [25], and these findings will be helpful in studying the evolution of *Geobacter sulfurreducens* PCA.

4. Conclusion

As the continuous increase in genomic data of *Geobacter* bacteria is available, comparative and functional analysis becomes essential. As these *Geobacter* spp. play a very important role in bioremediation and various environmentally important activities, they need to be analyzed carefully to identify the various aspects of their genome. Comparative analysis is one of the important methods in analyzing the sequence data obtained from databases. A pairwise comparison of the whole genome of six *Geobacter* species with *G. sulfurreducens* yielded homologous sequences between them and also genes specific to particular organism. This comparative analysis also helps in systems biology to fill the holes in pathways. The core genome of *Geobacter* was also identified and classified based on the pathway information. Based on the conservancy of genes, the functions of 74 hypothetical genes were predicted. The relationship between these *Geobacter* species was studied using 16S rRNA. *G. sulfurreducens* PCA is found to be very closely related to *G. metallireducens* GS-15 and distantly related to *G. lovleyi* SZ among the *Geobacter* species. And its evolutionary relationship is studied by constructing a tree for acetate kinase protein. And it is found to be closely related to Deferribacteraceae and Spirochaetaceae families which have

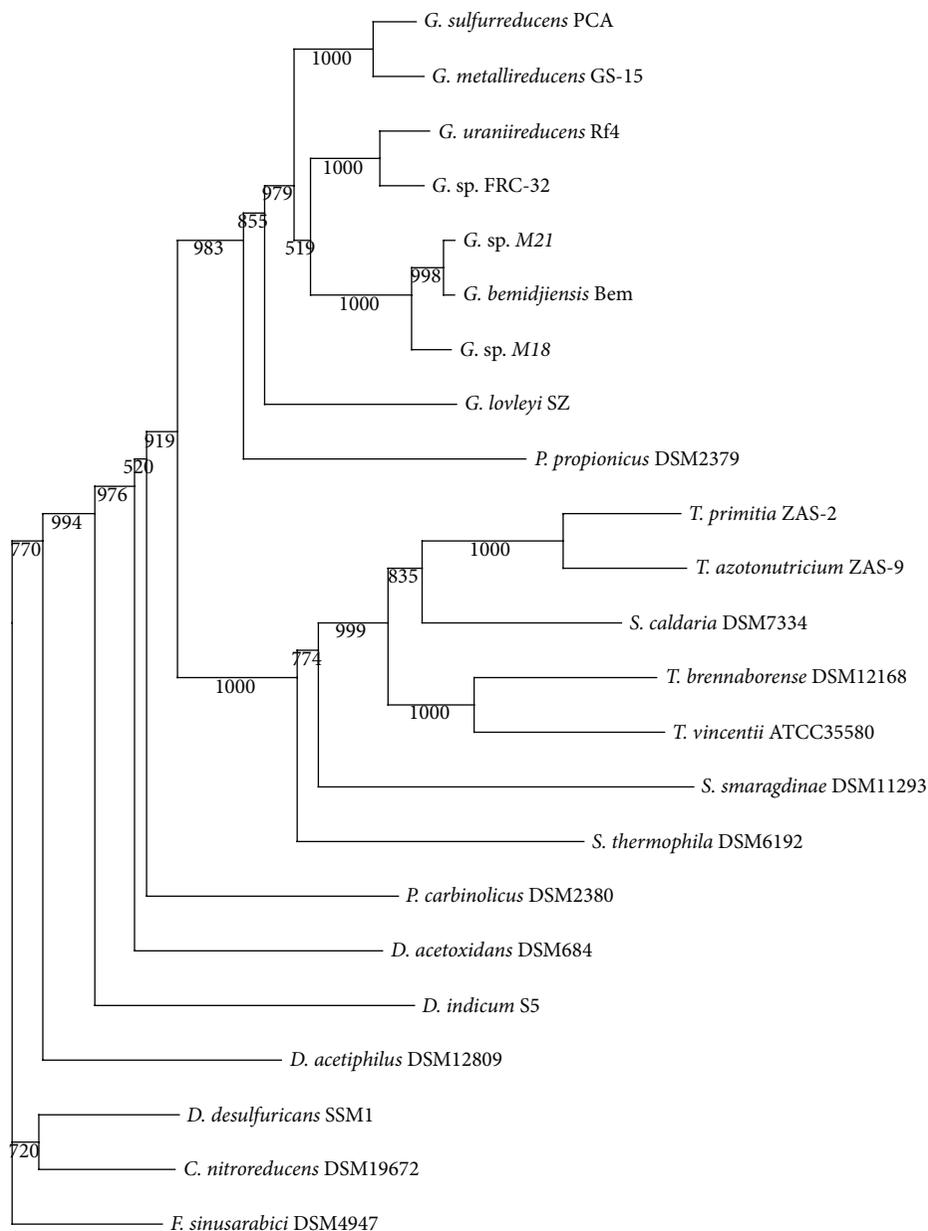


FIGURE 2: A neighbor-joining tree for acetate kinase protein conserved across *Geobacter* species. This tree shows the relationship between various families of bacteria and Geobacteraceae family. The Deferribacteraceae and Spirochaetaceae families are found to be closely related to Geobacteraceae family.

adopted different strategies to oxidize organic compounds. This evolutionary study indicates that these bacteria ancestors had organic compounds as substrate. These results will be useful in finding similarities and differences in the seven *Geobacter spp.* taken for the study and it helps in better understanding the genes involved in electricity production and bioremediation process.

Conflict of Interests

The authors declare that they do not have any conflict of interests.

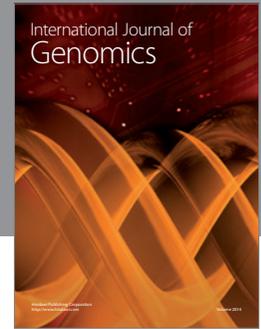
Acknowledgment

The authors would like to thank Dr. Venil N. Sumantran, Adjunct Faculty IITM, for giving valuable suggestions in writing this paper.

References

- [1] C. Leang, L. A. Adams, K.-J. Chin et al., "Adaptation to disruption of the electron transfer pathway for Fe(III) reduction in *Geobacter sulfurreducens*," *Journal of Bacteriology*, vol. 187, no. 17, pp. 5918–5926, 2005.

- [2] D. R. Bond and D. R. Lovley, "Electricity production by *Geobacter sulfurreducens* attached to electrodes," *Applied and Environmental Microbiology*, vol. 69, no. 3, pp. 1548–1555, 2003.
- [3] B. A. Methé, K. E. Nelson, J. A. Eisen et al., "Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments," *Science*, vol. 302, no. 5652, pp. 1967–1969, 2003.
- [4] R. T. Anderson, H. A. Vronis, I. Ortiz-Bernad et al., "Stimulating the in situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer," *Applied and Environmental Microbiology*, vol. 69, no. 10, pp. 5884–5889, 2003.
- [5] A. R. Schoen, "Carbon fiber electrode as an electron acceptor for a microbial fuel cell using geobacter," *Cantaurus*, vol. 15, pp. 24–26, 2007.
- [6] W. C. Lin, M. V. Coppi, and D. R. Lovley, "*Geobacter sulfurreducens* can grow with oxygen as a terminal electron acceptor," *Applied and Environmental Microbiology*, vol. 70, no. 4, pp. 2525–2528, 2004.
- [7] W. M. Fitch, "Homology personal view on some of the problems," *Trends in Genetics*, vol. 16, no. 5, pp. 227–231, 2000.
- [8] M. Richter, M. Kube, D. A. Bazylinski et al., "Comparative genome analysis of four magnetotactic bacteria reveals a complex set of group-specific genes implicated in magnetosome biomineralization and function," *Journal of Bacteriology*, vol. 189, no. 13, pp. 4899–4910, 2007.
- [9] L. Wei, Y. Liu, I. Dubchak, J. Shon, and J. Park, "Comparative genomics approaches to study organism similarities and differences," *Journal of Biomedical Informatics*, vol. 35, no. 2, pp. 142–150, 2002.
- [10] G. Sawa, J. Dicks, and I. N. Roberts, "Current approaches to whole genome phylogenetic analysis," *Briefings in Bioinformatics*, vol. 4, no. 1, pp. 63–74, 2003.
- [11] J. E. Butler, N. D. Young, and D. R. Lovley, "Evolution from a respiratory ancestor to fill syntrophic and fermentative niches: comparative genomics of six *Geobacteraceae* species," *BMC Genomics*, vol. 10, article 103, 2009.
- [12] E. W. Sayers, T. Barrett, D. A. Benson et al., "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 39, pp. 38–51, 2011.
- [13] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
- [14] UniProt Consortium, "Ongoing and future developments at the Universal Protein Resource," *Nucleic Acids Research*, vol. 39, pp. D214–D219, 2011.
- [15] M. A. Larkin, G. Blackshields, N. P. Brown et al., "Clustal W and clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.
- [16] J. Felsenstein, "PHYMLIP—phylogeny inference package (version 3.2)," *Cladistics*, vol. 5, pp. 164–166, 1989.
- [17] M. V. Han and C. M. Zmasek, "PhyloXML: XML for evolutionary biology and comparative genomics," *BMC Bioinformatics*, vol. 10, article 356, 2009.
- [18] D. Lee, O. Redfern, and C. Orengo, "Predicting protein function from sequence and structure," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 12, pp. 995–1005, 2007.
- [19] S. J. Callister, L. A. McCue, J. E. Turse et al., "Comparative bacterial proteomics: analysis of the core genome concept," *PLoS ONE*, vol. 3, no. 2, Article ID e1542, 2008.
- [20] D. W. Ussery and P. F. Hallin, "Genome update: length distributions of sequenced prokaryotic genomes," *Microbiology*, vol. 150, no. 3, pp. 513–516, 2004.
- [21] E. S. Shelobolina, H. A. Vronis, R. H. Findlay, and D. R. Lovley, "*Geobacter uraniireducens* sp. nov., isolated from subsurface sediment undergoing uranium bioremediation," *International Journal of Systematic and Evolutionary Microbiology*, vol. 58, no. 5, pp. 1075–1078, 2008.
- [22] F. Hildebrand, A. Meyer, and A. Eyre-Walker, "Evidence of selection upon genomic GC-content in bacteria," *PLoS Genetics*, vol. 6, no. 9, Article ID e1001107, 2010.
- [23] J. E. Clarridge III, "Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases," *Clinical Microbiology Reviews*, vol. 17, no. 4, pp. 840–862, 2004.
- [24] N. T. Trinh, J. H. Park, and B. Kim, "Increased generation of electricity in a microbial fuel cell using *Geobacter sulfurreducens*," *Korean Journal of Chemical Engineering*, vol. 26, no. 3, pp. 748–753, 2009.
- [25] G. E. Murphy, J. R. Leadbetter, and G. J. Jensen, "In situ structure of the complete *Treponema primitia* flagellar motor," *Nature*, vol. 442, no. 7106, pp. 1062–1064, 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

