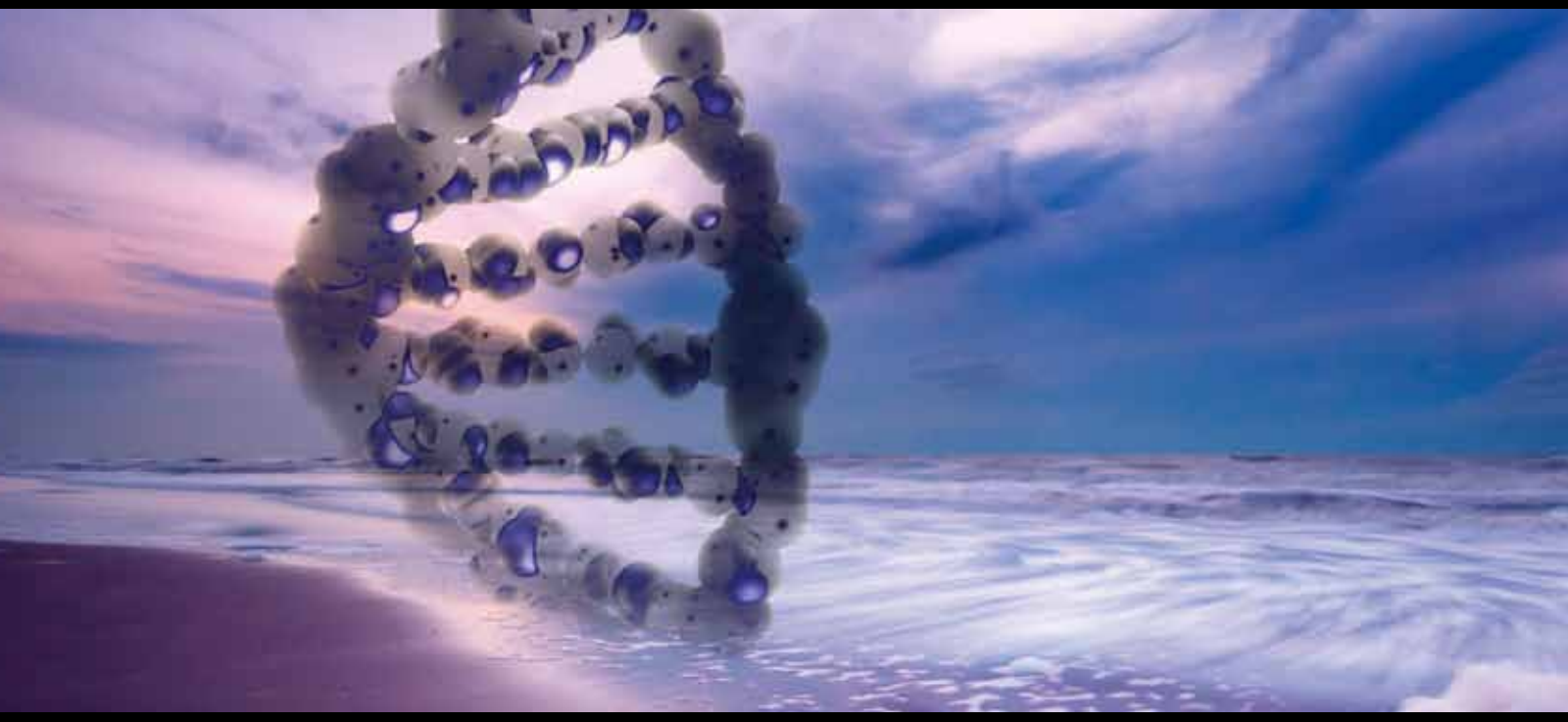
International Journal of Evolutionary Biology

Evolutionary Mechanisms of Microbial Genomes 2012

Guest Editors: Hiromi Nishida, Shinji Kondo, Hideaki Nojiri, Ken-ichi Noma, and Kenro Oshima



Evolutionary Mechanisms of Microbial Genomes 2012

International Journal of Evolutionary Biology

Evolutionary Mechanisms of Microbial Genomes 2012

Guest Editors: Hiromi Nishida, Shinji Kondo, Hideaki Nojiri, Ken-ichi Noma, and Kenro Oshima

Editorial Board

Contents

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 872768, 2 pages
doi:10.1155/2012/872768

Editorial

Evolutionary Mechanisms of Microbial Genomes 2012

Hiromi Nishida,[1] Shinji Kondo,[2] Hideaki Nojiri,[3] Ken-ichi Noma,[4] and Kenro Oshima[5]

[1] *Agricultural Bioinformatics Research Unit, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan*
[2] *Laboratory for Cellular Systems Modeling, RIKEN Research Center for Allergy and Immunology, Kanagawa 230-0045, Japan*
[3] *Laboratory of Environmental Biochemistry, Biotechnology Research Center, The University of Tokyo, Tokyo 113-8657, Japan*
[4] *Gene Expression and Regulation Program, The Wistar Institute, 3601 Spruce Street, Room 116, Philadelphia, PA 19104, USA*
[5] *Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan*

Correspondence should be addressed to Hiromi Nishida, hnishida@iu.a.u-tokyo.ac.jp

Received 11 July 2012; Accepted 11 July 2012

What is the driving force in the course of microbial genome evolution? What is the mechanism for distinguishing self-genome from others? These fundamental questions remain elusive although rigorous studies are underway by using comparative genomics. The special issue "*Evolutionary mechanisms of microbial genomes*" has been launched in 2011 and presented 11 original papers. Here, this new version in 2012 presents 10 papers (one review and nine research articles).

Two papers are presented in phylogenomics. K. Oshima et al. revealed a close relationship of Aquificales to Thermotogales based on the whole-genome comparison in "*Phylogenetic position of Aquificales based on the whole genome sequences of six Aquificales species*." An extensive and elaborate review of fish pathogenic bacteria has been presented by P. S. Sudheesh et al. in "*Comparative pathogenomics of bacteria causing infectious diseases in fish*."

Two papers are presented on subjects related to evolution of base composition in genomes. H. Nishida et al. in "*Genome signature difference between Deinococcus radiodurans and Thermus thermophilus*" observed distinct tetranucleotide frequencies between the genomes of *D. radiodurans* and *T. thermophilus*, potentially reflecting different evolutionary backgrounds of the two species after divergence from common ancestor. H. Nishida in "*Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids*" reported lower GC content (by up to ~10%) of plasmids compared to their host chromosomes and higher correlation of GC content and chromosome size in bacteria than in archaea.

Two papers are presented about horizontal gene transfer in genome evolution. M. Jalasvuori in "*Vehicles, replicators, and intercellular movement of genetic information: Evolutionary dissection of a bacterial cell*" discussed a hypothesis that any given biosphere comprising prokaryotic cell vehicles and genetic replicators may naturally evolve toward possessing horizontally moving replicators of various types. V. S. Pylro et al. described horizontal gene transfer events of the gene *dszC* involved in the cleavage of carbon-sulfur bonds in "*Detection of horizontal gene transfers from phylogenetic comparisons*."

An article about DNA mutation is presented by Y. Shiwa et al. in "*Whole-genome profiling of a novel mutagenesis technique using proofreading-deficient DNA polymerase δ*." They compared mutations created by the chemical mutagen ethyl methanesulfonate (EMS) and the proofreading-deficient DNA polymerase δ and found that the mutations created by the proofreading-deficient DNA polymerase δ generated more diverse amino acid substitution patterns than those by EMS.

Three papers are presented on subjects related to metabolic pathway. H. Nishida in "*Comparative analyses of homocitrate synthase genes of ascomycetous yeasts*" described gene duplications of the homocitrate synthase which have occurred multiple times during evolution of the ascomycetous yeasts. H. Nishida and M. Nishiyama in "*Evolution*

of lysine biosynthesis in the phylum Deinococcus-Thermus" reported that bacterial lysine biosynthesis genes of the common ancestor of the *Deinococcus-Thermus* phylum used the α-aminoadipate pathway instead of the diaminopimelate pathway. K. Ueda et al. in "*Dispensabilities of carbonic anhydrase in Proteobacteria*" analyzed the distribution of carbonic anhydrase (CA) in proteobacteria, compared CA-retaining and CA-deficient genomes, and found absence of coding sequence in some strains and frame shifts in others.

In closing this introduction to the special issue, we would like to express our full appreciation to all the authors and reviewers for their enormous efforts that have made the timely completion of our assignment successful. We sincerely hope that this special issue will stimulate further the investigation of evolutionary mechanisms of microbial genomes.

Hiromi Nishida
Shinji Kondo
Hideaki Nojiri
Ken-ichi Noma
Kenro Oshima

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 859264, 9 pages
doi:10.1155/2012/859264

Research Article

Phylogenetic Position of Aquificales Based on the Whole Genome Sequences of Six Aquificales Species

Kenro Oshima,[1] Yoko Chiba,[2] Yasuo Igarashi,[2] Hiroyuki Arai,[2] and Masaharu Ishii[2]

[1] *Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*
[2] *Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*

Correspondence should be addressed to Kenro Oshima, kenro@ims.u-tokyo.ac.jp

Received 10 February 2012; Revised 17 April 2012; Accepted 18 April 2012

Academic Editor: Hiromi Nishida

Species belonging to the order Aquificales are believed to be an early branching lineage within the Bacteria. However, the branching order of this group in single-gene phylogenetic trees is highly variable; for example, it has also been proposed that the Aquificales should be grouped with ε-proteobacteria. To investigate the phylogenetic position of Aquificales at the whole-genome level, here we reconstructed the phylogenetic trees of 18 bacteria including six Aquificales species based on the concatenated data of proteins shared by these bacteria. In the phylogenetic tree based on the whole-genome information, Aquificales was more closely related to Thermotogales than to Proteobacteria, suggesting that the Aquificales is a relatively early branching lineage within the Bacteria. Moreover, we classified the phylogenetic tree of each conserved orthologous protein by its topology. As a result, in the most major type of the phylogenetic trees, Aquificales was closely related to the Thermotogales. However, Aquificales was closely related to ε-proteobacteria in 21.0% of all phylogenetic trees, suggesting that many proteins phylogenetically related to the ε-proteobacteria may be encoded in the genomes of the members of the Aquificales. This unique feature may be responsible for the high variability in the branching order of Aquificales in single-gene phylogenetic trees.

1. Introduction

Species belonging to the order Aquificales are non-spore-forming, Gram-negative rods that are strictly thermophilic with optimal growth usually occurring above 65°C [1–3]. In terms of metabolism, most species of Aquificales are hydrogen-oxidizing bacteria that utilize hydrogen as the sole electron donor and oxygen as the electron acceptor [4]. Alternatively, thiosulfate or sulfur can also be used as a source of energy. Because of their thermostability, many enzymes found in this group are of interest for industrial and biotechnological applications [5].

Presently, the Aquificales species are believed to be the earliest branching lineage within the Bacteria [6–10]. However, the branching order of this group in single-gene phylogenetic trees is highly variable, and the deep branching of Aquificales is not supported by many protein phylogenies. For example, *Aquifex* has been shown to be close to ε-proteobacteria [11, 12] or the Chlamydiae group [6] in many protein phylogenies. Conserved inserts and deletions in a number of different proteins also provide evidence that the Aquificales is a late branching group within the Bacteria [13]. Many of these analyses suggest that Aquificales might be more closely related to Proteobacteria than to Thermotogales. Therefore, it is important to understand the phylogenetic position of Aquificales within the bacterial phylogeny.

The phylogenetic tree based on 16S rRNA sequences provides the presently accepted framework for understanding the evolutionary relationships among bacteria [14]. However, phylogenetic analysis at the single-gene level may provide only a limited understanding of the relationships and evolutionary history of bacteria, especially the closely related species that diverged at almost the same time [15]. In addition, species phylogenies derived from comparisons of different genes do not always concur, which may be attributed to lateral gene transfer [16], saturation with respect to amino acid substitutions [17], or highly variable

rates of evolution of individual genes [18, 19]. Therefore, it is believed that comparative studies based on the complete sequences of bacterial genomes should form the basis for phylogeny and, ultimately, taxonomy [20].

The phylogenies inferred from concatenated data of housekeeping proteins amplified the resolving power for delineating the phylogenetic relationships among prokaryotes [21–23]. The complete genome of *Aquifex aeolicus* was sequenced in 1998 [3], and the genomes of five Aquificales species (*Hydrogenobacter thermophilus* TK-6, *Hydrogenobaculum* sp. Y04AAS1, *Persephonella marina*, *Sulfurihydrogenibium azorense* and *Sulfurihydrogenibium* sp. YO3AOP1) have recently been sequenced [24, 25]. Here we reconstructed the phylogenetic trees of 18 bacteria including six Aquificales bacteria based on the concatenated data of proteins shared by these bacteria. Moreover, the phylogenetic relationship between Aquificales and ε-proteobaceria was analyzed at the whole genome level.

2. Materials and Methods

In this study, we used 18 genome sequences from *Aquifex aeolicus*, *Hydrogenobacter thermophilus* TK-6, *Hydrogenobaculum* sp. Y04AAS1, *Persephonella marina*, *Sulfurihydrogenibium azorense*, *Sulfurihydrogenibium* sp. YO3AOP1, *Bacillus subtilis* subsp. *subtilis* str. 168, *Burkholderia mallei* ATCC 23344, *Campylobacter jejuni* subsp. *jejuni* NCTC 11168, *Chlamydophila pneumoniae* CWL029, *Deinococcus radiodurans* R1, *Thermus thermophilus* HB8, *Escherichia coli* str. K-12 substr. MG1655, *Salmonella enterica* subsp. *enterica* serovar Typhimurium LT2, *Helicobacter pylori* 26695, *Pyrococcus horikoshii* OT3, *Thermotoga maritima* MSB8 and *Thermotoga petrophila* RKU-1. These genome sequences was obtained from GenomeNet (http://www.genome.jp/).

First, BLASTP searches (each protein encoded in the genome of *Hydrogenobacter thermophilus* TK-6 was used as a query) were performed against 18 whole genomes by using stand-alone BLAST program [26]. If 18 different proteins from all 18 bacteria occupied the top 18 proteins of the result of the BLAST search, additional BLASTP searches were performed against 18 whole genomes by using each of top 18 proteins as a query. If the top 18 proteins in all 18 BLAST search are the same, we defined these 18 proteins as a conserved orthologous protein. This procedure enabled us to define 62 sets of orthologous proteins from the 18 genomes in our study. (see Supplementary Table 1 in supplementary material available online at doi:10.1155/2012/859264).

Next, we constructed 62 multiple-alignments using MUSCLE [27]. After that, a concatenated multiple alignment inferred from the 62 multiple alignments was generated. The concatenated alignment had 31,542 amino acid sites, including 15,442 gap/insertion sites that were not considered in this analysis. To avoid a potential cause for long branch attraction, we removed the most saturated sites from the whole multiple alignments according to the previously described method by Boussau et al. [28] as follows. First, PhyML [29] was used to build a starting phylogeny based on the whole multiple alignments, using the JTT model and a gamma law discretized in four classes to account for variation in the evolutionary rates. Second, to estimate how sites were modeled by the discretized gamma law, we plotted the distribution of expected relative evolutionary rates across sites as found by BppML (Supplementary Figure 1). Third, to reduce risks of long branch attraction, we decided to discard sites whose evolutionary rate was above the threshold of 2.0 (red line, Supplementary Figure 1). Finally, phylogenetic analyses were performed based on 10,000 amino acid sites. Based on the multiple alignments, a maximum likelihood (ML) tree was reconstructed using the PhyML [29] based on the JTT model and a gamma law discretized in four classes to account for variation in the evolutionary rates. *Pyrococcus horikoshii* was used as an outgroup. The confidence values (%) were estimated with the bootstrap sampling method (200 replications).

In addition, to reduce the influence of compositional bias, we recoded the alignment without saturated sites in 4 states based on the physicochemical properties of the amino acids [28] as follows: aromatic (FWY) and hydrophobic (MILV) amino acids were grouped in a single state, basic amino-acids (HKR) in another, acidic (DENQ) amino acids in one more state, and the fourth state contained all other amino acids (AGPST) to the exception of cysteine which was coded as missing data. The ML tree was constructed with this recoded alignment by the GTR model, an estimated proportion of invariant sites, a gamma law discretized in 5 categories with its alpha parameter estimated, and 200 bootstrap replicates [28].

To construct the phylogenetic tree of six Aquificales species, two Thermotogales species, two γ-proteobacteria, and two ε-proteobacteria, we used 12 genome sequences from *Aquifex aeolicus*, *Hydrogenobacter thermophilus* TK-6, *Hydrogenobaculum* sp. Y04AAS1, *Persephonella marina*, *Sulfurihydrogenibium azorense*, *Sulfurihydrogenibium* sp. YO3AOP1, *Campylobacter jejuni* subsp. *jejuni* NCTC 11168, *Deinococcus radiodurans* R1, *Escherichia coli* str. K-12 substr. MG1655, *Salmonella enterica* subsp. *enterica* serovar Typhimurium LT2, and *Helicobacter pylori* 26695.

To construct the phylogenetic tree of Thermales-Deinococcales species, Thermotogales species, γ-proteobacteria, and ε-proteobacteria, we used 8 genome sequences from *Campylobacter jejuni* subsp. *jejuni* NCTC 11168, *Deinococcus radiodurans* R1, *Thermus thermophilus* HB8, *Escherichia coli* str. K-12 substr. MG1655, *Salmonella enterica* subsp. *enterica* serovar Typhimurium LT2, *Helicobacter pylori* 26695, *Thermotoga maritima* MSB8, and *Thermotoga petrophila* RKU-1. The phylogenetic tree based on the concatenate data of the whole conserved orthologous proteins was constructed by the same method as described above. The ML trees of individual proteins were constructed using PhyML [28].

3. Results and Discussion

3.1. Phylogenetic Position of Aquificales Based on Whole-Genome Sequences. First, we constructed ML trees based on the 16S rRNA sequences of 18 bacteria (Figure 1(a)). This

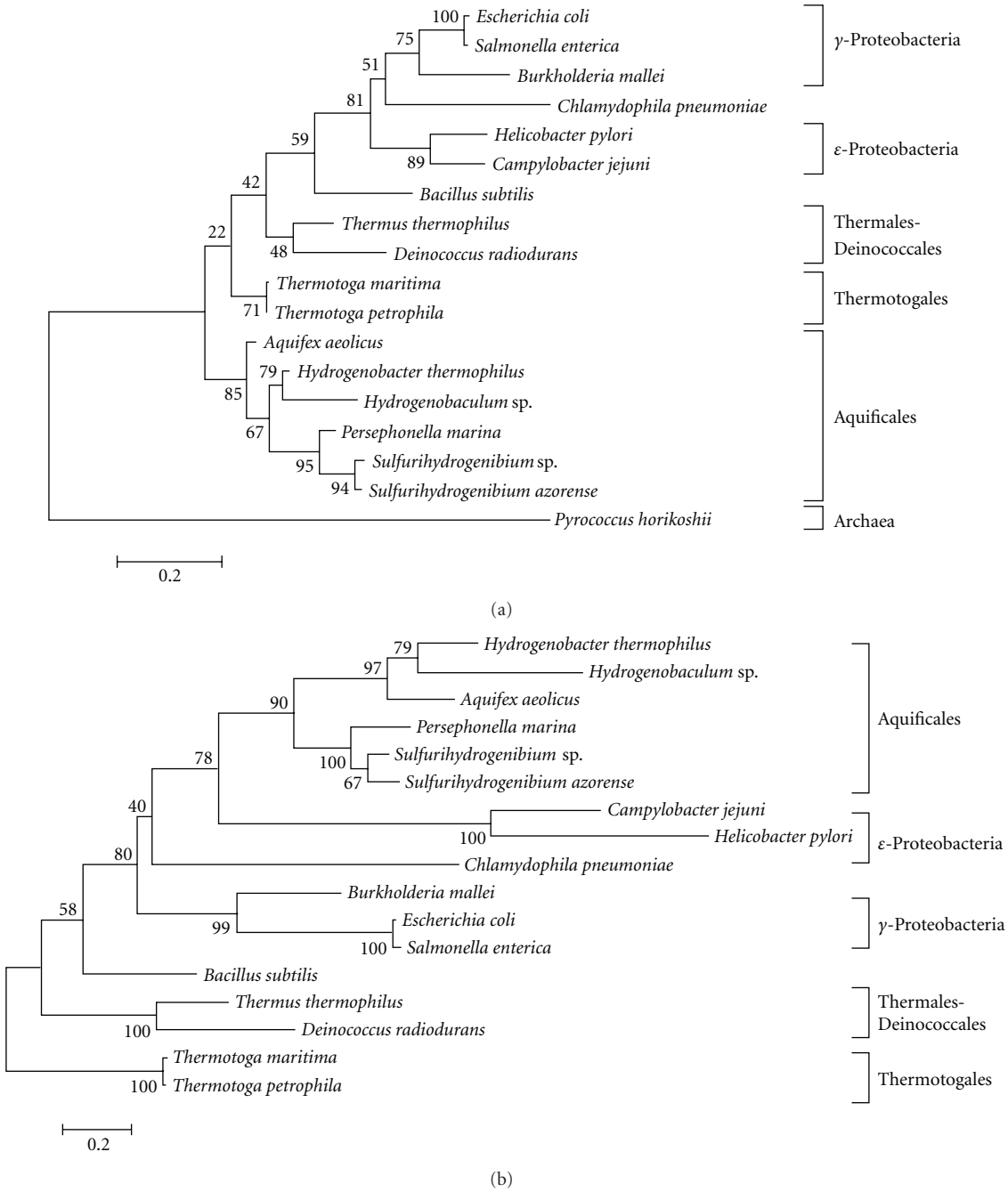


Figure 1: (a) Maximum likelihood tree based on the 16S rRNA sequence comparison. The number at each node represents the percentage in the bootstrap analysis (1000 replicates). (b) Maximum likelihood tree based on the amino acid sequence of the transcription elongation factor. The number at each node represents the percentage in the bootstrap analysis (1000 replicates).

phylogenetic tree indicated that each bacteria belonging to Archaea, Aquificales, Thermales-Deinococcales, and Thermotogales was clustered as a clade. Thermales-Deinococcales was clustered with proteobacteria with 74% bootstrap support. The closest species to Archaea was the bacteria belonging to Thermotogales, and the second nearest neighbor was Aquificales, suggesting that the Aquificales species are an early branching lineage within the Bacteria. In contrast, the topology of the phylogenetic tree based on the amino acid sequences of transcription elongation factor (NusA) (Figure 1(b)) differed from that of the 16S rRNA gene. For example, Aquificales species were clustered with ε-proteobacteria with 80% bootstrap support, suggesting that the Aquificales is a late branching group within the Bacteria.

To investigate the phylogenetic position of Aquificales species at the whole-genome level, we constructed a phylogenetic tree based on 18 whole genomes of Archaea, Aquificales, Thermales-Deinococcales, Thermotogales, and related

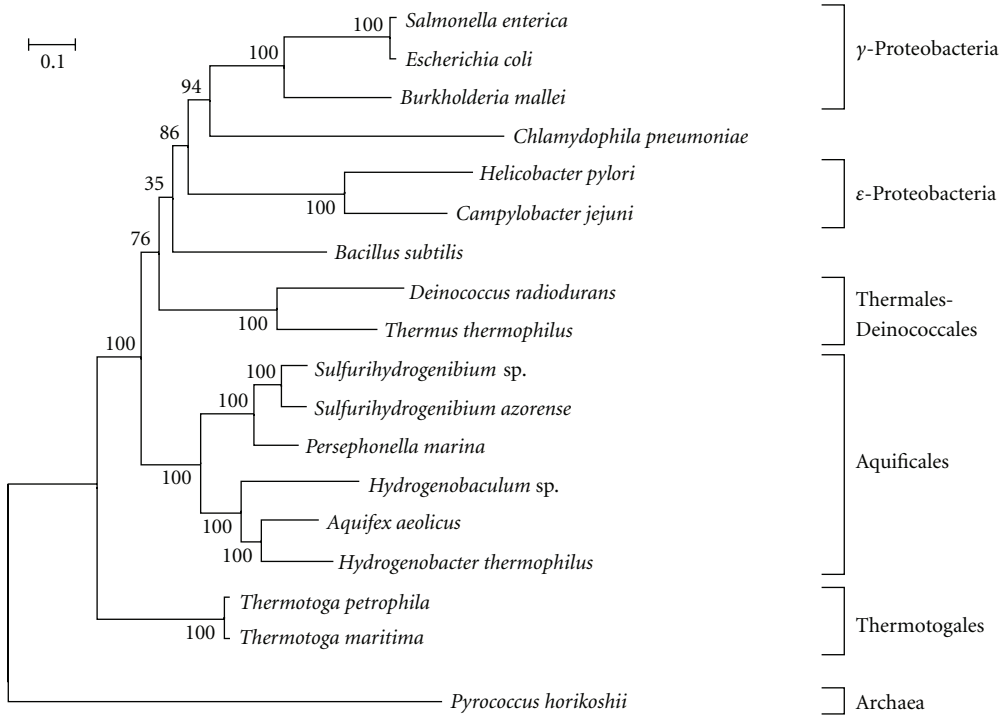


Figure 2: Maximum likelihood tree based on the comparison of 62 proteins; 10,000 amino acid sites were considered (see Section 2). The number at each node represents the percentage in the bootstrap analysis (200 replicates).

bacteria. First, 62 orthologous gene families that are shared by all 18 bacteria were selected (Supplementary Table 1). To avoid a potential cause for long branch attraction, we removed the most saturated sites from the whole multiple alignments according to the previously described method by Boussau et al. [28]. As a result, 10,000 amino acid sites were considered in the maximum likelihood analysis. The phylogenetic tree based on the whole-genome information indicated that the 18 bacteria were divided into six major groups (Archaea, Aquificales, Thermotogales, Thermales-Deinococcales, γ-proteobacteria, and ε-proteobacteria) with 100% bootstrap support. Analysis of signature sequences (consisting of conserved inserts or deletions) in highly conserved proteins suggested that the Aquificales diverged after the branching of Thermotogales, Thermales-Deinococcales, Cyanobacteria, Spirochetes, and Chlamydiae, but before the emergence of Proteobacteria [13]. However, in the phylogenetic tree based on the whole-genome information, the Archaea group was evolutionarily closely related to the Thermotogales, and Aquificales was a neighbor to Thermotogales with 76% bootstrap value (Figure 2). These analyses suggest that Aquificales is more closely related to Thermotogales than to Proteobacteria, which is consistent with the phylogenetic relationship showed by Boussau et al. [28]. To reduce the influence of compositional bias, we recoded the concatenated protein alignment in 4 states based on the physicochemical properties of the amino acids, and constructed a phylogenetic tree. As a result, although *Bacillus subtilis* was clustered with Thermales-Deinococcales, the ML tree obtained by the recoded alignment (Supplemetary Figure 2) was very similar to the previous tree (Figure 2), implying that the Aquificales-Thermotogales grouping does not seem to result from compositional biases. These results suggest that the Aquificales species are a relatively early branching lineage within the Bacteria.

3.2. Phylogenetic Relationships between the Aquificales and ε-Proteobacteria. It has been proposed that the Aquificales should be grouped with the ε-proteobacteria [12], which is supported by the phylogenetic analysis of single protein sequences such as the transcription elongation factor (Figure 1(b)). However, the late branching of the Aquificales is not supported by the 16S rRNA gene sequence tree (Figure 1(a)) and the phylogenetic tree based on the whole-genome information (Figure 2). To investigate the phylogenetic relationships between the Aquificales and ε-proteobacteria, we reconstructed phylogenetic trees of 12 bacteria including six Aquificales species, two Thermotogales species, two γ-proteobacteria, and two ε-proteobacteria based on the concatenated data of proteins shared by these bacteria. First, 271 orthologous gene families that are shared by all 12 bacteria were selected. As a result, 16,532 amino acid sites were considered in the ML analysis. The phylogenetic tree based on this whole-genome information indicated that the 12 bacteria were divided into four major groups (Aquificales, Thermotogales, γ-proteobacteria, and ε-proteobacteria) with 100% bootstrap support (Figure 3(a)).

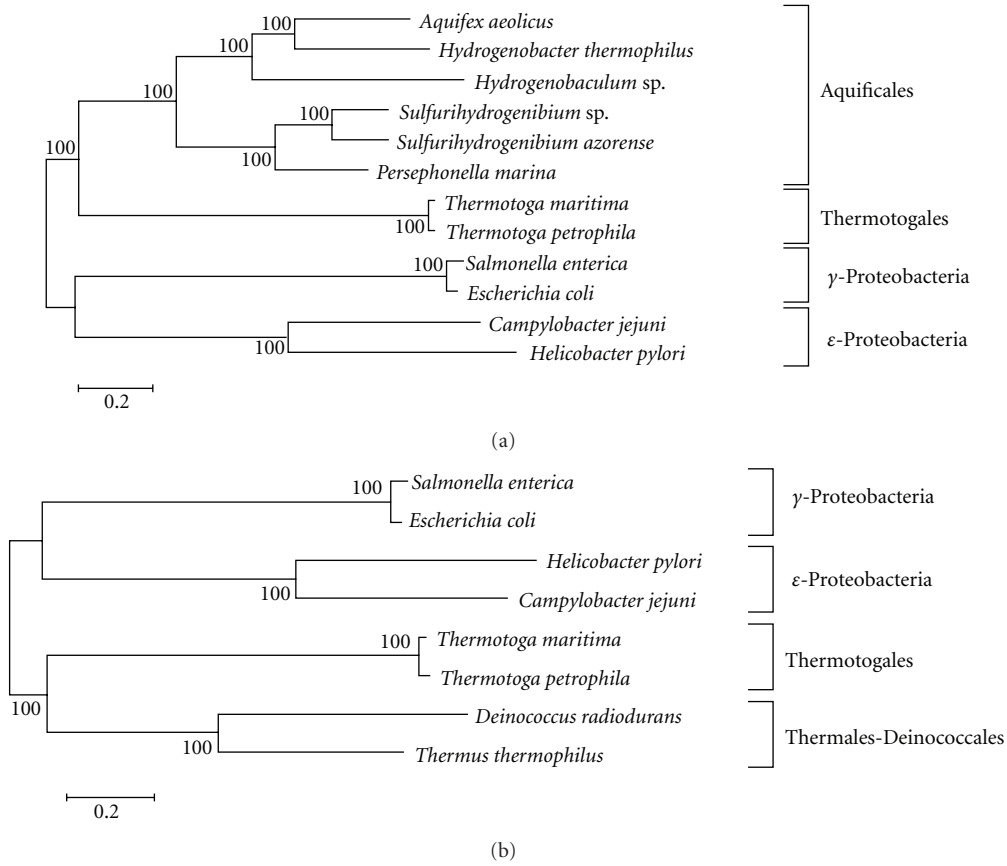


Figure 3: Unrooted maximum likelihood tree based on whole-genome information by using (a) the 271 conserved proteins among Aquificales, Thermotogales, γ-proteobacteria and ε-proteobacteria, or (b) the 259 conserved proteins among Thermales, Thermotogales, γ-proteobacteria and ε-proteobacteria. The number at each node represents the percentage in the bootstrap analysis (200 replicates).

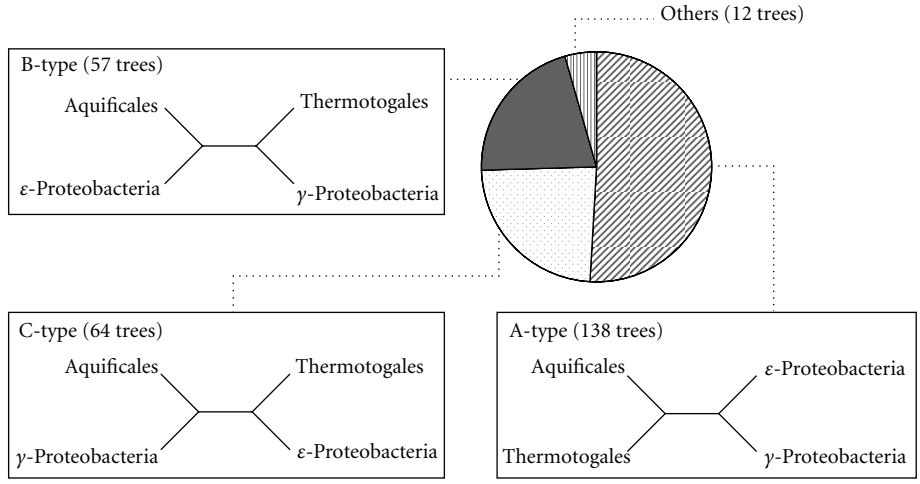


Figure 4: Distribution of topology of the phylogenetic trees of the 271 conserved proteins among Aquificales, Thermotogales, γ-proteobacteria, and ε-proteobacteria.

Table 1: List of B-type conserved proteins that the Aquificales is clustered with ε-proteobacteria in the phylogenetic analysis (Figure 4). Accession numbers of conserved proteins of *Hydrogenobacter thermophilus* TK-6 are indicated.

Accession number	Putative function
YP_003431690	transcription elongation factor
YP_003432239	ribosomal protein S9
YP_003432379	ribosomal protein L18
YP_003432892	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
YP_003432936	ATP-dependent protease
YP_003433556	UDP-N-acetylglucosamine pyrophosphorylase
YP_003431738	putative metalloprotease
YP_003431749	diaminopimelate decarboxylase
YP_003431809	dihydrodipicolinate reductase
YP_003431998	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
YP_003432481	ribosomal protein S20
YP_003432953	queuine tRNA-ribosyltransferase
YP_003431834	ATP-dependent protease La
YP_003431839	tRNA delta(2)-isopentenylpyrophosphate transferase
YP_003431873	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
YP_003431915	ribonuclease III
YP_003432036	riboflavin synthase alpha chain
YP_003432044	DNA polymerase I
YP_003432149	2-methylthioadenine synthetase
YP_003432165	folylpolyglutamate synthase
YP_003432232	DNA polymerase III beta subunit
YP_003432262	UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
YP_003432385	methionine aminopeptidase
YP_003432408	methionyl-tRNA synthetase
YP_003432463	fatty acid/phospholipid synthesis protein
YP_003433015	carboxyl-terminal protease
YP_003433058	rRNA methylase
YP_003433377	3-phosphoshikimate 1-carboxyvinyltransferase
YP_003433542	arginyl-tRNA synthetase
YP_003431843	F0F1-type ATP synthase gamma subunit
YP_003431889	signal recognition particle GTPase
YP_003432507	ribosomal protein L22
YP_003432144	Holliday junction resolvase
YP_003432824	DNA processing protein
YP_003432257	GTP-binding protein
YP_003432274	triosephosphate isomerase
YP_003432330	aspartate 1-decarboxylase
YP_003432353	uridylate kinase
YP_003432374	ribosomal protein L24
YP_003432380	ribosomal protein S5
YP_003432524	transcription antitermination protein
YP_003432640	methionyl-tRNA formyltransferase
YP_003433333	ribosomal protein L20
YP_003432384	adenylate kinase
YP_003432390	ribosomal protein S4
YP_003432414	thiol peroxidase

Table 1: Continued.

Accession number	Putative function
YP_003432533	orotidine 5′-phosphate decarboxylase
YP_003432615	S-adenosyl-methyltransferase
YP_003432911	carbamoyl-phosphate synthase small subunit
YP_003432886	dihydrodipicolinate synthase
YP_003432967	membrane protein
YP_003432968	GMP synthase
YP_003433028	hypothetical protein HTH_1376
YP_003433124	homoserine kinase
YP_003433221	UDP-glucose-4-epimerase
YP_003433380	pantothenate metabolism flavoprotein
YP_003433549	cell cycle protein

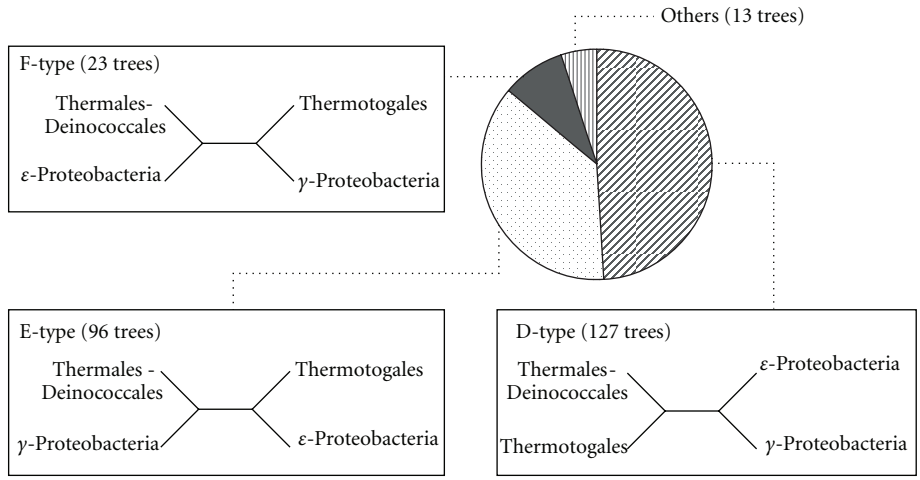


Figure 5: Distribution of topology of the phylogenetic trees of the 259 conserved proteins among Thermales, Thermotogales, γ-proteobacteria and ε-proteobacteria.

In addition, the phylogenetic tree based on the whole-genome information indicated that the Aquificales group was clustered with the Thermotogales group with 100% bootstrap support (Figure 3(a)).

Next, to investigate the contribution of each protein to the whole-genome phylogenetic tree, we constructed 271 ML trees from 271 protein sets. We classified these trees into the following three types (Figure 4): A-type, the Aquificales group is more closely related to the Thermotogales group; B-type, the Aquificales group is more closely related to the ε-proteobacteria group; C-type, the Aquificales group is more closely related to the γ-proteobacteria group. The most frequent type of these phylogenetic trees was A-type (138 trees), which is consistent with the results obtained from the phylogenetic tree based on the 271 conserved proteins (Figure 3(a)). Interestingly, B-type trees occupied 21.0% (57 trees) of all phylogenetic trees (Table 1). For example, the Aquificales was clustered with the ε-proteobacteria with 94% bootstrap support in the phylogenetic tree of DNA polymerase I (Supplementary Figure 3). These results suggest that many proteins phylogenetically related to the ε-proteobacteria may be encoded in the genomes of the members of the Aquificales order.

To compare this profile with that of other bacteria, we performed the same phylogenetic analysis against 259 conserved proteins among Thermales-Deinococcales, Thermotogales, γ-proteobacteria, and ε-proteobacteria. As a result, the Thermales-Deinococcales group was clustered with the Thermotogales group with 100% bootstrap support in the phylogenetic tree based on the whole genome conserved proteins (Figure 3(b)). Next, we classified the phylogenetic tree of each protein into the following three types (Figure 5); D-type, the Thermales-Deinococcales group is more closely related to the Thermotogales group; E-type, the Thermales-Deinococcales group is more closely related to the γ-proteobacteria group; F-type, the Thermales-Deinococcales group is more closely related to the ε-proteobacteria group. The most frequent type of these phylogenetic trees was D-type (127 trees), and E-type trees occupied 37.1% (96 trees) of all phylogenetic trees (Figure 5). In contrast, F-type trees occupied only 8.9% (23 trees) of all phylogenetic trees (Figure 5), suggesting that the phylogenetic relationship

between the Thermales-Deinococcales and ε-proteobacteria may be low compared to the Aquificales.

These results support the hypothesis that many proteins phylogenetically close to the ε-proteobacteria may be encoded in the genomes of the Aquificales. This unique feature may be responsible for the high variability in the branching order of Aquificales in single-gene phylogenetic trees. Moreover, these results raised the possibility that a large horizontal gene transfer had occurred between the Aquificales and ε-proteobacteria, which was suggested by Boussau et al. [28]. This hypothesis might be supported by the fact that ε-proteobacteria include hydrogen-oxidizing bacteria and sulfur-oxidizing bacteria [30] which occupy the same ecological niche with Aquificales.

Several house-keeping proteins have often been used for the phylogenetic analyses of bacteria [31–33]. However, our results suggest that the phylogenetic position of single proteins is highly variable even for transcription elongation factor and DNA polymerase I. Therefore, whole-genome level phylogenetic approaches are extremely important and will possibly play a crucial role in the future studies of microbial evolution.

Acknowledgments

The authors thank anonymous reviewer for valuable comments. This work was supported in part by a Grant-in-Aid for scientific research (A) (21248010) from the Japan Society for the Promotion of Science (JSPS), a Grant-in-Aid for JSPS fellows (23-3030), and a funding program for Next Generation World-Leading Researchers (GS005).

References

[1] W. Eder and R. Huber, "New isolates and physiological properties of the Aquificales and description of Thermocrinis albus sp. nov," *Extremophiles*, vol. 6, no. 4, pp. 309–318, 2002.

[2] T. Kawasumi, Y. Igarashi, T. Kodama, and Y. Minoda, "Hydrogenobacter thermophilus gen. nov., sp. nov., an extremely thermophilic, aerobic, hydrogen-oxidizing bacterium," *International Journal of Systematic Bacteriology*, vol. 34, no. 1, pp. 5–10, 1984.

[3] G. Deckert, P. V. Warren, T. Gaasterland et al., "The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*," *Nature*, vol. 392, no. 6674, pp. 353–358, 1998.

[4] R. Huber and W. Eder, "Aquificales," in *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community. Release 3.8*, M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, Eds., Springer, New York, NY, USA, 2002.

[5] H. Huber and K. O. Stetter, "Hyperthermophiles and their possible potential in biotechnology," *Journal of Biotechnology*, vol. 64, no. 1, pp. 39–52, 1998.

[6] S. Burggraf, G. J. Olsen, K. O. Stetter, and C. R. Woese, "A phylogenetic analysis of *Aquifex pyrophilus*," *Systematic and Applied Microbiology*, vol. 15, no. 3, pp. 352–356, 1992.

[7] G. J. Olsen, C. R. Woese, and R. Overbeek, "The winds of (evolutionary) change: breathing new life into microbiology," *Journal of Bacteriology*, vol. 176, no. 1, pp. 1–6, 1994.

[8] M. Bocchetta, S. Gribaldo, A. Sanangelantoni, and P. Cammarano, "Phylogenetic depth of the bacterial genera *Aquifex* and *Thermotoga* inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences," *Journal of Molecular Evolution*, vol. 50, no. 4, pp. 366–380, 2000.

[9] J. R. Brown and W. F. Doolittle, "Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 7, pp. 2441–2445, 1995.

[10] C. Pitulle, Y. Yang, M. Marchiani et al., "Phylogenetic position of the genus *Hydrogenobacter*," *International Journal of Systematic Bacteriology*, vol. 44, no. 4, pp. 620–626, 1994.

[11] E. J. Deeds, H. Hennessey, and E. I. Shakhnovich, "Prokaryotic phylogenies inferred from protein structural domains," *Genome Research*, vol. 15, no. 3, pp. 393–402, 2005.

[12] T. Cavalier-Smith, "The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification," *International Journal of Systematic and Evolutionary Microbiology*, vol. 52, no. 1, pp. 7–76, 2002.

[13] E. Griffiths and R. S. Gupta, "Signature sequences in diverse proteins provide evidence for the late divergence of the Order Aquificales," *International Microbiology*, vol. 7, no. 1, pp. 41–52, 2004.

[14] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 12, pp. 4576–4579, 1990.

[15] G. E. Fox, J. D. Wisotzkey, and P. Jurtshuk, "How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity," *International Journal of Systematic Bacteriology*, vol. 42, no. 1, pp. 166–170, 1992.

[16] P. Forterre and H. Philippe, "Where is the root of the universal tree of life?" *BioEssays*, vol. 21, no. 10, pp. 871–879, 1999.

[17] W. F. Doolittle, "Phylogenetic classification and the universal tree," *Science*, vol. 284, no. 5423, pp. 2124–2128, 1999.

[18] M. Hasegawa, H. Kishino, and T. A. Yano, "Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea," *Journal of Human Evolution*, vol. 18, no. 5, pp. 461–476, 1989.

[19] M. A. Huynen and P. Bork, "Measuring genome evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 11, pp. 5849–5856, 1998.

[20] S. Razin, D. Yogev, and Y. Naot, "Molecular biology and pathogenicity of mycoplasmas," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 4, pp. 1094–1156, 1998.

[21] J. R. Brown, C. J. Douady, M. J. Italia, W. E. Marshall, and M. J. Stanhope, "Universal trees based on large combined protein sequence data sets," *Nature Genetics*, vol. 28, no. 3, pp. 281–285, 2001.

[22] K. Oshima and H. Nishida, "Phylogenetic relationships among mycoplasmas based on the whole genomic information," *Journal of Molecular Evolution*, vol. 65, no. 3, pp. 249–258, 2007.

[23] K. Oshima and H. Nishida, "Detection of the genes evolving under *Ureaplasma* selection," *Journal of Molecular Evolution*, vol. 66, no. 5, pp. 529–532, 2008.

[24] H. Arai, H. Kanbe, M. Ishii, and Y. Igarashi, "Complete genome sequence of the thermophilic, obligately chemolithoautotrophic hydrogen-oxidizing bacterium *Hydrogenobacter thermophilus* TK-6," *Journal of Bacteriology*, vol. 192, no. 10, pp. 2651–2652, 2010.

[25] A. L. Reysenbach, N. Hamamura, M. Podar et al., "Complete and draft genome sequences of six members of the *Aquificales*,"

Journal of Bacteriology, vol. 191, no. 6, pp. 1992–1993, 2009.

[26] 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.

[27] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[28] B. Boussau, L. Guéguen, and M. Gouy, "Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of aquificales in the phylogeny of Bacteria," *BMC Evolutionary Biology*, vol. 8, no. 1, p. 272, 2008.

[29] S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0," *Systematic Biology*, vol. 59, no. 3, pp. 307–321, 2010.

[30] J. W. Voordeckers, M. H. Do, M. Hügler, V. Ko, S. M. Sievert, and C. Vetriani, "Culture dependent and independent analyses of 16S rRNA and ATP citrate lyase genes: a comparison of microbial communities from different black smoker chimneys on the Mid-Atlantic Ridge," *Extremophiles*, vol. 12, no. 5, pp. 627–640, 2008.

[31] W. Ludwig, O. Strunk, S. Klugbauer et al., "Bacterial phylogeny based on comparative sequence analysis," *Electrophoresis*, vol. 19, no. 4, pp. 554–568, 1998.

[32] M. Falah and R. S. Gupta, "Phylogenetic analysis of mycoplasmas based on Hsp70 sequences: cloning of the dnaK (hsp70) gene region of *Mycoplasma capricolum*," *International Journal of Systematic Bacteriology*, vol. 47, no. 1, pp. 38–45, 1997.

[33] V. Kamla, B. Henrich, and U. Hadding, "Phylogeny based on elongation factor Tu reflects the phenotypic features of mycoplasmas better than that based on 16S rRNA," *Gene*, vol. 171, no. 1, pp. 83–87, 1996.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 813015, 7 pages
doi:10.1155/2012/813015

Research Article

Detection of Horizontal Gene Transfers from Phylogenetic Comparisons

Victor Satler Pylro,[1] Luciano de Souza Vespoli,[2] Gabriela Frois Duarte,[3] and Karla Suemy Clemente Yotoko[4]

[1] *Laboratório de Biodiversidade e Biotecnologia para o Meio Ambiente, Departamento de Microbiologia, Universidade Federal de Viçosa, Avenida P.H. Rolfs, s/n, Viçosa, MG, Brazil*
[2] *Laboratório de Biotecnologia, CCTA, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Avenida Alberto Lamego, 2000, Parque Califórnia, Campos dos Goytacazes, RJ, Brazil*
[3] *Centro Universitário Newton Paiva, Escola de Engenharia e Arquitetura, Rua José Cláudio Rezende, 420, Estoril, CEP 30494-230 Belo Horizonte, MG, Brazil*
[4] *Laboratório de Bioinformática e Evolução, Universidade Federal de Viçosa, Avenida P.H. Rolfs, s/n, Viçosa, MG, Brazil*

Correspondence should be addressed to Victor Satler Pylro, victor.pylro@ufv.br

Received 9 February 2012; Revised 22 March 2012; Accepted 23 March 2012

Academic Editor: Hiromi Nishida

Bacterial phylogenies have become one of the most important challenges for microbial ecology. This field started in the mid-1970s with the aim of using the sequence of the small subunit ribosomal RNA (16S) tool to infer bacterial phylogenies. Phylogenetic hypotheses based on other sequences usually give conflicting topologies that reveal different evolutionary histories, which in some cases may be the result of horizontal gene transfer events. Currently, one of the major goals of molecular biology is to understand the role that horizontal gene transfer plays in species adaptation and evolution. In this work, we compared the phylogenetic tree based on 16S with the tree based on *dsz*C, a gene involved in the cleavage of carbon-sulfur bonds. Bacteria of several genera perform this survival task when living in environments lacking free mineral sulfur. The biochemical pathway of the desulphurization process was extensively studied due to its economic importance, since this step is expensive and indispensable in fuel production. Our results clearly show that horizontal gene transfer events could be detected using common phylogenetic methods with gene sequences obtained from public sequence databases.

1. Introduction

The discussion concerning bacteria phylogenies has become one of the most important aspects of microbial ecology. In the mid-1970s, Woese and his collaborators proposed and began assembling a significant database of sequence information based on small subunit ribosomal RNA (SSU rRNA 16S). The current universal tree is based on this [1–6], since it is easily sequenced (±1,500 nucleotides) and widely available in sequence databases (Gen-Bank, EMBL) [7, 8]. However, phylogenetic hypotheses based on several other genes result in conflicting topologies and reveal different evolutionary histories. In many cases, especially within bacteria, these may be the result of horizontal gene transfers (HGTs) [9, 10], which are regarded as a crucial mechanism of increasing genetic variability among bacteria [11–13]. Currently, one of the major goals of molecular biology is to understand the role that HGT plays in species adaptation and evolution [10, 14, 15]. The presence of HGT in bacteria has been reported for several years, suggesting that for some genes the tree of life becomes a net [16]. HGT is dominant among various groups of genes in prokaryotes such as antibiotic resistance, carbon source utilization, organic contaminant degradation, and freeze tolerance genes [12, 13]. However, there is some evidence of HGT in housekeeping genes such as those for replication, transcription, and translation as well [10, 17, 18].

Sulphur is the third most abundant element in petroleum (after carbon and hydrogen), and its release contributes to air pollution by causing acid rain [19, 20]. For this reason,

sulphur regulations have continued to become more stringent and it is necessary to remove sulphur oxides from fossil fuels during the refining process. Most inorganic and simple organic sulphur can be removed by hydrodesulphurization, the technique currently used by most petroleum refineries, but, in petroleum, the majority of sulphur is found in dibenzothiophene (DBT) and its derivatives, which can only be removed through a specific biological mechanism called biodesulphurization [21]. Several studies have investigated the development of aerobic microbial desulphurisation pathways [22–24].

Some bacteria can desulphurize DBT to 2-hydroxybiphenyl (2-HBP) through the sulphur-specific degradation pathway (4S pathway) without destroying the hydrocarbon skeleton [22, 24–26]. In natural environments, the cleavage of carbon-sulfur bonds in molecules such as DBT liberates sulfur, making it available as a nutrient to support the growth of bacteria in environments poor in mineral sulfur [27]. These bacteria have been assigned to a number of genera including *Rhodococcus* [21, 28], *Acinetobacter*, and *Pseudomonas* [24]. Species of other bacteria genera, such as *Brevibacterium* sp. strain DO [29], strains identified as *Arthrobacter* spp. [30], and *Gordonia* sp. strain CYKS1 [31], are also able to use this pathway.

The pSOX plasmid [28], or *dsz* genes [32, 33], responsible for the sulfur oxidation in DBT, have been cloned, sequenced, and studied, generating considerable knowledge of these pathway enzymes [28, 32, 34]. The 4S pathway consists of three genes designated *dsz* A, B, and C. Studies have shown that the product of *dsz*C directly converts DBT to DBTO2 and the products of *dsz*A and *dsz*B act together to convert DBTO2 to 2-HBP. The operon *dsz* occupies a 4 kb gene locus in a 120 kb linear plasmid in bacteria *Rhodococcus erythropolis* strain IGTS8 [28, 32, 35–37]. The plasmid nature of the *dsz* genes increases the probability of successful transfers, and the availability of the *dsz*C sequences in GenBank allows the construction of phylogenetic hypothesis based on this gene, in order to compare it with the 16S.

In this work, we aim to demonstrate the utility of phylogenetic methods based on molecular data to help in studies of horizontal transfer of functional genes in bacteria.

2. Materials and Methods

2.1. Nucleotide Sequences. The nucleotide sequences used in this study were obtained from the National Center for Biotechnology Information-GenBank (http://www.ncbi.nlm.nih.gov). For analyses involving the *dsz*C gene, 18 sequences were selected (Table 1), representing all genera and/or species with *dsz*C sequences available in the database as of March 2012. We also searched for the other two genes of the operon, *dsz*A and *dsz*B, but they are underrepresented in GenBank and phylogenetic trees could not be constructed based on these genes. For the 16S gene, we chose 39 sequences, including at least two sequences of at least 1400 bp from each genera and/or species in the *dsz*C tree (Table 2).

Table 1: Bacteria species names and NCBI accession number of *dsz*C sequences used.

Bacteria species	NCBI accession numbers
Acidovorax delafieldii	DQ062154.1
Agrobacterium tumefaciens	AY960127.1
Bacillus subtilis	AB076745.1
Mycobacterium sp. strain G3	AB070603.1
Brevibacillus brevis	DQ062161.1
Gordonia alkanivorans strain 1B	AY678116.1
Gordonia alkanivorans	AY714057.1
Gordonia alkanivorans strain RIPI90A	EU364831.1
Gordonia sp. strain CYKS2	AY396519.1
Mycobacterium goodii strain X7B	EU527978.1
Rhodococcus erythropolis	AY714058.1
Rhodococcus erythropolis	AY294404.1
Rhodococcus sp. strain IGTS8	L37363.1
Rhodococcus sp. strain IIPS7	DQ198086.1
Rhodococcus sp. strain SDUZAWQ	AY789136.1
Rhodococcus sp. strain XP	AY278323.1
Synthetic construct dibenzothiophene monooxygenase (sequence from *Rhodococcus* sp. LY822)	EF570783.1

2.2. Phylogenetic Analysis. Phylogenetic analyses were performed with four different methods: neighbour joining (NJ) using the program MEGA 5.0 [38]; maximum parsimony (MP) and maximum likelihood (ML) using the program PAUP* [39]; Bayesian inference (BA) using the program MrBayes [40]. For NJ, ML, and BA, we chose the best nucleotide substitution model using the programs Mega 5.0 [38], ModelTest [41], and MrModelTest [42]. The chosen models are shown in Table 3. We used the Tree Bisection and Reconnection heuristic search method to search for the MP and ML trees. The MP tree started with a random tree, while the ML tree started with an NJ tree. To infer the tree through the BA, we run two independent analyses with four chains each (one cold and three hot chains), started with four different random trees modified through 5,000,000 generations of MCMC. We checked the likelihood of the resulting topologies and burned-out 25% of the trees (to keep those within the area of the best likelihoods) to construct the consensus tree. The robustness of each node of the tree was obtained by the bootstrap test (MV, MP, and NJ); the posterior probability was calculated by the frequency of each node in the consensus BA tree.

2.3. Phylogenetic Network Estimation of dszC Genes. Given the phylogenetic hypothesis for the *dsz*C gene, we constructed a network using the most related haplotypes with statistical-parsimony analyses [43]. The graphic network was constructed using TCS vers. 1.21 [44]. This method starts by calculating the overall limits of parsimony for the complete

Table 2: Bacteria species names and NCBI accession number of 16S sequences used.

Bacteria species	NCBI accession number
Acidovorax delafieldii strain 179	EU730925.1
Acidovorax delafieldii strain NBGD35	HQ003420.1
Agrobacterium tumefaciens strain NBGD13	HQ003411.1
Agrobacterium tumefaciens strain SJ61	GQ140318.1
Agrobacterium tumefaciens strain SJ22	GQ140317.1
Bacillus subtilis strain DmB55	HQ111354.1
Bacillus subtilis strain CCM7	HQ108184.1
Bacillus subtilis strain ANctcri3	HQ286641.1
Bacillus subtilis strain 13B	HQ335318.1
Brevibacillus brevis strain NBGD26	HQ003422.1
Brevibacillus brevis strain H2	HM449127.1
Brevibacillus brevis strain EIF87	HM480358.1
Brevibacillus brevis strain Hot-1	EU327889.1
Gordonia alkalivorans	Y18054.1
Gordonia alkanivorans strain DSM 44187	AY995556.1
Gordonia alkanivorans strain TPR13	EU373422.1
Gordonia alkanivorans strain HKI 0136	NR_026488.1
Gordonia amicalis strain IEGM	NR_028735.1
Gordonia amicalis strain CC-MJ-2a	EU266484.1
Gordonia amicalis strain CC-MJ-15b	EU266486.1
Gordonia amicalis	AF101418.1
Mycobacterium avium strain M214	GU142929.1
Mycobacterium avium complex strain 27497	EF611344.1
Mycobacterium avium strain ATCC 19698	EF521896.1
Mycobacterium avium strain ATCC 25291	EF521895.1
Mycobacterium avium strain IWGMT49	EF521892.1
Mycobacterium goodii	Y12872.1
Mycobacterium goodii strain M069	NR_029341.1
Mycobacterium goodii strain X7B	AF513815.1
Rhodococcus erythropolis strain ZJB-0910	GU726138.1
Rhodococcus erythropolis strain MJ2	GU991529.1
Rhodococcus erythropolis strain 13648E	EU741153.1
Rhodococcus erythropolis strain e1	EU434599.1
Rhodococcus sp. NKCM 2512	AB591806.1
Rhodococcus sp. BY44	FR690460.1
Rhodococcus sp. ITP08	FR667175.1
Rhodococcus sp. SH15	HM590053.1
Rhodococcus equi strain ATCC 6939	FJ468344.1
Rhodococcus erythropolis strain XP	DQ074453.1

data set using a statistic from neutral coalescent theory [45, 46]. Although this method has been used extensively with restriction site and nucleotide sequence data to estimate population level genealogies when divergences are low (intraspecific data) [46, 47], it also proved to be reliable at higher divergences, outperforming parsimony and parsimony with bootstrapping [48].

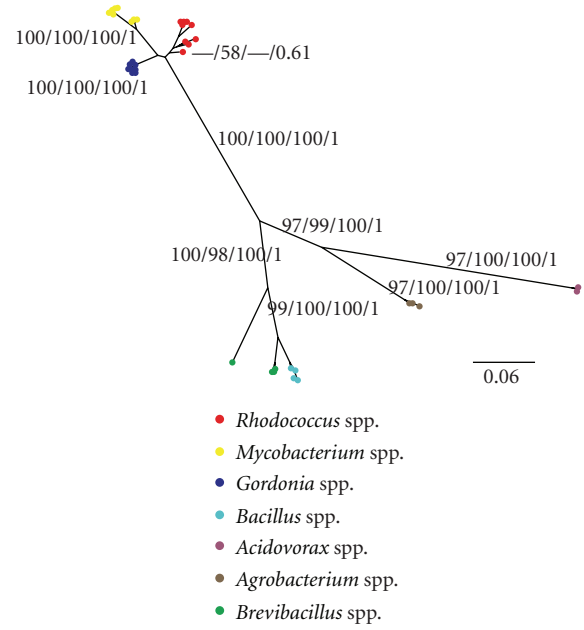


Figure 1: Tree obtained by BA analysis from sequences of the 16S gene. The values on the branches represent bootstrap values of NJ, MV, and MP and posterior probability of the BA analysis.

3. Results and Discussion

The BA hypotheses for the 16S gene are shown in Figure 1, which presents the expected pattern of species grouped within their respective genera. The different phylogenetic methods resulted in very similar tree topologies (data not shown) and robust bootstrap values for NJ, ML, MP, and BA posterior probability of the branches. The only exception was the branch containing representatives from the *Rhodococcus* spp. (in red), which showed low bootstrap value in MP (58%), low posterior probability value in BA (0.61), and bootstrap values for NJ and ML < 50%. Although the convergence of results using different phylogenetic methods has been considered good evidence that the correct phylogeny was obtained [49], total genome phylogenies show that different phylogenetic methods can provide incongruent phylogenies [50, 51]. However, the comparison of 16S sequences is still considered a powerful and accepted tool for deducing phylogenetic and evolutionary relationships among bacteria and is routinely used [4, 52–54]. In fact, most of bacteria systematics is based on the topologies generated by this gene [3].

Figure 2, on the other hand, did not group the species by genera. Instead, this figure presents only three branches: the first includes *Mycobacterium* sp. (strain G3—AB070603.1) and *Bacillus subtilis* (AB076745.1) sequences; the second groups two sequences of *Gordonia alkanivorans* (strain 1 B—AY678116.1 and strain RIPI90A—EU364831.1); the third clusters all remaining sequences belonging to all genera included in this work except for *Bacillus*. It is expected that molecular phylogenies based on single genes lead to apparently conflicting results with alternative branches that present low bootstrap (or posterior probability) values [50].

Table 3: Software, nucleotide substitution models and criteria used for phylogenetic analysis of 16S and *dsz*C genes in each tested method.

Method	Gene	Nucleotide substitution model	Gamma distribution	Invariable sites proportion
NJ	*dsz*C	Tamura-Nei	0.71	—
	16S	Tamura-Nei	0.69	—
ML	*dsz*C	GTR+G	0.8291	—
	16S	GTR+G+I	0.8125	0.3667
BA	*dsz*C	GTR+G	0.8291	—
	16S	GTR+G+I	0.8126	0.3667

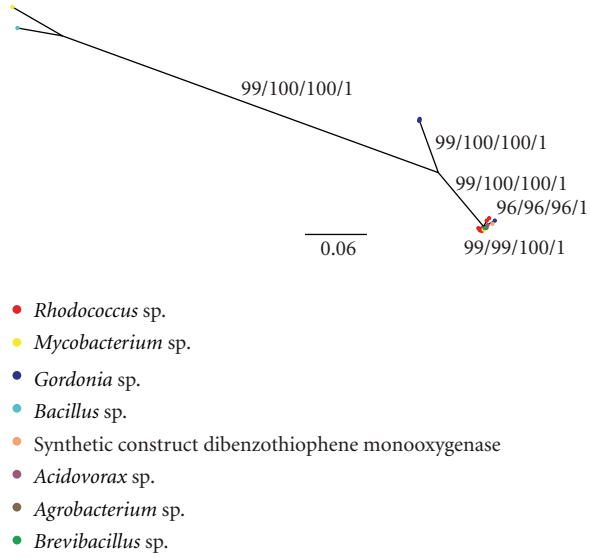


Figure 2: Tree obtained by BA analysis from sequences of the *dsz*C gene. The values on the branches represent bootstrap values of NJ, MV, and MP and posterior probability of the BA analysis.

However, the conflicting topologies shown in Figures 1 and 2 present high bootstrap and posterior probability values in alternative branches, strongly suggesting that the *dszC* was indeed subjected to horizontal transfer events among these bacteria.

The phylogenetic network estimation (Figure 3) of the *dsz*C haplotypes of the most specious cluster shown in Figure 2 emphasizes the fact that *Acidovorax delafieldii* (Seq 1), *Agrobacterium tumefaciens* (Seq 2), *Brevibacillus brevis* (Seq 3), and *Rhodococcus* sp. (Seq 11) present identical sequences, which were grouped together as a square within Figure 3, while other haplotypes are displayed as ovals connected to the square by lines with black circles to indicate the maximum number of steps between each pair of haplotypes.

The results presented here, based solely on GenBank data, provide strong evidence that the *dsz*C gene was horizontally transferred among different evolutionary lineages of bacteria. This evidence is reinforced by the fact that the *dczC* gene is generally found in conjugative plasmids, in the vicinity of insertion sequences, transcribed in the same direction and under the control of a single promoter [35, 37, 55, 56]. Furthermore, another evidence of *dsz*C horizontal transfer is the significant difference of the C+G content of this gene with the C+G content of the entire chromosome of some species studied here (data not shown).

Our results reinforce the importance of public sequence repositories (such as GenBank), which result from a successful policy of requiring the inclusion of gene sequences in public databases in order to publish any research article containing sequence analyses [57, 58]. In addition to the DNA sequence of each entry, GenBank and other public databases include associated metadata, which provide relevant information about the organism whose sequence is available, generally by linking to the articles with the respective sequence [58]. However, the public databases also contain several molecular sequences submitted by researchers who have not published their results. In these cases, there is neither citation information nor any relevant data about the organisms from which the sequences were made, which in most cases makes the sequences useless for *in silico* works, since diverse knowledge about a given molecular sequence provides an essential first step in developing research hypotheses.

It is easy to generate new sequences and add them to the GenBank database, which contains about 150 million gene sequences as of February 2012. However, GenBank, along with its INSDC (International Nucleotide Sequence Database Collaboration) partners (EMBL & DDBJ), should be treated not only as archival stores of molecular sequence data (a task at which it has been very successful) but also as a starting point for future studies. In this context, it would be helpful if the process of submitting sequences required a minimum of information about the organism from which the sequences were made, as well as the details of the gene sequenced, in order to substantiate future research.

In this sense, our study could be improved if flanking DNA sequences of functional genes such as *dsz*C were available in the databases, since we could then evaluate if one set of *dsz* genes is flanked by a particular insertion sequence while another cluster is not.

Although laboratory data that demonstrate the transfer by conjugation of plasmids containing *dsz* genes or transposition of these genes are scarce, their distribution in bacterial cultures strongly supports the hypothesis that these genes are commonly subject to horizontal transfer in nature as evidenced in the present work. For this reason, we conclude that phylogenetic tools can be useful for inferring horizontal

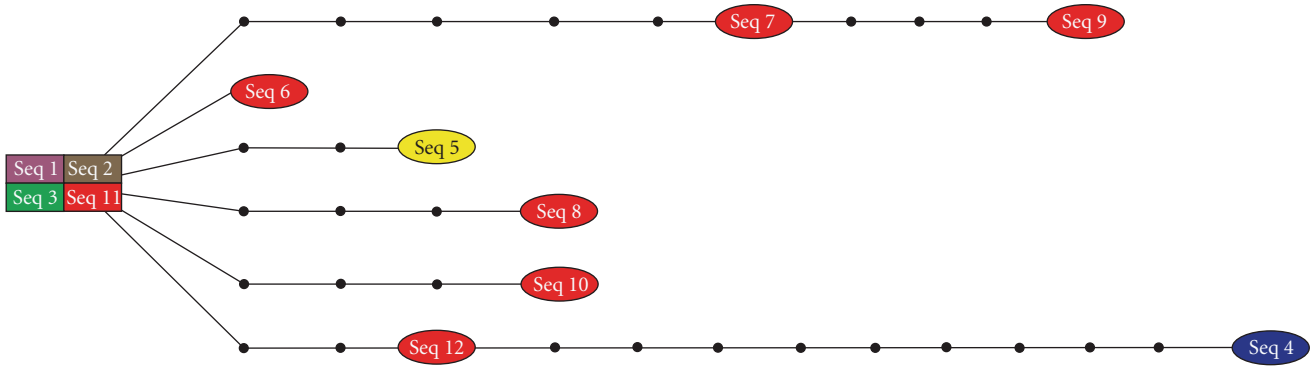


Figure 3: Phylogenetic network estimation of *dsz*C genes that remained grouped after phylogenetic analyses. Seq 1: *Acidovorax delafieldii* (DQ062154.1); Seq 2: *Agrobacterium tumefaciens* (AY960127.1); Seq 3: *Brevibacillus brevis* (DQ062161.1); Seq 4: *Gordonia alkanivorans* (AY714057.1); Seq 5: *Mycobacterium goodii* (EU527978.1); Seq 6: *Rhodococcus erythropolis* (AY714058.1); Seq 7: *Rhodococcus erythropolis* (AY294404.1); Seq 8: *Rhodococcus* sp. (L37363.1); Seq 9: *Rhodococcus* sp. (DQ198086.1); Seq 10: *Rhodococcus* sp. (AY789136.1); Seq 11: *Rhodococcus* sp. (AY278323.1); Seq 12: Synthetic construct dibenzothiophene monooxygenase (EF570783.1).

transfer events of functional genes such as *dsz*C. Phylogenetic comparisons with other genes traditionally used for this purpose, such as 16S, can provide good information about evolution and functional gene distribution.

Lateral gene transfer events provide a venue for bacterial diversification by rearranging existing capabilities. Because bacterial genomes can maintain only a finite amount of information, they are sampling rather than accumulating sequences, counterbalancing gene acquisition with gene loss. As a result, lateral gene transfer can redefine the ecological niche of a microorganism, in effect promoting bacterial speciation [58]. Although a potential result of interspecific recombination is the uncertainty of species boundaries, the increased mixing of genes and the observed phylogenetic inconsistencies show the history of a gene-transfer-mediated diversification of microorganisms.

Acknowledgments

The authors would like to thank M.S. Marcos M. D. Carneiro for help in reviewing the English of this work, M.S. Gilberto de O. Mendes for his help in editing the figures, M.S. Mateus Ferreira Santana for his help in phylogenetic network analyses, and FAPEMIG, CNPq, and CAPES for financial support. This paper is dedicated *in memoriam* of Aurea C. S. Pylro.

References

[1] C. R. Woese and G. E. Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 11, pp. 5088–5090, 1977.

[2] C. R. Woese, "Bacterial evolution," *Microbiological Reviews*, vol. 51, no. 2, pp. 221–271, 1987.

[3] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 12, pp. 4576–4579, 1990.

[4] S. Winker and C. R. Woese, "A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics," *Systematic and Applied Microbiology*, vol. 14, no. 4, pp. 305–310, 1991.

[5] G. J. Olsen, N. Larsen, and C. R. Woese, "The ribosomal RNA database project," *Nucleic Acids Research*, vol. 19, pp. 2017–2021, 1991.

[6] B. L. Maidak, G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese, "The RDP (Ribosomal Database Project)," *Nucleic Acids Research*, vol. 25, no. 1, pp. 109–110, 1997.

[7] H. L. Coutinho, V. M. De Oliveira, G. P. Manfio, and A. S. Rosado, "Evaluating the microbial diversity of soil samples: methodological innovations.," *Anais da Academia Brasileira de Ciências*, vol. 71, no. 3, pp. 491–503, 1999.

[8] D. A. Stahl, "Molecular approaches for the measurement of density, diversity and phylogeny," in *Manual of Environmental Microbiology*, C. J. Hurst, G. R. Knudsen, M. J. McInerney, L. D. Stetzenbach, and M. V. Walter, Eds., pp. 102–114, ASM Press, Washington, DC, USA, 1997.

[9] W. Martin, "Mosaic bacterial chromosomes: a challenge en route to a tree of genomes," *BioEssays*, vol. 21, no. 2, pp. 99–104, 1999.

[10] W. F. Doolittle, "Phylogenetic classification and the universal tree," *Science*, vol. 284, no. 5423, pp. 2124–2128, 1999.

[11] S. Paytubi, J. García, and A. Juárez, "Bacterial Hha-like proteins facilitate incorporation of horizontally transferred DNA," *Central European Journal of Biology*, vol. 6, no. 6, pp. 879–886, 2011.

[12] R. E. Collins and J. W. Deming, "Abundant dissolved genetic material in Arctic sea ice Part II: viral dynamics during autumn freeze-up," *Polar Biology*, vol. 34, no. 12, pp. 1819–1830, 2011.

[13] R. C. Baños, A. Vivero, S. Aznar et al., "Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS," *PLoS Genetics*, vol. 5, no. 6, Article ID e1000513, 2009.

[14] E. V. Koonin, "Horizontal gene transfer: The path to maturity," *Molecular Microbiology*, vol. 50, no. 3, pp. 725–727, 2003.

[15] A. Boc and V. Makarenkov, "Towards an accurate identification of mosaic genes and partial horizontal gene transfers," *Nucleic Acids Research*, vol. 39, no. 21, article e144, 2011.

[16] E. Hilario and J. P. Gogarten, "Horizontal transfer of ATPase genes—the tree of life becomes a net of life," *BioSystems*, vol. 31, no. 2-3, pp. 111–119, 1993.

[17] J. N. Bautista-Zapanta, H. H. Arafat, K. Tanaka, H. Sawada, and K. Suzuki, "Variation of 16S-23S internally transcribed spacer sequence and intervening sequence in rDNA among the three major Agrobacterium species," *Microbiological Research*, vol. 164, no. 6, pp. 604–612, 2009.

[18] T. Asai, D. Zaporojets, C. Squires, and C. L. Squires, "An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 5, pp. 1971–1976, 1999.

[19] Y. Tanaka, T. Matsui, J. Konishi, K. Maruhashi, and R. Kurane, "Biodesulfurization of benzothiophene and dibenzothiophene by a newly isolated *Rhodococcus* strain," *Applied Microbiology and Biotechnology*, vol. 59, no. 2-3, pp. 325–328, 2002.

[20] T. Ohshiro and Y. Izumi, "Microbial desulfurization of organic sulfur compounds in petroleum," *Bioscience, Biotechnology, and Biochemistry*, vol. 63, no. 1, pp. 1–9, 1999.

[21] F. Davoodi-Dehaghani, M. Vosoughi, and A. A. Ziaee, "Biodesulfurization of dibenzothiophene by a newly isolated *Rhodococcus erythropolis* strain," *Bioresource Technology*, vol. 101, no. 3, pp. 1102–1105, 2010.

[22] J. H. Chang, Y. K. Chang, K. S. Cho, and H. N. Chang, "Desulfurization of model and diesel oils by resting cells of Gordona sp.," *Biotechnology Letters*, vol. 22, no. 3, pp. 193–196, 2000.

[23] G. F. Duarte, A. S. Rosado, L. Seldin, W. De Araujo, and J. D. Van Elsas, "Analysis of Bacterial Community Structure in Sulfurous-Oil-Containing Soils and Detection of Species Carrying Dibenzothiophene Desulfurization (dsz) Genes," *Applied and Environmental Microbiology*, vol. 67, no. 3, pp. 1052–1062, 2001.

[24] D. Boniek, D. Figueiredo, V. S. Pylro, and G. F. Duarte, "Characterization of bacterial strains capable of desulphurisation in soil and sediment samples from Antarctica," *Extremophiles*, vol. 14, no. 5, pp. 475–481, 2010.

[25] E. S. Olson, D. C. Stanley, and J. R. Gallagher, "Characterization of intermediates in the microbial desulfurization of dibenzothiophene," *Energy & Fuels*, vol. 7, no. 1, pp. 159–164, 1993.

[26] S. Nekodzuka, T. Nakajima-Kambe, N. Nomura, J. Lu, and T. Nakahara, "Specific desulfurization of dibenzothiophene by *Mycobacterium* sp. strain G3," *Biocatalysis and Biotransformation*, vol. 15, no. 1, pp. 17–27, 1997.

[27] J. J. Kilbane Ii and S. Le Borgne, "Petroleum biorefining: the selective removal of sulfur, nitrogen, and metals," in *Studies in Surface Science and Catalysis*, V.-D. Rafael and Q.-R. Rodolfo, Eds., Chapter 2, pp. 29–65, Elsevier, 2004.

[28] S. A. Denome, C. Oldfield, L. J. Nash, and K. D. Young, "Characterization of the desulfurization genes from *Rhodococcus* sp. strain IGTS8," *Journal of Bacteriology*, vol. 176, no. 21, pp. 6707–6716, 1994.

[29] M. Van Afferden, S. Schacht, J. Klein, and H. G. Truper, "Degradation of dibenzothiophene by Brevibacterium sp.DO," *Archives of Microbiology*, vol. 153, no. 4, pp. 324–328, 1990.

[30] M. K. Lee, J. D. Senius, and M. J. Grossman, "Sulfur-specific microbial desulfurization of sterically hindered analogs of dibenzothiophene," *Applied and Environmental Microbiology*, vol. 61, no. 12, pp. 4362–4366, 1995.

[31] S. K. Rhee, J. H. Chang, Y. K. Chang, and H. N. Chang, "Desulfurization of dibenzothiophene and diesel oils by a newly isolated Gordona strain, CYKS1," *Applied and Environmental Microbiology*, vol. 64, no. 6, pp. 2327–2331, 1998.

[32] C. S. Piddington, B. R. Kovacevich, and J. Rambosek, "Sequence and molecular characterization of a DNA region encoding the dibenzothiophene desulfurization operon of *Rhodococcus* sp. strain IGTS8," *Applied and Environmental Microbiology*, vol. 61, no. 2, pp. 468–475, 1995.

[33] M. Z. Li, C. H. Squires, D. J. Monticello, and J. D. Childs, "Genetic analysis of the dsz promoter and associated regulatory regions of *Rhodococcus erythropolis* IGTS8," *Journal of Bacteriology*, vol. 178, no. 22, pp. 6409–6418, 1996.

[34] S. A. Denome, E. S. Olson, and K. D. Young, "Identification and cloning of genes involved in specific desulfurization of dibenzothiophene by *Rhodococcus* sp. strain IGTS8," *Applied and Environmental Microbiology*, vol. 59, no. 9, pp. 2837–2843, 1993.

[35] C. Oldfield, O. Pogrebinsky, J. Simmonds, E. S. Olson, and C. F. Kulpa, "Elucidation of the metabolic pathway for dibenzothiophene desulphurization by *Rhodococcus* sp. strain IGTS8 (ATCC 53968)," *Microbiology*, vol. 143, no. 9, pp. 2961–2973, 1997.

[36] J. J. Kilbane, "Desulfurization of coal: the microbial solution," *Trends in Biotechnology*, vol. 7, no. 4, pp. 97–101, 1989.

[37] C. Denis-Larose, D. Labbé, H. Bergeron et al., "Conservation of plasmid-encoded dibenzothiophene desulfurization genes in several rhodococci," *Applied and Environmental Microbiology*, vol. 63, no. 7, pp. 2915–2919, 1997.

[38] K. Tamura et al., "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.

[39] D. L. Swofford, *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4, Sinauer Associates, Sunderland, Mass, USA, 2002.

[40] J. P. Huelsenbeck and F. Ronquist, "MRBAYES: Bayesian inference of phylogenetic trees," *Bioinformatics*, vol. 17, no. 8, pp. 754–755, 2001.

[41] D. Posada and K. A. Crandall, "MODELTEST: testing the model of DNA substitution," *Bioinformatics*, vol. 14, no. 9, pp. 817–818, 1998.

[42] J. A. A. Nylander, "MrModeltest v2. Program distributed by the auhor," Evolutionary Biology Centre, Uppsala University, Sweden, 2004.

[43] A. R. Templeton, K. A. Crandall, and C. F. Sing, "A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation," *Genetics*, vol. 132, no. 2, pp. 619–633, 1992.

[44] M. Clement, D. Posada, and K. A. Crandall, "TCS: a computer program to estimate gene genealogies," *Molecular Ecology*, vol. 9, no. 10, pp. 1657–1659, 2000.

[45] R. R. Hudson, "How often are polymorphic restriction sites due to a single mutation?" *Theoretical Population Biology*, vol. 36, no. 1, pp. 23–33, 1989.

[46] J. Gómez-Zurita, E. Petitpierre, and C. Juan, "Nested cladistic analysis, phylogeography and speciation in the Timarcha goettingensis complex (Coleoptera, Chrysomelidae)," *Molecular Ecology*, vol. 9, no. 5, pp. 557–570, 2000.

[47] C. Vilà, I. R. Amorim, J. A. Leonard et al., "Mitochondrial DNA phylogeography and population history of the grey wolf Canis lupus," *Molecular Ecology*, vol. 8, no. 12, pp. 2089–2103, 1999.

[48] K. A. Crandall, "Intraspecific cladogram estimation: accuracy at higher levels of divergence," *Systematic Biology*, vol. 43, no. 2, pp. 222–235, 1994.

[49] A. Rokas, B. I. Williams, N. King, and S. B. Carroll, "Genome-scale approaches to resolving incongruence in molecular phylogenies," *Nature*, vol. 425, no. 6960, pp. 798–804, 2003.

[50] O. Jeffroy, H. Brinkmann, F. Delsuc, and H. Philippe, "Phylogenomics: the beginning of incongruence?" *Trends in Genetics*, vol. 22, no. 4, pp. 225–231, 2006.

[51] K. S. C. Yotoko and S. L. Bonatto, "A phylogenomic appraisal of the evolutionary relationship of Mycoplasmas," *Genetics and Molecular Biology*, vol. 30, no. 1, pp. 270–276, 2007.

[52] W. G. Weisburg, S. M. Barns, D. A. Pelletier, and D. J. Lane, "16S ribosomal DNA amplification for phylogenetic study," *Journal of Bacteriology*, vol. 173, no. 2, pp. 697–703, 1991.

[53] S. M. Hinsa-Leasure, L. Bhavaraju, J. L. M. Rodrigues, C. Bakermans, D. A. Gilichinsky, and J. M. Tiedje, "Characterization of a bacterial community from a Northeast Siberian seacoast permafrost sample," *FEMS Microbiology Ecology*, vol. 74, no. 1, pp. 103–113, 2010.

[54] D. Davolos and B. Pietrangeli, "Phylogenetic analysis on the arsenic-resistant bacteria isolated from three different freshwater environments," *Chemistry and Ecology*, vol. 27, no. 1, pp. 79–87, 2011.

[55] K. D. Young, "Identification and activity of two insertion sequence elements in *Rhodococcus* sp. strain IGTS8," *Gene*, vol. 161, no. 1, pp. 33–38, 1995.

[56] P. A. Williams, "Catabolic plasmids: fast-track bacterial evolution to combat pollution," *Microbiology Today*, vol. 31, pp. 168–171, 2004.

[57] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank," *Nucleic Acids Research*, vol. 37, no. 1, pp. D26–D31, 2009.

[58] H. Miller, C. N. Norton, and I. N. Sarkar, "GenBank and PubMed: how connected are they?" *BMC Research Notes*, vol. 2, article no. 101, 2009.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 457264, 16 pages
doi:10.1155/2012/457264

Review Article

Comparative Pathogenomics of Bacteria Causing Infectious Diseases in Fish

Ponnerassery S. Sudheesh, Aliya Al-Ghabshi, Nashwa Al-Mazrooei, and Saoud Al-Habsi

Microbiology Laboratory, Fishery Quality Control Center, Ministry of Agriculture and Fisheries Wealth, P.O. Box 427, 100 Muscat, Oman

Correspondence should be addressed to Ponnerassery S. Sudheesh, sudheeshps@yahoo.com

Received 6 February 2012; Accepted 20 March 2012

Academic Editor: Shinji Kondo

Fish living in the wild as well as reared in the aquaculture facilities are susceptible to infectious diseases caused by a phylogenetically diverse collection of bacterial pathogens. Control and treatment options using vaccines and drugs are either inadequate, inefficient, or impracticable. The classical approach in studying fish bacterial pathogens has been looking at individual or few virulence factors. Recently, genome sequencing of a number of bacterial fish pathogens has tremendously increased our understanding of the biology, host adaptation, and virulence factors of these important pathogens. This paper attempts to compile the scattered literature on genome sequence information of fish pathogenic bacteria published and available to date. The genome sequencing has uncovered several complex adaptive evolutionary strategies mediated by horizontal gene transfer, insertion sequence elements, mutations and prophage sequences operating in fish pathogens, and how their genomes evolved from generalist environmental strains to highly virulent obligatory pathogens. In addition, the comparative genomics has allowed the identification of unique pathogen-specific gene clusters. The paper focuses on the comparative analysis of the virulogenomes of important fish bacterial pathogens, and the genes involved in their evolutionary adaptation to different ecological niches. The paper also proposes some new directions on finding novel vaccine and chemotherapeutic targets in the genomes of bacterial pathogens of fish.

1. Introduction

Genome sequencing has provided us with powerful insights into the genetic makeup of the microbial world. The microbial genomics today has progressed from the long drawnout individual genome sequencing projects in the past to a level of technological advancement, where sequencing and comparing the genomes of several strains of a single pathogen is accomplished in a very short period of time [1, 2]. We are currently passing through a period of explosive developments in the field and an overwhelming glut in the genome sequence data of microorganisms. To date, over 1800 microbial genomes have been published and the sequencing of more than 5200 microbial genome are in different stages of completion (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi).

The genomics information has categorically disproved the earlier thinking that microbial genomes are static and has demonstrated that genomic evolutionary processes are much more flexible and dynamic than previously thought. This has led to the emergence of new ideas such as "uprooting the tree of life" and the concept of "horizontal genomics" [3–8]. This new thinking about microbial genome evolution has emerged from the observations of lineage-specific genome reduction and horizontal gene transfer (HGT), frequently occurring in bacterial genomes. Increasingly, genome sequencing projects have identified an unexpected level of diversity among bacteria, which can often be linked to recombination and gene transfer between a variety of prokaryotic organisms.

There is large variation in size and content of bacterial genomes between different genera and species, and also among strains of the same species. Known genome sizes of bacteria range from under 0.6 to 10 megabases (Mb). The smallest bacterial genomes reported are for the mycoplasmas and related bacteria, with sizes as low as 530 kilobases

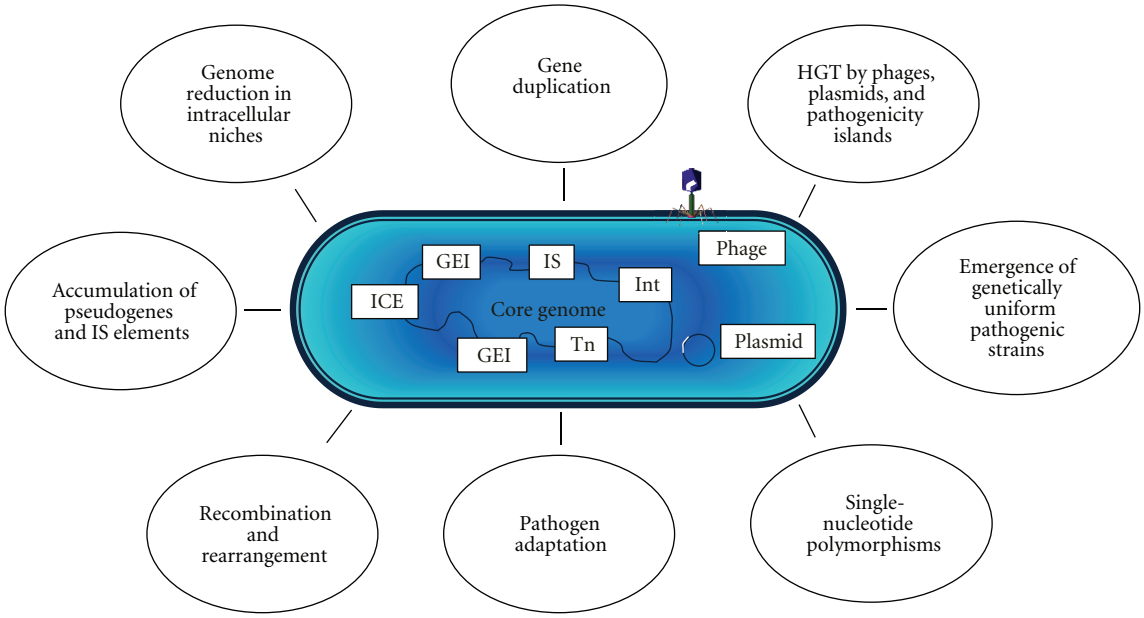


Figure 1: Major factors responsible for the pathogenomic evolution of bacteria (modified from [14, 15]; HGT: horizontal gene transfer, GEIs: genomic islands, ICEs: integrative conjugative elements, Int: integrons, Tn: conjugative transposons, IS: IS elements.

[9]. It has been emphasized that the adaptive capability ("versatility") of bacteria directly correlates with genome size [10].

Genome sequencing of bacterial pathogens has produced exciting information on evolutionary relationships between pathogenic and nonpathogenic species and has demonstrated how each has developed special adaptations advantageous for each of their unique infectious lifestyles. In the longer term, an understanding of their genome and biology will enable scientists to design means of disrupting their infectious lifestyles.

The genomes of bacteria are made up of circular or linear chromosomes, extrachromosomal linear or circular plasmids as well as different combinations of these molecules. The functionally related genes are clustered together in very close proximity to each other, and those genes located on the "core" part of the chromosome present a relatively uniform G+C content and a specific codon usage. Closely related bacteria generally have very similar genomes [11].

The stability and integrity of the "core" sequences of the genome, however, is often interrupted by the presence of DNA fragments with a G+C content and a codon usage markedly different from those of the "core" genome. The "flexible" gene pool or the so-called "mobilome" [12], is created by the acquisition of strain-specific "assortments" of genetic information mainly represented by mobile genetic elements (MGE), such as plasmids, bacteriophages, genomic/pathogenicity islands (GEIs/PAIs), integrons, IS elements (ISEs), and transposons (see Figure 1). The flexible genes scattered in the genome provide the microbes with an additional repertoire of arsenal, for example, resistance to antibiotics, production of toxic compounds as well as other virulence factors [13].

A fundamental question in biology is to define the minimum number of genes or functions to support cellular life. The size of bacterial genomes is primarily the result of two counteracting processes: the acquisition of new genes by gene duplication or by horizontal gene transfer; the deletion of nonessential genes. Genomic flux created by these gains and losses of genetic information can substantially alter gene content. This process drives divergence of bacterial species and eventually adaptation to new ecological niches [16].

Bacterial pathogens are a major cause of infectious diseases and mortality in wild fish stocks and fish reared in confined conditions. Disease problems constitute the largest single cause of economic losses in aquaculture [17]. Concurrent with the rapid growth and intensification of aquaculture, increased use of water bodies, pollution, globalization, and transboundary movement of aquatic fauna, the list of new pathogenic bacterial species isolated from fish has been steadily increasing [18]. In addition, the virulence and host range of existing pathogens has also been increasing, posing considerable challenge to fish health researchers, who are actively looking for more efficient vaccines and therapeutic drugs to combat bacterial fish diseases. The current treatment methods are ineffective and have many practical difficulties.

At the level of host-pathogen interaction, there is considerable pressure on pathogens to adapt to the harsh host environment as well as to adapt and evolve along with the ever changing external environment. The interplay between the host and the pathogen is a complex one, each driven by the need to secure the success of the species. Adaptations by one partner to exploit new environments will often stimulate the other to modify its characteristics to take advantage of the change. As a consequence of this cycle of interaction created by changing environments, new strains of pathogen

Table 1: Currently sequenced genomes of bacterial pathogens of fish.

Organisms	Size (Mb)	CDS**	Unknown/ Hypothetical genes (%)	Pseudogenes prophages, ISE/GEI	% GC	Chromosomes	Plasmids
Vibrio anguillarum 775 serotype O1	4.117	3880	26	92	44.3	2	1
Vibrio anguillarum 96F serotype O1	4.065	3766	26	38	42	2	0
Vibrio anguillarum RV22 serotype O2β	4.022	3949	26	68	43.1	2	0
Vibrio ordalii ATCC 33509	3.415	3281	—	31	43.3	2	0
Vibrio harveyi ATCC BAA-1116*	6.054	—	—	—	45.4	2	1
Vibrio vulnificus YJ016 biotype 1	5.26	5028	34	—	46.1	2	1
Vibrio splendidus strain LGP32	4.974	4498	24.8	—	43.8	2	0
Aliivibrio salmonicida strain LFI1238	4.655	4286	—	1179	38.3	2	4
Flavobacterium psychrophilum JIP02/86	2.862	2432	45.3	94	32.5	1	1
Flavobacterium branchiophilum FL-15	3.56	2867	—	54	32.9	1	—
Flavobacterium columnare ATCC 49512*	3.2	2896	—	—	32.0	1	—
Edwardsiella tarda EIB202	3.76	3486	28	97	59.7	1	1
Edwardsiella ictaluri 93–146*	3.812	3783	—	100	57.4	1	—
Aeromonas hydrophila ATCC 7966	4.744	5195	27.7	7	61.5	1	0
Aeromonas salmonicida A449	4.702	4437	—	258	58.5	1	5
Aeromonas veronii Strain B565	4.551	4057	—	—	58.7	1	—
A. caviae Ae398	4.43	—	—	6	61.4	1	1
Renibacterium salmoninarum ATCC 33209	3.155	3507	25.3	151	56.3	1	0
Streptococcus parauberis	2.143	2641	21.3	—	35.6	1	0
Lactococcus garvieae UNIUD074	2.172	2101	21.8	224	38.7	1	0
Mycobacterium marinum M	6.636	5424	26	65	62.5	1	1

* Unpublished.
** Coding sequences.

will evolve. Over time, these strains may emerge as new species with characteristic disease symptoms.

The use of antibiotics to control fish diseases has met with limited success and has the potential danger of antibiotic resistance development in aquatic bacteria (World Health Organization antimicrobial resistance fact sheet 194, http://www.who.int/inf-fs/en/fact194.html) [19]. As aquaculture is one of the fastest growing food production industries in the world, demand for sustainable ways of combating fish diseases is gaining significance. There is tremendous scope for developing novel vaccines and therapeutic drugs against bacterial fish pathogens.

Genomic evolution and adaptive strategies of bacterial fish pathogens are poorly understood and lags far behind that of human and terrestrial animal pathogens. A detailed knowledge of the genome sequences of bacterial fish pathogens and how the genomes of the pathogenic species or strains evolved from nonpathogenic ancestors or counterparts will help us better understand their pathogenicity mechanisms and strategies of host adaptations. This information will help identifying novel vaccine and drug targets in the genomes of pathogens.

Recently, genome sequencing of a number of bacteria pathogenic to fish and other aquatic organisms have been completed. The genome sequence and genome characteristics of important bacterial fish pathogens completed and published to date are summarized in Table 1.

The main aim of this paper is to put together and summarize the scattered genome sequencing information on important bacterial fish pathogens available in the literature to date. We sincerely believe that this paper will provide a genomic perspective on the adaptive evolutionary strategies of bacterial fish pathogens in different ecological niches and will help better understand the virulence mechanisms and pathogenesis of infections. It is hoped that this will lead to finding the most appropriate vaccine and therapeutic drug targets in the genomes and developing efficient control and treatment methods for fish diseases.

2. Bacterial Pathogens of Fish

Although pathogenic species representing majority of existing bacterial taxa have been implicated in fish diseases, only a relatively small number of pathogens are responsible for important economic losses in cultured fish worldwide. Major bacterial pathogens responsible for infectious disease outbreaks in different species of fish are listed in Table 2. Major groups of bacteria causing infectious diseases in fish and the important genome characteristics of these bacteria are described in the following sections.

3. *Vibrios*

Bacteria in the genus *Vibrio* are mainly pathogenic to marine and brackish water fish. However, they are occasionally

Table 2: Major bacterial pathogens of economically important fish.

Causative agent/species	Disease	Main host fish
Gram-negatives		
Vibrio anguillarum	Vibriosis	Salmonids, turbot, sea bass, striped bass, eel, ayu, cod, and red sea bream
Aliivibrio salmonicida (formerly *Vibrio salmonicida*)	Vibriosis	Atlantic salmon, cod
Vibrio vulnificus	Vibriosis	Eels, tilapia
Vibrio ordalii	Vibriosis	Salmonids
Vibrio carchariae (syn.: *Vibrio harveyi*)	Vibriosis, infectious gastroenteritis	Shark, abalone, red drum, sea bream, sea bass, cobia, and flounder
Moritella viscosa (formerly *Vibrio viscosus*)	Winter ulcer	Atlantic salmon
Photobacterium damselae subsp. piscicida (formerly *Pasteurella piscicida*)	Photobacteriosis (pasteurellosis)	Sea bream, sea bass, sole, striped bass, and yellowtail
Pasteurella skyensis	Pasteurellosis	Salmonids and turbot
Tenacibaculum maritimum (formerly *Flexibacter maritimus*)	Flexibacteriosis	Turbot, salmonids, sole, sea bass, gilthead sea bream, red sea bream, and flounder
Flavobacterium psychrophilum	Coldwater disease	Salmonids, carp, eel, tench, perch, ayu
Flavobacterium branchiophila	Bacterial gill disease	A broad range of cultured cold water and warm water salmonid and nonsalmonid fishes
Flavobacterium columnare	Columnaris disease	cyprinids, salmonids, silurids, eel, and sturgeon
Pseudomonas anguilliseptica	Pseudomonadiasis, winter disease	Sea bream, eel, turbot, and ayu
Aeromonas salmonicida	Furunculosis	salmon, trout, goldfish, koi and a variety of other fish species
Aeromonas hydrophila *Aeromonas veronii* Biovar Sobria *Aeromonas sobria* Biovar Sobria (Motile aeromonads)	Motile aeromonas septicemia (MAS), hemorrhagic septicemia, ulcer disease or red-sore disease, and epizootic ulcerative syndrome (EUS)	A wide variety of salmonid and nonsalmonid fish, sturgeon, tilapia, catfish, striped bass, and eel
Edwardsiella ictaluri	Enteric septicemia	Catfish and tilapia
Edwardsiella tarda	Edwardsiellosis	Salmon, carps, tilapia, catfish, striped bass, flounder, and yellowtail
Yersinia ruckeri	Enteric redmouth	Salmonids, eel, minnows, sturgeon, and crustaceans
Piscirickettsia salmonis	Piscirickettsiosis	Salmonids
Gram-positives		
Lactococcus garvieae (formerly *Enterococcus seriolicida*)	Streptococcosis or lactococcosis	Yellowtail and eel
Streptococcus iniae	Streptococcosis	Yellowtail, flounder, sea bass, and barramundi
Streptococcus parauberis	Streptococcosis	Turbot
Streptococcus phocae	Streptococcosis	Atlantic salmon
Renibacterium salmoninarum	Bacterial kidney disease	Salmonids
Mycobacterium marinum	Mycobacteriosis	Sea bass, turbot, and Atlantic salmon

reported in freshwater species as well [20, 21]. The distribution of vibriosis is worldwide and causes great economic loss to the aquaculture industry [22]. Vibriosis, one of the major bacterial diseases affecting fish, bivalves, and crustaceans, is mainly caused by pathogenic species such as *Vibrio anguillarum*, *V. harveyii* (Syn. *V. carchariae*), *V. ordalii*, and *Aliivibrio salmonicida* (formerly *Vibrio salmonicida*) [23, 24].

Other species such as *V. vulnificus* [25, 26] *and Moritella viscosa* (formerly *Vibrio viscosus*) [27] have been implicated in fish diseases such as septicemia and winter ulcer, respectively; more pathogenic species have been isolated frequently and reported in the literature [28].

Genome sequences of four major fish pathogenic vibrios, *V. anguillarum, V. ordalii, Aliivibrio salmonicida, and V.vulnificus* have been completed and published [29–31]. Generally, they have two chromosomes, one larger and one smaller. The majority of genes that encode cell functions and pathogenic factors are located in the large one. The

small chromosome usually contains genes for environmental adaptation.

Vibrio anguillarum is the most studied aetiological agent of vibriosis [32]. *V. anguillarum* typically causes a hemorrhagic septicemia. The O1 and O2 serotypes are the virulent strains frequently isolated from diseased fish [33, 34]. Many O1 serotype strains harbor 65 kb pJM1-type plasmids, which carry the siderophore anguibactin biosynthesis and transport genes, a main virulence factor of *V. anguillarum*, while one of the O1 serotype strains and other serotypes, such as all of the O2 strains, are plasmidless [28, 35, 36]. The O1 serotype strains cause disease in salmonid fish, whereas O2 β strains are usually isolated from cod and other nonsalmonids [28, 32].

Vibrio ordalii is a very close relative of *V. anguillarum* [37] and was previously recognized as *V. anguillarum* biotype 2. Vibriosis caused by these two species are strikingly different based on histological evidences [38]. *V. anguillarum* has a special affinity for blood and loose connective tissue, whereas *V. ordali* is mostly present as aggregates in skeletal and cardiac muscles. *V. ordalii* has a lesser affinity for blood and develops bacteremia only at late stages of disease.

Genomic sequences of three different strains of *V. anguillarum* (the strain 775 containing plasmid pJM1, serotype O1 strain 96F, and plasmidless serotype O2 β strain RV22) and *V. ordali* have recently been published [31]. The pJM1 plasmid in the strain 775 contains 65 genes including the anguibactin biosynthesis and transport genes that are unique for the strain.

V. anguillarum 775 contains more transposase genes (about 53) than 96F (about 23), RV22 (about 42), and *V. ordalii* (about 18).

The genome comparison of *V. anguillarum* serotypes has revealed some interesting differences in the genomic composition, indicating horizontal acquisition of virulence genes and the evolution of different potential virulence mechanisms among the closely related serotypes [31]. The *V. anguillarum* 96F strain has a type III secretion system 2 (T3SS2) cluster, which is absent in the 775 strain. The T3SS2 genes are highly conserved with other T3SS2 genes reported in *V. parahaemolyticus, V. cholera*, and *V. mimicus* [39–41]. In the 775 strain, three transposase genes are present at the T3SS2 chromosomal location, one of which probably originated from the pJM1, indicating that the gene cluster is inactivated by a transposition, deletion, or inversion event [31]. The 775 strain also contains 10 genomic islands including integrase, transposase, and some novel sequences conferring genomic plasticity to adapt to specific ecological niches.

The strain RV22 genome contains the toxin-antitoxin systems, and genes encoding the accessory *V. cholerae* enterotoxin (Ace) and the *Zonula occludens* toxin (Zot), which is not present in the 775 strain. The yersiniabactin-like siderophore cluster, which is highly conserved in many *Vibrio* species and *Photobacterium damselae* subspecies piscicida [42], is present in strain RV22 and *V. ordalii*.

A striking feature of *V. ordali* genome is its significant reduction in size (3.4 Mb) compared to the *V. anguillarum* strains 775 (4.1 Mb), 96F (4.0 Mb), and RV22 (4.0 Mb). *V. ordali* lacks the ABC transporter genes, the type VI secretion systems, and the gene for microbial collagenase. The Syp biofilm formation cluster, which is conserved in many *Vibrio* species such as *V. fischeri*, *V. vulnificus*, and *V. parahaemolyticus* [43, 44], is present only in *V. ordalii*. Thus, it is probable that the transition of *V. anguillarum* to *V. ordalii* is mediated by genome reductive evolution to become an endosymbiotic organism; *V. ordali* has the smallest genome of all vibrios.

Vibrio vulnificus includes three distinct biotypes. Biotype 1 strains cause human disease, while biotype 2 infects primarily eels, and biotype 3 infections has been associated with persons handling Tilapia, although the source and reservoir of biotype 3 have yet to be identified [45]. In another classification the terms clade 1 and clade 2 are used based on the multilocus sequence typing (MLST) [46]. Biotype 1 strains are present in both clades, whereas biotype 2 strains are present only in clade 1, and biotype 3 strains appear to be a hybrid between clades 1 and 2. Clade 1 strains are most often isolated from environmental samples, while clade 2 strains are mostly associated with human disease and are considered more virulent. Recent comparative genomic analysis of these biotypes or clades has clearly differentiated them based on the possession of an array of clade-specific unique genes including the presence of a virulence-associated genomic island XII in the highly virulent strains [30].

Aliivibrio salmonicida (formerly *Vibrio salmonicida*) causes coldwater vibriosis in marine fish such as farmed Atlantic salmon (*Salmo salar*), sea-farmed rainbow trout (*Oncorhynchus mykiss*), and captive Atlantic cod (*Gadus morhua*) [47]. The Gram-negative bacterium causes tissue degradation, hemolysis, and sepsis *in vivo*. Genome sequencing of *Aliivibrio salmonicida* has revealed a mosaic structure of the genome caused by large intrachromosomal rearrangements, gene acquisition, deletion, and duplication of DNA within the chromosomes and between the chromosomes and the plasmids [29].

The genome has many genes that appear to be recently acquired by HGT, and large sections of over 300 coding sequences (CDS) are disrupted by IS elements or contain point mutations causing frame shifts or premature stop codons [29]. The genomic islands (GIs) identified in the bacteria include major virulence-related genes encoding T6SS and Flp-type pilus and genes that appear to provide new functions to the bacteria. The Tad system has been proposed to represent a new subtype of T2SS and is essential for biofilm formation, colonization, and pathogenesis [48].

The genome analysis has unequivocally confirmed that *Aliivibrio salmonicida* has undergone extensive rearrangement of its genome by losing massive functional genes and acquiring new genes and become host-restricted, allowing the pathogen to adapt to new niches. IS expansion has been related to genome reduction in the evolution and emergence of pathogenicity [49], and accumulation of pseudogenes has been described for several other host restricted pathogens [50, 51].

4. *Aeromonads*

Aeromonas hydrophila and other motile aeromonads are among the most common bacteria in a variety of aquatic

environments worldwide, including bottled water, chlorinated water, well water, sewage, and heavily polluted waters, and are frequently associated with severe disease among cultured and feral fishes, amphibians, reptiles, and birds [52]. Aeromonads are also considered serious emerging pathogens of human beings [53]. Determination of the etiology of diseases involving aeromonad infections has been complicated by the genetic, biochemical, and antigenic heterogeneity of members of this group.

The genus *Aeromonas* has been conveniently divided into a group of nonmotile, psychrophilic species, prominently represented by *Aeromonas salmonicida*, which is an obligate fish pathogen and a second group of mostly human pathogenic, motile, and mesophilic species including *A. hydrophila*.

Genome sequencing of *A. hydrophila* ATCC 7966^T, *A. salmonicida* subsp. salmonicida A449, *A. veronii* strain B565, and *A. caviae* [54–57] has helped in resolving their taxonomic confusion and has brought new insights into the way these bacteria adapt to a myriad of ecological niches, their host adaptive evolution and virulence mechanisms.

Aeromonas salmonicida, the causative agent of furunculosis in salmonid and nonsalmonid fish, is a non-motile, Gram-negative bacterium; furunculosis is an important disease in wild and cultured stocks of fish inflicting heavy losses to aquaculture industry worldwide [58, 59]. *A. hydrophila* causes a septicemic disease in fish known variously as "motile aeromonas septicemia" (MAS), "hemorrhagic septicemia," "ulcer disease," or "red-sore disease" [60]. The disease caused by this bacterium primarily affects freshwater fish such as catfish, several species of bass, and many species of tropical or ornamental fish. *A. veronii* is the causative agent of bacterial hemorrhagic septicemia in fish and is becoming a major economic problem in the fish-farming industry [23].

Genome sequencing of the fish pathogen *A. salmonicida* A449 has confirmed the presence of fully functional genes for a type III secretion system (T3SS) that has been shown to be required for virulence in *A. salmonicida* [61], and genes for a type VI secretion system (T6SS), which is disrupted by an IS element [55]. The ancestral state of the T3SS in *A. salmonicida* A449 is ambiguous because of the absence of the genes in *A. hydrophila* ATCC 7966^T, while other *A. hydrophila* strains carry T3SS operons on the chromosome [62]. The genome contains a multitude of virulence-related genes including several types of adhesins (e.g., surface layer, flagella, and pili), toxin genes (aerolysin, hemolysin, repeats in toxin (RTX) protein, and cytolytic delta-endotoxin), secreted enzymes (protease, phospholipase, nuclease, amylase, pullulanase, and chitinase), antibiotic resistance genes (*tetA*, *β-lactamase* gene, and efflux pumps), and genes involved in iron acquisition and quorum sensing.

Most of the above genes are present in *A. hydrophila* ATCC 7966^T genome and an expansion of gene families (paralogs) of ABC transporters, two-component signal transduction systems (TCSs), transcriptional regulators, FeS cluster-binding proteins involved in energy transduction at the membrane, and methyl-accepting chemotaxis proteins (MCPs). Interestingly, transposase, resolvase, or insertion sequence element sequences were not discovered in the *A. hydrophila* ATCC 7966^T genome, whereas these have been identified in *A. salmonicida* and *A. caviae* genomes. *A. salmonicida* possesses 88 copies of 10 different IS elements whereas *A. caviae* Ae398 has only five different IS elements, and *A. hydrophila* completely lacks IS elements.

Although *A. hydrophila* ATCC 7966^T has been demonstrated to be the second most virulent species among *Aeromonas* [63], a very important virulence determinant, T3SS, which is present in *A. salmonicida* A449 is strikingly absent in *A. hydrophila* ATCC 7966^T genome. *A. caviae* contains many putative virulence genes, including those encoding a type 2 secretion system, an RTX toxin, and polar flagella.

The genome of *A. veronii* strain B565 contains some putative virulence factors, such as chitinase, RTX protein, adhesion factor, flagella, and mannose-sensitive hemagglutinin (MSHA), all of which are shared with *A. hydrophila* ATCC 7966^T and *A. salmonicida* A449. On the other hand, 346 genes including some important putative virulence factors such as hemolysins and the type III secretion protein, which are shared by the latter two species are absent in *A. veronii* strain B565.

Many unique genes in *A. hydrophila* ATCC 7966^T and *A. salmonicida* A449 are virulence genes and often form large clusters, such as the *rtx* cluster in ATCC 7966^T and the flagellar gene cluster in A449, or are involved in mobile elements such as phages and transposons, highlighting their lateral transfer history [56].

The *A. hydrophila* ATCC 7966^T and *A. salmonicida* A449 genomes appear to be very closely related, encoding similar number of proteins with only 9% difference in gene content. However, there are many transposons, phage-related genes, and unique CDS in *A. salmonicida* A449 genome that are different from *A. hydrophila* ATCC 7966^T sequences, showing their distinct lineages and adaptive evolution that occurred while segregating into different species of the genus.

In sharp contrast to *A. hydrophila* ATCC 7966^T genome, the *A. salmonicida* A449 genome is characterized by the presence of large numbers of several different types of IS elements in multiple copies, with more than 20 genes being interrupted by IS elements. *A. hydrophila* ATCC 7966^T genome has no IS elements.

There is a higher tendency for genomic reduction in *A. salmonicida* A449 with the formation of many pseudogenes, and *A. hydrophila* ATCC 7966^T has only seven pseudogenes. The formation of pseudogenes has resulted in the loss of function of many genes including flagella and type IV pili, transcriptional regulators, genes encoding carbohydrate synthesis, and modification enzymes and genes for basic metabolic pathways, which are some characteristic features of pathogenomic evolution.

Thus, *A. salmonicida* A449 appears to have evolved much faster than *A. hydrophila* ATCC 7966^T through genetic rearrangements, genomic reduction, and HGT from common ancestral lineages by acquiring and forming multiple plasmids, prophages, a battery of IS elements, pseudogenes, and several individual genes and operons.

5. *Flavobacterium*

The genus *Flavobacterium* includes over 30 species of which *Flavobacterium psychrophilum*, *F. branchiophilum*, and *F. columnare* are important disease agents for salmonids, catfish, and other cultured species [64, 65]. Flavobacteria are significant as they are ubiquitous in the soil, freshwater, and marine environments and are noted for their novel gliding motility and ability to degrade polymeric organic matter such as hydrocarbons [66].

F. psychrophilum is the etiological agent of bacterial cold-water disease (BCWD). It is a serious fish pathogen causing substantial economic losses and rearing difficulties to both commercial and conservation aquaculture. *F. psychrophilum* infections are found throughout the world. Juvenile rainbow trout and coho salmon are particularly susceptible to BCWD. However, *F. psychrophilum* infections have been reported in a wide range of hosts, Anguilla *japonica*, *A. anguilla*, *Cyprinus carpio*, *Carassius carassius, Tinca tinca*, *Plecoglossus altivelis*, *Perca fluviatilis*, and *Rutilus rutilus* [64, 67]. Fry and fingerlings with BCWD often have skin ulcerations on the peduncle, anterior to the dorsal fin, at the anus, or on the lower jaw and mortalities can go up to 70% [68].

F. branchiophilum is the causative organism of bacterial gill disease (BGD) in several parts of the world [69]. This disease is characterized by explosive morbidity and mortality rates attributable to massive bacterial colonization of gill lamellar surfaces and progressive branchial pathology stemming from high rates of lamellar epithelial necrosis [70].

F. columnare (formerly *Cytophaga columnaris*; *Flexibacter columnaris*) is the causative agent of columnaris disease of salmonids and other fishes in commercial aquaculture, the ornamental fish industry, and wild fish populations worldwide [71]. Classically, during outbreaks, its morbidity and mortality rates escalate more gradually than for BGD. Additionally, unlike the pattern of necrosis in BGD, fish with columnaris will have severe necrosis of all parts of the gill as the bacterium invades inwardly [72].

The taxonomy of the three species was initially based on phenotypic characteristics and has been revised several times during the years. The latest classification based on G+C content, DNA-ribosomal ribonucleic acid (rRNA) hybridisation, and fatty acid and protein profiles, has confirmed that all the three species now belong to the phylum/division *Cytophaga-Flavobacterium-Bacteroides*, family *Flavobacteriaceae*, and genus *Flavobacterium* [73].

The whole genome sequences of *F. psychrophilum* and *F. branchiophilum* have been published [74, 75]. The *F. columnare* genome sequence is yet to be completed and published [76].

Prominent features of *F. psychrophilum* infection include the strong adhesion to fish epithelial tissues followed by gliding motility, rapid and mass tissue destruction, and severe muscle tissue ulcerations. Hence, the identification of multiple genes encoding secreted proteases, adhesins, and gliding motility (*gld*) genes in *F. psychrophilum* genome indicates their possible involvement in the virulence of the pathogen. However, the gene sequence of a secreted collagenase was disrupted by an insertion sequence of the IS256 family in several strains isolated from rainbow trout [74] indicating the clonal dissemination of strains containing the disrupted gene. The *F. psychrophilum* seems to have horizontally acquired virulence associated genes from other unrelated bacteria. It has a hemolysin similar to the toxin VAH5, which is a virulence factor in *Vibrio anguillarum* [77]. It also has a gene encoding a protein that is similar to domains 1–3 of thiol-activated cytolysin family of pore-forming toxins (TACYs), which has been implicated in the pathogenicity of several Gram-positive bacteria [78]. Interestingly, *F. psychrophilum* lacks the type III and IV secretion systems usually present in Gram-negative pathogens; but, it has genes encoding PorT and PorR proteins, which are involved in transport and anchoring of virulence factors of the bacteria [79, 80]. In addition, the *F. psychrophilum* genome contains a large repertoire of genes involved in aerobic respiration, psychrotolerance, and stress response.

The sequencing of *F. branchiophilum* genome has revealed the existence of virulence mechanisms distinctly different from the closest species, *F. psychrophilum*. The *F. branchiophilum* genome has the first cholera-like toxin in a nonproteobacteria and an array of adhesins. A comparative analysis of its genome with genomes of other *Flavobacterium* species revealed a smaller genome size, large differences in chromosome organization, and fewer rRNA and tRNA genes, fitting with its more fastidious growth. In addition, identification of certain virulence factors, genomic islands, and CRISPR (clustered regularly interspaced short palindromic repeats) systems points to the adaptive evolution of *F. branchiophilum* by horizontal acquisition of genes.

6. *Edwardsiella*

The genus *Edwardsiella* belongs to subgroup 3 of γ-proteobacteria, encompassing a group of Gram-negative enteric bacteria pathogenic to a variety of animals [81]. Two very closely related species, *Edwardsiella tarda* and *E. ictaluri* are important fish pathogens. Both are Gram-negative motile rods that are cytochrome oxidase negative and ferment glucose with production of acid and gas. The two species can be differentiated biochemically in that *E. tarda* produces both indol and hydrogen sulfide, whereas *E. ictaluri* produces neither. Moreover, the two species do not cross-react serologically. *E. tarda* has been isolated from many warm water fishes and some coldwater fishes, whereas *E. ictaluri* has been isolated only from a few species of warm water fishes (Table 2). Additionally, *E. tarda* causes disease in such other animals as marine mammals, pigs, turtles, alligators, ostriches, skunks, and snakes [81]. It has also occasionally infected humans [82, 83]. In contrast, *E. ictaluri* is limited to fish, and survivors of epizootics probably become carriers. The geographic range of *E. tarda* is worldwide, whereas that of *E. ictaluri* is still confined to the catfish growing areas in the United States [84].

E. tarda causes a disease condition in fish called systemic hemorrhagic septicemia with swelling skin lesions as well as ulcer and necrosis in internal organs such as liver, kidney,

spleen, and musculature [85]. It has the ability of invading and multiplying in epithelial cells and macrophages in order to subvert the host immune system and to survive in the fish [86].

E. ictaluri is the causative agent of enteric septicemia of catfish (ESC), a major disease affecting the catfish industry. The disease can manifest as an acute form that is characterized by hemorrhagic enteritis and septicemia and a chronic disease that is characterized by meningoencephalitis [87]. Gross external symptoms include hemorrhages on the body, especially around the mouth and fins. Other signs include pale gills, exophthalmia, and small ulcerations on the body [84].

The whole genome sequencing of the two species has recently been completed and published allowing comparative genomic analysis of these very important fish pathogens [88, 89]. The genome sequencing of the two closely related species *E. tarda* and *E. ictaluri* has revealed a high level of genomic plasticity with a high content of mobile genetic elements, IS elements, genomic islands, phage-like products, integrases, or recombinases. *E. ictaluri* displays high biochemical homogeneity with only one serotype, but possess many IS elements in the genome. In addition, highly variable G+C content and a large quantity of variable number of tandem repeats (VNTRs) or direct repeat sequences were identified in the *E. tarda* genome indicating the rapid genomic evolution undergoing in the species [88]. An interesting feature is the identification of insertion sequence IS *Saen1* of *Salmonella enterica* serovar Enteritidis [90] in both *E. tarda* EIB202 and *E. ictaluri* 93–146 genomes. Conversely, the difference in genomic islands among the three species may partially explain their rapid evolutionary changes and diverging lineage from a common ancestor.

The *E. tarda* genome has a gene cluster sharing high similarities to the *pvsABCDE-psuA-pvuA* operon, which encodes the proteins for the synthesis and utilization of vibrioferrin, an unusual type of siderophore requiring nonribosomal peptide synthetase (NRPS) independent synthetases (NIS) and usually mediating the iron uptake systems in *V. parahaemolyticus* and *V. alginolyticus* [91, 92]. But *E. ictaluri* genome lacks siderophore biosynthesis genes, even though it possesses heme binding/transport genes.

E. tarda genome is smaller than that of *E. ictaluri* and other sequenced genomes of *Enterobacteriaceae*, justifying the hypothesis that *E. tarda* may not be present as a free living microorganism in natural waters but multiply intracellularly in protozoans and transmitted to fish, reptile, and other animals or humans [81].

The *E. tarda* and *E. ictaluri* genomes have a multitude of virulence factors including P pilus, type 1 fimbriae, nonfimbrial adhesins, invasins and hemagglutinins and various secretion pathways including sec-dependent transport system, the components of the main terminal branch of the general secretory pathway (GSP), the signal recognition particle (SRP), and the sec-independent twin arginine transport (Tat), T1SS, TTSS, and T6SS indicating their evolutionary fitness and ability to adapt to a variety of demanding ecological niches and harsh host intracellular environments.

7. *Yersinia ruckeri*

Yersinia ruckeri, the causal agent of enteric redmouth (ERM) disease, which is a systemic bacterial infection of fishes, but is principally known for its occurrence in rainbow trout, *Salmo gairdneri* [93]. *Y. ruckeri* was initially isolated from rainbow trout in the Hagerman Valley, Idaho, USA, in the 1950s [94] and is now widely found in fish populations throughout North America, Australia, South Africa, and Europe [95]. Outbreaks of ERM usually begin with low mortalities which slowly escalate and may result in high losses. The problem may become large-scaled if chronically infected fish are exposed to stressful conditions such as high stocking densities and poor water quality [96]. *Y. ruckeri* is a nonspore-forming bacterium which does not possess a capsule, but often has a flagellum [97].

Historically, *Y. ruckeri* is fairly homogenous in biochemical reactions. However, *Y. ruckeri* strains have recently been grouped into clonal types on the basis of biotype, serotype, and outer membrane protein (OMP) profiles [98]. Strains of serovars I and II [99], equivalent to serotypes O1a and O2b, respectively [100], cause most epizootic outbreaks in cultured salmonids, serovar I being predominant in rainbow trout [101]. Within serovar I, six clonal OMP types have been recognized, but only two are associated with major disease outbreaks: clonal group 5, which includes the so-called Hagerman strain and clonal group 2 [98, 102]. Clonal group 5 comprises the majority of isolates, all of them motile and with a widespread distribution (Europe, North America, and South Africa). Clonal group 2 includes only nonmotile strains isolated in the UK.

More recently, multilocus sequence typing has revealed distinct phylogenetic divergence of *Y. ruckeri* from the rest of the *Yersinia* genus raising doubts about its taxonomic position [103]. This view has gained credibility after the genome sequencing of *Y. ruckeri*, which has a substantially reduced total genome size (3.58 to 3.89 Mb), compared with the 4.6 to 4.8 Mb seen in the genus generally [104]. In addition, *Y. ruckeri* was found to be the most evolutionarily distant member of the genus with a number of features distinct from other members of the genus.

Several common *Yersinia* genes were missing in *Y. ruckeri*. These included genes involved in xylose utilization, urease activity, B12-related metabolism, and the *mtnKADCBEU* gene cluster that comprises the majority of the methionine salvage pathway [104]. The genomic reduction achieved by losing these and other genes is suggestive of its means of adaptation to an obligatory life style in fish hosts.

8. *Renibacterium salmoninarum*

Renibacterium salmoninarum is a small Gram-positive diplobacillus, and the causative agent of bacterial kidney disease (BKD), which is a slowly progressive, systemic infection in salmonid fishes with a protracted course and an insidious nature [105]. The pathogen can be transmitted from fish to fish [106] or from adults to their progeny via eggs [107]. Infected fish may take months to show signs of disease. bacterial kidney disease is one of the most difficult bacterial

diseases of fish to treat [108], mainly due to its ability to evade phagocytosis and invade and survive in host cells [109, 110]. *R. salmoninarum* is very slow growing, and it is extremely difficult to apply genetic manipulation techniques to study its gene functions.

R. salmoninarum, despite being an obligate intracellular pathogen of fish, is phylogenetically closest to the nonpathogenic environmental *Arthrobacter* species [51]. Based on 16S rRNA phylogenetic analysis, *R. salmoninarum* has been included in the actinomycetes subdivision and was found related to a subgroup harboring morphologically and chemotaxonomically rather heterogeneous taxa, including *Arthrobacter, Micrococcus, Cellulomonas, Jonesia, Promicromonospora, Stomatococcus*, and *Brevibacterium* [111]. In fact, *Arthrobacter davidanieli* is commercially used as a vaccine (commercially known as Renogen) and can provide significant cross-protection in Atlantic salmon, though not in Pacific salmon [112]. The genome sequencing of *R. salmoninarum* ATCC 33209 strain and two *Arthrobacter* strains, the TC1 and FB24, has revealed many interesting aspects of how this obligates fish pathogen evolved, via genomic reduction and horizontal gene acquisition, from members of the nonpathogenic genus *Arthrobacter* [51, 113]. A total of 1562 ORF clusters were similar in *R. salmoninarum* and *Arthrobacter* spp. demonstrating the genetic basis for the efficiency and cross-protection of the *A. davidanieli* vaccine.

There is significant genome reduction in *R. salmoninarum* genome, which is 1.44 Mb smaller than the chromosome of TC1 and 1.55 Mb smaller than the chromosome of FB24. The two *Arthrobacter* strains have several large plasmids that are not present in the ATCC 33209 strain. In addition, these plasmids do not have high levels of similarity to sequences in the *R. salmoninarum* chromosome [51].

The presence of many IS elements, pseudogenes, and genomic islands in *R. salmoninarum* genome coupled with a lack of restriction-modification systems contribute to the extensive disruption of ORFs as a strategy to reduce many pathways in the bacteria. Moreover, the highly homogeneous nature of *R. salmoninarum* with respect to the overall genomic structure, biochemical properties, and surface antigens [114, 115] points to the evolution of this pathogen towards a strictly intracellular life style.

Several virulence factors including capsular synthesis genes, heme acquisition operons, genes encoding possible hemolysins, and the poorly characterized *msa* genes identified in the *R. salmoninarum* genome seems to be horizontally acquired. *Arthrobacter* spp. lacks most of these gene sequences, thus underlining the differential evolution and adaptation of these two very closely related species to contrasting ecological niches.

9. *Streptococcus and Lactococcus*

Gram-positive cocci belonging to the genera *Streptococcus and Lactococcus* are increasingly being recognized as important fish pathogens all over the world [116]. There are several different species of Gram-positive cocci, including *Streptococcus parauberis*, *S. iniae*, *S. agalactiae* (syn. *Streptococcus difficilis*), *S. phocae* [117, 118], *Lactococcus garvieae* (syn. *Enterococcus seriolicida)* [119], *L. piscium* [120–123], *Vagococcus salmoninarum*, and *Carnobacterium piscicola* [124], implicated in infectious diseases of warm water as well as cold water fishes.

Streptococcosis appears to have very few limitations in regard to geographic boundaries or host range, with outbreaks occurring in aquaculture facilities worldwide and in many different cultured species. *S. iniae, S.parauberis*, *S. agalactiae*, and *L. garvieae* are known as the major pathogens of streptococcosis and lactococcosis in *Oncorhynchus mykiss*, *Seriola quinqueradiata, Siganus canaliculatus*, and *Tilapia* spp. [125]. *Recently, S. iniae* and *L. garvieae* are also recognized as emerging zoonotic pathogens, causing diseases in both fish and human beings [23, 126].

S. iniae is a β-haemolytic, Gram-positive coccus that causes generalized septicaemia and meningoencephalitis in a variety of warm water fishes [127], whereas *S. parauberis* is an α-hemolytic, Gram-positive coccus, mainly pathogenic in cultured turbot (*Scophthalmus maximus*) and olive flounder, *Paralichthys olivaceus*. *L. garvieae* causes a hyperacute and haemorrhagic septicemia in fishes particularly during the summer time. General pathological symptoms of streptococcosis and lactococcosis in fishes are hemorrhage, congestion, lethargy, dark pigmentation, erratic swimming, and exophthalmos with clouding of the cornea [117, 128].

Complete genome sequences of different strains of *S. parauberis* and *L. garvieae*, important pathogenic species isolated from both fish and human, have been published [129–132].

S. parauberis is recognized as the dominant etiological agent of streptococcosis in fish [117], whereas both *S. parauberis* and *S. uberis* are involved the causation of bovine mastitis in dairy cow [133, 134].

S. parauberis is closer to *S. uberis* than with other *Streptococcus* spp. and is biochemically and serologically indistinguishable from *S. uberis* [135]. Both species were earlier considered as type I and II of *S. uberis*, but later shown to be phylogenetically distinct and renamed the type I as *S. uberis* and type II as *S. parauberis* [134].

The *S. parauberis* strain KCTC11537BP genome size falls in the middle of the 1.8 to 2.3 Mb range of streptococcal genomes sequenced to date and the average G+C content of 35.6% is significantly lower than those of *S. pyogenes* [132]. About 78% of genes are shared between the genomes of *S. parauberis* strain KCTC11537BP and *S. uberis* NC_012004, but they differ significantly at two regions of the genome, demonstrating the genomic basis for their separation into two species.

S. parauberis genome encodes an M-like protein of *S. iniae* (SiM), which is an important virulence factor in *S. iniae* [136]. It also encodes *has*A and *has*B genes that may be involved in capsule production for resistance against phagocytosis. The genome analysis indicates that *S. parauberis* could possibly possess the ability to regulate the metabolism of more carbohydrates than other *Streptococcus* species and to synthesize all the aminoacids and regulatory factors required to adapt and survive in a highly hostile host environment.

Complete genome sequences of *L. garvieae* strain UNIUD074, isolated from diseased rainbow trout in Italy, a virulent strain Lg2 (serotype KG2) and a nonvirulent strain ATCC 49156 (serotype KG+), both isolated from diseased yellowtail in Japan have recently been published [130, 131]. In addition, genome sequence of *L. garvieae* strain 21881, isolated from a man suffering from septicemia has been published [129].

The strains Lg2 and ATCC 49156 have 99% sequence identity and share 1944 orthologous genes, but are different in 24 Lg2-specific genes that were absent in the ATCC 49156 genome. One of the Lg2-specific genes is a 16.5 kb capsule gene cluster, which confirms the earlier transmission electron microscopic finding that Lg2 is encapsulated, and ATCC 49156 is nonencapsulated [137]. In fact, the capsule gene cluster has the features of a horizontally acquired genomic island conferring virulence to the Lg2 strain but might have been lost from the ATCC 49156 strain while subculturing in the laboratory [131]. Both genomes carried three types of IS elements, prophage sequences, and integrase genes and were found smaller than those of at least five sequenced *L. lactis* genomes. The Lg2 genome lacks several aminoacid biosynthesis genes, which is a characteristic feature of pathogenic bacteria with reduced genomes. The Lg2 strain contains hemolysins, NADH oxidase and superoxide dismutase (SOD), adhesins and sortase, which are known virulence factors [137–139]. It also encodes a gene for phosphoglucomutase, a virulence factor conferring the resistance to peptide antimicrobials in *S. iniae* [140].

Although *L. garvieae* and *L. lactis* genomes share 75% CDS, about 25% genes are Lg2-specific hypothetical proteins and proteins of unknown functions, which may be involved in the virulence of the Lg2 strain. These findings indicate that *L. garvieae* and *L. lactis* have significantly diverged from the common ancestor, and the *L. garvieae* is evolving into a pathogenic species equipped with virulence features suitable for living in the host environment.

10. Mycobacteria

Chronic infections in fish caused by different species of mycobacteria have been well recognized [23, 141, 142]. Several slow growing as well as fast growing species of mycobacteria such as *Mycobacterium marinum*, *M. fortuitum*, *M. chelonae*, and *M. avium* have been isolated from wild and cultured fish suffering from mycobacteriosis in different parts of the world [143–145]. Among them, *M. marinum* is the most important fish pathogen, frequently isolated from a variety of fish species with granulomas [146]. It is also a known zoonotic pathogen, transmitted to man though fish handling in aquariums and aquaculture tanks, producing superficial and self-limiting lesions called "fish tank or aquarium tank granuloma" involving the cooler parts of the body such as hands, forearms, elbows, and knees [147, 148]. Although strain variation has been reported [149], there is significant intraspecies sequence homogeneity among different *M. mrinum* strains [150]. However, it is hypothesized that only certain strains of *M. marinum* have zoonotic potential [151]. Phylogenetic studies have shown that *M. marinum* is most closely related to *M. ulcerans* followed by *M. tuberculosis* [150]. Owing to this, *M. marinum* and *M. tuberculosis* share many virulence factors and significant pathological features and respond to similar antibiotics [152, 153]. Hence, *M. marinum* is also an important model organism to study the pathogenesis of tuberculosis [152, 153].

Interestingly, the genome of *M. marinum* is 50% bigger than that of *M. tuberculosis* and seems to have acquired a number of genes encoding NRPSs and the huge repertoire of PE, PPE, and ESX systems probably by HGT [154]. Both species might have evolved differently from a common environmental mycobacteria. *M. tuberculosis* might have adapted to its host intracellular life by extensive genome reduction and *M. marinum*, by and large retained or obtained genes required for its dual lifestyle and broad-host range.

11. Genome Sequencing to Find Novel Vaccine and Drug Targets in Fish Pathogens

Our understanding of the molecular basis of virulence of certain well-studied fish bacterial pathogens has increased dramatically during the past decade. This has resulted from the application of recombinant DNA technology and cell biology to investigate bacterial infections, and the development of genetic techniques for identifying virulence genes.

More recently, genome sequence information of several bacterial fish pathogens has become available from genome sequencing projects. There is strong reason to believe that this understanding will be exploited to develop new interventions against fish bacterial infections.

The relevance of sequencing projects for drug and vaccine discovery is obvious. During the "pregenomic" era, the vaccine candidate genes were individually identified by tedious gene knockout studies and virulence attenuation. But now, the complete genome sequencing provides information on every virulence gene and all potential vaccine candidates, and the sequence databases will become indispensable for research in fish vaccinology and drug development.

After sequencing, the open reading frames (ORFs) are searched against available databases for sequence similarity with genes of known functions in other organisms. There are several strategies for gene annotation employing the tools of predictive bioinformatics programs combined with analyses of the published literature.

Multiple target vaccine candidate genes can be chosen and deleted simultaneously by various strategies including global transposon mutagenesis and gene replacement techniques [155, 156] to study their effect on virulence and essentiality. A number of important virulence determinants identified in the sequenced genome can be targeted. For example, the sortase enzyme in Gram-positive fish pathogens would be a very attractive universal vaccine and therapeutic drug target, as it mediates covalent anchoring of many surface displayed antigenic and/or virulence related proteins in Gram-positive bacteria [139]. The inactivation or inhibition of the sortase enzyme can simultaneously prevent the surface display of a number of virulence factors, thus effectively attenuating the virulence of the pathogen [110, 157].

The availability of sequences of the complete surface antigenic repertoire of pathogens, including protein and noprotein antigens would facilitate strategies for rational design of vaccines and drugs. In addition, the recent availability of large collections of the "virulogenome" of fish bacterial pathogens will provide enormous virulence sequence information for DNA vaccination studies. The whole complement of IS elements, prophages, and pathogenicity islands that can harbor virulence, and antimicrobial resistance gene clusters can be easily identified in the genomes. The comparison of genomes of different strains of the same bacteria or closely related species can reveal how these strains or species behave differently while infecting fish hosts, thus opening exciting opportunities for functional genomic analysis of infection processes and pathogenesis. However, experimental validation of predicted functions of genes identified from sequencing projects has lagged far behind the speed of annotation, and the major challenge of researchers in the field today is to understand the functional framework of the sequenced genomes.

12. Conclusions

There has been a steady increase in the number of species of bacteria implicated in fish diseases. The common fish pathogenic bacterial species belong to the genera *Vibrio, Aeromonas, Flavobacterium, Yersinia, Edwardsiella, Streptococcus, lactococcus, Renibacterium*, *and Mycobacterium* [23]. However, there is growing indications that the pathogenic species spectrum as well as the geographic and host range is widening among fish pathogens [158–161], leading to the emergence of new pathogens. Unlike the situation in human and animal medicine, fish diseases pose unique and daunting challenges. Fish are always bathed in a continuous medium of water, and fish disease treatment is essentially a population medicine. In addition, the current treatment methods are largely ineffective, and the biology and genetics of most fish bacterial pathogens are poorly understood, limiting the application of modern science-based pathogen intervention strategies.

Rapid growth and expansion of genome sequencing of human and animal pathogens enabled better understanding of their biology, evolution, and host adaptation strategies, and helped in combating many major diseases. Unfortunately, such developments and progress in the genomics and functional genomics of fish pathogenic bacteria have been very slow. However, recent availability of cost-effective high-throughput sequencing technologies has set the pace of sequencing of more fish pathogenic bacteria. Genome sequencing of a number of important bacterial pathogens of fish has helped us to better understand their biology and genetics. The sequencing projects have unearthed exciting new information on the adaptive evolution of fish pathogens, for example, how the nonpathogenic and ubiquitous soil bacteria such as *Arthrobacter* sp. has evolved into a strictly obligate fish pathogen, *R. salmoninarum*, by shedding functional genes through genomic reduction to lead to a very cosy intracellular life style.

On the other hand, phenotypically similar strains of the same species differ in certain set of virulence gene clusters, acquired through HGT and become highly virulent. The capsule gene cluster in the *L. garvieae* Lg2 strain confers virulence compared to noncapsulated ATCC 49156, which lacks the gene cluster. Nonpathogenic strains acquire genomic islands from distantly related pathogenic species and emerge as new pathogens of fish.

Comparative pathogenomics of closely related bacteria has increased our knowledge of how they vary in their virulence and their ability to adapt to different ecological niches. This is clearly evident in the difference in virulence of various strains of *V. anguillarum* and *V. vulnificus,* and among the closely related species of the genus *Flavobacterium*. As more strain-specific sequence information on bacterial pathogens of fish becomes available, we will have a better understanding of the subtle genomic differences among strains with varying virulence characteristics.

The typical pathogen evolutionary strategy of acquiring, shuffling and shedding genes mediated by IS elements, pseudogenes, prophage sequences, and HGT is also observed in most bacterial pathogens of fish. It is certain that the new genomic information will bring paradigm changes in bacterial pathogenesis and should provide new perspectives to our current thinking on the evolutionary and adaptive strategies of aquatic bacteria and how they colonize and establish in wider ecological niches and new host species. Moreover, the identification of key virulence factors in pathogenic strains should help us design efficient drugs and vaccines to combat major bacterial pathogens of fish.

However, it should be stressed that the genomic information will provide only a snapshot of the microorganism. Highly virulent clones armed with one or more acquired virulence factors can suddenly develop from the existing harmless microorganisms in the face of environmental, antibiotic, and host-induced selective pressures.

More intriguingly, about 40% of the genes in sequenced bacterial genomes constitute new putative genes and hypothetical proteins with mysterious functions and are conserved among several different species of bacteria. Even in *Escherichia coli,* the most studied of all bacteria, only 54% genes have currently been functionally characterized based on experimental evidence [162]. A close scrutiny of the sequenced genomes of fish pathogens reveals that the above situation is essentially true for these pathogens as well. Although current advances in functional genomics, structural genomics and bioinformatics have contributed immensely to deciphering and extracting useful biological information from the vast genomic data, understanding and assigning functionality to the unique and new gene sequences discovered in the genomes will be the major task of genome biologists in the coming years.

Acknowledgments

This work was financially supported by various research fundings from the Ministry of Agriculture and Fisheries Wealth, Oman.

References

[1] J. M. Rothberg and J. H. Leamon, "The development and impact of 454 sequencing," *Nature Biotechnology*, vol. 26, no. 10, pp. 1117–1124, 2008.

[2] J. Zhang, R. Chiodini, A. Badr, and G. Zhang, "The impact of next-generation sequencing on genomics," *Journal of Genetics and Genomics*, vol. 38, no. 3, pp. 95–109, 2011.

[3] E. Pennisi, "Genome data shake tree of life," *Science*, vol. 280, no. 5364, pp. 672–674, 1998.

[4] W. F. Doolittle, "Phylogenetic classification and the universal tree," *Science*, vol. 284, no. 5423, pp. 2124–2128, 1999.

[5] W. F. Doolittle, "Lateral genomics," *Trends in Cell Biology*, vol. 9, no. 12, pp. M5–M8, 1999.

[6] E. Pennisi, "Is it time to uproot the tree of life?" *Science*, vol. 284, no. 5418, pp. 1305–1307, 1999.

[7] W. F. Doolittle, "Uprooting the tree of life," *Scientific American*, vol. 282, no. 2, pp. 90–95, 2000.

[8] E. V. Koonin, K. S. Makarova, and L. Aravind, "Horizontal gene transfer in prokaryotes: quantification and classification," *Annual Review of Microbiology*, vol. 55, pp. 709–742, 2001.

[9] N. A. Moran, "Microbial minimalism: genome reduction in bacterial pathogens," *Cell*, vol. 108, no. 5, pp. 583–586, 2002.

[10] A. Mira, L. Klasson, and S. G. E. Andersson, "Microbial genome evolution: sources of variability," *Current Opinion in Microbiology*, vol. 5, no. 5, pp. 506–512, 2002.

[11] L. Holm, "Codon usage and gene expression," *Nucleic Acids Research*, vol. 14, no. 7, pp. 3075–3087, 1986.

[12] L. S. Frost, R. Leplae, A. O. Summers, and A. Toussaint, "Mobile genetic elements: the agents of open source evolution," *Nature Reviews Microbiology*, vol. 3, no. 9, pp. 722–732, 2005.

[13] U. Dobrindt, B. Hochhut, U. Hentschel, and J. Hacker, "Genomic islands in pathogenic and environmental microorganisms," *Nature Reviews Microbiology*, vol. 2, no. 5, pp. 414–424, 2004.

[14] U. Dobrindt and J. Hacker, "How bacterial pathogens were constructed," in *Bacterial Virulence: Basic Principles, Models and Global Approaches*, P. Sansonetti, Ed., pp. 3–15, Wiley-VCH, GmbH & Co. KGaA, Weinheim, Germany, 2010.

[15] M. J. Pallen and B. W. Wren, "Bacterial pathogenomics," *Nature*, vol. 449, no. 7164, pp. 835–842, 2007.

[16] J. G. Lawrence and J. R. Roth, "Genomic flux: genome evolution by gene loss and acquisition," in *Organization of the Prokaryotic Genome*, R. L. Charlebois, Ed., pp. 263–289, American Society for Microbiology, Washington, DC, USA, 1999.

[17] F. P. Meyer, "Aquaculture disease and health management.," *Journal of Animal Science*, vol. 69, no. 10, pp. 4201–4208, 1991.

[18] C. D. Harvell, K. Kim, J. M. Burkholder et al., "Emerging marine diseases—climate links and anthropogenic factors," *Science*, vol. 285, no. 5433, pp. 1505–1510, 1999.

[19] R. Subasinghe, "Fish health and quarantine; review of the state of the World aquaculture," in *FAO Fisheries Circular*, no. 886, pp. 45–49, Rome, Italy, Food and Agriculture Organization of the United Nations, 1997.

[20] B. Hjeltnes and R. J. Roberts, "Vibriosis," in *Bacterial Diseases of Fish*, R. J. Roberts, N. R. Bromag, and V. Inglis, Eds., pp. 109–121, Blackwell Scientific, Oxford, UK, 1993.

[21] D. V. Lightner and R. M. Redman, "Shrimp diseases and current diagnostic methods," *Aquaculture*, vol. 164, no. 1–4, pp. 201–220, 1998.

[22] O. Bergh, F. Nilsen, and O. B. Samuelsen, "Diseases, prophylaxis and treatment of the Atlantic halibut *Hippoglossus hippoglossus*: a review," *Diseases of Aquatic Organisms*, vol. 48, no. 1, pp. 57–74, 2001.

[23] B. Austin and D. A. Austin, *Bacterial Fish Pathogens: Disease in Farmed and Wild Fish*, Springer, New York, NY, USA, 3rd edition, 1999.

[24] B. Austin and X. H. Zhang, "*Vibrio harveyi*: a significant pathogen of marine vertebrates and invertebrates," *Letters in Applied Microbiology*, vol. 43, no. 2, pp. 119–124, 2006.

[25] C. Amaro, E. G. Biosca, C. Esteve et al., "Comparative study of phenotypic and virulence properties in *Vibrio vulnificus* biotypes 1 and 2 obtained from a European eel farm experiencing mortalities," *Diseases of Aquatic Organisms*, vol. 13, pp. 29–35, 1992.

[26] E. G. Biosca, H. Llorens, E. Garay, and C. Amaro, "Presence of a capsule in *Vibrio vulnificus* biotype 2 and its relationship to virulence for eels," *Infection and Immunity*, vol. 61, no. 5, pp. 1611–1618, 1993.

[27] E. Benediktsdóttir, L. Verdonck, C. Spröer, S. Helgason, and J. Swings, "Characterization of *Vibrio viscosus* and *Vibrio wodanis* isolated at different geographical locations: a proposal for reclassification of *Vibrio viscosus* as *Moritella viscosa* comb. nov.," *International Journal of Systematic and Evolutionary Microbiology*, vol. 50, no. 2, pp. 479–488, 2000.

[28] L. A. Actis, M. E. Tolmasky, and J. H. Crosa, "Vibriosis," in *Fish Diseases and Disorders, Viral, Bacterial, and Fungal Infections*, P. T. K. Woo and D. W. Bruno, Eds., vol. 3, pp. 570–605, CABI International, Oxfordshire, UK, 2nd edition, 2011.

[29] E. Hjerde, M. Lorentzen, M. T. G. Holden et al., "The genome sequence of the fish pathogen *Aliivibrio salmonicida* strain LFI1238 shows extensive evidence of gene decay," *BMC Genomics*, vol. 9, article 616, 2008.

[30] P. A. Gulig, V. D. Crécy-Lagard, A. C. Wright, B. Walts, M. Telonis-Scott, and L. M. McIntyre, "SOLiD sequencing of four *Vibrio vulnificus* genomes enables comparative genomic analysis and identification of candidate clade-specific virulence genes," *BMC Genomics*, vol. 11, no. 1, article 512, 2010.

[31] H. Naka, G. M. Dias, C. C. Thompson, C. Dubay, F. L. Thompson, and J. H. Crosa, "Complete genome sequence of the marine fish pathogen *Vibrio anguillarum* harboring the pJM1 virulence plasmid and genomic comparison with other virulent strains of *V. anguillarum* and *V. ordalii*," *Infection and Immunity*, vol. 79, no. 7, pp. 2889–2900, 2011.

[32] J. L. Larsen, K. Pedersen, and I. Dalsgaard, "*Vibrio anguillarum* serovars associated with vibriosis in fish," *Journal of Fish Diseases*, vol. 17, no. 3, pp. 259–267, 1994.

[33] U. B. S. Sorensen and J. L. Larsen, "Serotyping of *Vibrio anguillarum*," *Applied and Environmental Microbiology*, vol. 51, no. 3, pp. 593–597, 1986.

[34] A. E. Toranzo and J. L. Barja, "A review of the taxonomy and seroepizootiology of *Vibrio anguillarum*, with special reference to aquaculture in the NorthWest Spain," *Diseases of Aquatic Organisms*, vol. 9, pp. 73–82, 1990.

[35] J. H. Crosa, "A plasmid associated with virulence in the marine fish pathogen *Vibrio anguillarum* specifies an iron-sequestering system," *Nature*, vol. 284, no. 5756, pp. 566–568, 1980.

[36] A. E. Toranzo, J. L. Barja, S. A. Potter et al., "Molecular factors associated with virulence of marine vibrios isolated from striped bass in Chesapeake Bay," *Infection and Immunity*, vol. 39, no. 3, pp. 1220–1227, 1983.

[37] M. H. Schiewe, T. J. Trust, and J. H. Crosa, "*Vibrio ordalii* sp. nov.: a causative agent of vibriosis in fish," *Current Microbiology*, vol. 6, no. 6, pp. 343–348, 1981.

[38] D. P. Ransom, C. N. Lannan, J. S. Rohovec, and J. L. Fryer, "Comparison of histopathology caused by *Vibrio anguillarum* and *Vibrio ordalii* in three species of Pacific salmon," *Journal of Fish Diseases*, vol. 7, pp. 107–115, 1984.

[39] K. Makino, K. Oshima, K. Kurokawa et al., "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*," *The Lancet*, vol. 361, no. 9359, pp. 743–749, 2003.

[40] N. Okada, S. Matsuda, J. Matsuyama et al., "Presence of genes for type III secretion system 2 in *Vibrio mimicus* strains," *BMC Microbiology*, vol. 10, article 302, 2010.

[41] A. Alam, K. A. Miller, M. Chaand, J. S. Butler, and M. Dziejman, "Identification of *Vibrio cholerae* type III secretion system effector proteins," *Infection and Immunity*, vol. 79, no. 4, pp. 1728–1740, 2011.

[42] C. R. Osorio, S. Juiz-Rio, and M. L. Lemos, "A siderophore biosynthesis gene cluster from the fish pathogen *Photobacterium damselae* subsp. *piscicida* is structurally and functionally related to the *Yersinia* high-pathogenicity island," *Microbiology*, vol. 152, no. 11, pp. 3327–3341, 2006.

[43] E. S. Yip, B. T. Grublesky, E. A. Hussa, and K. L. Visick, "A novel, conserved cluster of genes promotes symbiotic colonization and σ-dependent biofilm formation by *Vibrio fischeri*," *Molecular Microbiology*, vol. 57, no. 5, pp. 1485–1498, 2005.

[44] H. S. Kim, S. J. Park, and K. H. Lee, "Role of NtrC-regulated exopolysaccharides in the biofilm formation and pathogenic interaction of *Vibrio vulnificus*," *Molecular Microbiology*, vol. 74, no. 2, pp. 436–453, 2009.

[45] E. Sanjuan, B. Fouz, J. D. Oliver, and C. Amaro, "Evaluation of genotypic and phenotypic methods to distinguish clinical from environmental *Vibrio vulnificus* strains," *Applied and Environmental Microbiology*, vol. 75, no. 6, pp. 1604–1613, 2009.

[46] N. Bisharat, D. I. Cohen, M. C. Maiden, D. W. Crook, T. Peto, and R. M. Harding, "The evolution of genetic structure in the marine pathogen, *Vibrio vulnificus*," *Infection, Genetics and Evolution*, vol. 7, no. 6, pp. 685–693, 2007.

[47] M. B. Schrøder, S. Espelid, and T. Ø. Jørgensen, "Two serotypes of *Vibrio salmonicida* isolated from diseased cod (*Gadus morhua* L.); virulence, immunological studies and vaccination experiments," *Fish and Shellfish Immunology*, vol. 2, no. 3, pp. 211–221, 1992.

[48] M. Tomich, P. J. Planet, and D. H. Figurski, "The tad locus: postcards from the widespread colonization island," *Nature Reviews Microbiology*, vol. 5, no. 5, pp. 363–375, 2007.

[49] P. Siguier, J. Filee, and M. Chandler, "Insertion sequences in prokaryotic genomes," *Current Opinion in Microbiology*, vol. 9, no. 5, pp. 526–531, 2006.

[50] J. Parkhill, G. Dougan, K. D. James et al., "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18," *Nature*, vol. 413, no. 6858, pp. 848–852, 2001.

[51] G. D. Wiens, D. D. Rockey, Z. Wu et al., "Genome sequence of the fish pathogen *Renibacterium salmoninarum* suggests reductive evolution away from an environmental *Arthrobacter* ancestor," *Journal of Bacteriology*, vol. 190, no. 21, pp. 6970–6982, 2008.

[52] A. Martin-Carnahan and S. W. Joseph, "Aeromonadaceae," in *Bergey's Manual of Systematic Bacteriology*, G. M. Garrity, Ed., vol. 2, pp. 556–580, Springer, New York, NY, USA, 2nd edition, 2005.

[53] M. J. Figueras, "Clinical relevance of *Aeromonas* sM503," *Reviews in Medical Microbiology*, vol. 16, no. 4, pp. 145–153, 2005.

[54] R. Seshadri, S. W. Joseph, A. K. Chopra et al., "Genome sequence of *Aeromonas hydrophila* ATCC 7966T: jack of all trades," *Journal of Bacteriology*, vol. 188, no. 23, pp. 8272–8282, 2006.

[55] M. E. Reith, R. K. Singh, B. Curtis et al., "The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: Insights into the evolution of a fish pathogen," *BMC Genomics*, vol. 9, article 427, 2008.

[56] Y. Li, Y. Liu, Z. Zhou et al., "Complete genome sequence of *Aeromonas veronii* strain B565," *Journal of Bacteriology*, vol. 193, no. 13, pp. 3389–3390, 2011.

[57] S. A. Beatson, M. D. G. de Luna, N. L. Bachmann et al., "Genome sequence of the emerging pathogen *Aeromonas caviae*," *Journal of Bacteriology*, vol. 193, no. 5, pp. 1286–1287, 2011.

[58] E. M. Bernoth, "Furunculosis: the history of the diseases and of disease research," in *Furunculosis, Multidisciplinary Fish Disease Research*, E. M. Bernoth, A. E. Ellis, P. J. Midtlyng, G. Olivier, and P. Smith, Eds., pp. 1–20, Academic Press, London, UK, 1997.

[59] M. Hiney and G. Olivier, "Furunculosis (*Aeromonas salmonicida*)," in *Fish Diseases and Disorders III: Viral, Bacterial and Fungal Infections*, P. T. K. Woo and D. W. Bruno, Eds., pp. 341–425, CAB Publishing, Oxford, UK, 1999.

[60] C. Paniagua, O. Rivero, J. Anguita, and G. Naharro, "Pathogenicity factors and virulence for rainbow trout (*Salmo gairdneri*) of motile *Aeromonas* spp. isolated from a river," *Journal of Clinical Microbiology*, vol. 28, no. 2, pp. 350–355, 1990.

[61] S. E. Burr, D. Pugovkin, T. Wahli, H. Segner, and J. Frey, "Attenuated virulence of an *Aeromonas salmonicida* subsp. *salmonicida* type III secretion mutant in a rainbow trout model," *Microbiology*, vol. 151, no. 6, pp. 2111–2118, 2005.

[62] J. Sha, L. Pillai, A. A. Fadl, C. L. Galindo, T. E. Erova, and A. K. Chopra, "The type III secretion system and cytotoxic enterotoxin alter the virulence of *Aeromonas hydrophila*," *Infection and Immunity*, vol. 73, no. 10, pp. 6446–6457, 2005.

[63] J. M. Janda and R. P. Kokka, "The pathogenicity of *Aeromonas* strains relative to genospecies and phenospecies identification," *FEMS Microbiology Letters*, vol. 90, no. 1, pp. 29–33, 1991.

[64] J. Madetoja, I. Dalsgaard, and T. Wiklund, "Occurrence of *Flavobacterium psychrophilum* in fish-farming environments," *Diseases of Aquatic Organisms*, vol. 52, no. 2, pp. 109–118, 2002.

[65] A. Nematollahi, A. Decostere, F. Pasmans, and F. Haesebrouck, "*Flavobacterium psychrophilum* infections in salmonid fish," *Journal of Fish Diseases*, vol. 26, no. 10, pp. 563–574, 2003.

[66] J. G. Leahy and R. R. Colwell, "Microbial degradation of hydrocarbons in the environment," *Microbiological Reviews*, vol. 54, no. 3, pp. 305–315, 1990.

[67] J. Lehmann, D. Mock, F. J. Stuerenberg, and J. F. Bernardet, "First isolation of *Cytophaga psychrophila* from a systemic disease in eel and cyprinids," *Diseases of Aquatic Organisms*, vol. 10, pp. 217–220, 1991.

[68] R. A. Holt, *Cytophaga psychrophila, the causative agent of bacterial cold-water disease in salmonid fish*, Ph.D. thesis, Oregon State University, Corvallis, Ore, USA, 1987.

[69] G. J. Heo, K. Kasai, and H. Wakabayashi, "Occurrence of *Flavobacterium branchiophila* associated with bacterial gill disease at a trout hatchery," *Fish Pathology*, vol. 25, pp. 99–105, 1990.

[70] V. E. Ostland, J. S. Lumsden, D. D. MacPhee, and H. W. Ferguson, "Characteristics of *Flavobacterium branchiophilum*, the cause of salmonid bacterial gill disease in Ontario," *Journal of Aquatic Animal Health*, vol. 6, no. 1, pp. 13–26, 1994.

[71] J. F. Bernardet, "*Flexibacter columnaris*: first description in France and comparison with bacterial strains from other origins," *Diseases of Aquatic Organisms*, vol. 6, pp. 37–44, 1989.

[72] D. J. Speare and H. W. Ferguson, "Clinical and pathological features of common gill diseases of cultured salmonids in Ontario," *Canadian Veterinary Journal*, vol. 30, pp. 882–887, 1989.

[73] J. F. Bernardet, P. Segers, M. Vancanneyt, F. Berthe, K. Kersters, and P. Vandamme, "Cutting a gordian knot: emended classification and description of the genus *Flavobacterium*, emended description of the family *Flavobacteriaceae*, and proposal of *Flavobacterium hydatis* nom. nov. (basonym, Cytophaga aquatilis Strohl and Tait 1978)," *International Journal of Systematic Bacteriology*, vol. 46, no. 1, pp. 128–148, 1996.

[74] E. Duchaud, M. Boussaha, V. Loux et al., "Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum*," *Nature Biotechnology*, vol. 25, no. 7, pp. 763–769, 2007.

[75] M. Touchon, P. Barbier, J. F. Bernardet et al., "Complete genome sequence of the fish pathogen *Flavobacterium branchiophilum*," *Applied and Environmentl Microbiology*, vol. 77, no. 21, pp. 7656–7662, 2011.

[76] M. L. Lawrence, A. Karsi, H. C. Tekedar et al., "Comparative genomics of *Flavobacterium columnare* isolates from two genetic divisions and with different pathogenic potential for channel catfish (*Ictalurus punctatus*)," in *Proceedings of the Conference on FLAVOBACTERIUM*, Paris, France, September 2009.

[77] C. Rodkhum, I. Hirono, J. H. Crosa, and T. Aoki, "Four novel hemolysin genes of *Vibrio anguillarum* and their virulence to rainbow trout," *Microbial Pathogenesis*, vol. 39, no. 4, pp. 109–119, 2005.

[78] M. Palmer, "The family of thiol-activated, cholesterol-binding cytolysins," *Toxicon*, vol. 39, no. 11, pp. 1681–1689, 2001.

[79] M. Shoji, D. B. Ratnayake, Y. Shi et al., "Construction and characterization of a nonpigmented mutant of *Porphyromonas gingivalis*: Cell surface polysaccharide as an anchorage for gingipanis," *Microbiology*, vol. 148, no. 4, pp. 1183–1191, 2002.

[80] K. Sato, E. Sakai, P. D. Veith et al., "Identification of a new membrane-associated protein that influences transport/maturation of gingipains and adhesins of *Porphyromonas gingivalis*," *The Journal of Biological Chemistry*, vol. 280, no. 10, pp. 8668–8677, 2005.

[81] S. L. Abbott and J. M. Janda, "The genus *Edwardsiella*," *Prokaryotes*, vol. 6, pp. 72–89, 2006.

[82] J. E. Clarridge, D. M. Musher, V. Fanstein, and R. J. Wallace Jr., "Extra-intestinal human infection caused by *Edwardsiella tarda*," *Journal of Clinical Microbiology*, vol. 11, no. 5, pp. 511–514, 1980.

[83] P. Nagel, A. Serritella, and T. J. Layden, "*Edwardsiella tarda* gastroenteritis associated with a pet turtle," *Gastroenterology*, vol. 82, no. 6, pp. 1436–1437, 1982.

[84] W. A. Rogers, "Edwardsiellosis in fishes," in *Antigens of Fish Pathogens. Les Antigenes des Microorganisms Pathogenes des Poissons*, D. P. Anderson, M. Dorson, and Ph. Dubourget, Eds., pp. 153–159, Fondation Marcel Merieux, Lyon, France, 1983.

[85] F. P. Meyer and G. L. Bullock, "*Edwardsiella tarda*, a new pathogen of channel catfish (*Ictalurus punctatus*)," *Applied Microbiology*, vol. 25, no. 1, pp. 155–156, 1973.

[86] S. H. M. Ling, X. H. Wang, L. Xie, T. M. Lim, and K. Y. Leung, "Use of green fluorescent protein (GFP) to study the invasion pathways of *Edwardsiella tarda* in *in vivo* and *in vitro* fish models," *Microbiology*, vol. 146, no. 1, pp. 7–19, 2000.

[87] J. C. Newton, L. G. Wolfe, J. M. Grizzle, and J. A. Plumb, "Pathology of experimental enteric septicaemia in channel catfish, *Ictalurus punctatus* (Rafinesque), following immersion-exposure to *Edwardsiella ictaluri*," *Journal of Fish Diseases*, vol. 12, no. 4, pp. 335–347, 1989.

[88] Q. Wang, M. Yang, J. Xiao et al., "Genome sequence of the versatile fish pathogen *Edwardsiella tarda* provides insights into its adaptation to broad host ranges and intracellular niches," *PLoS One*, vol. 4, no. 10, Article ID e7646, 2009.

[89] M. L. Williams, A. F. Gillaspy, D. W. Dyer et al., "Genome sequence of *Edwardsiella ictaluri* 93–146, a strain associated with a natural channel catfish outbreak of enteric septicemia of catfish," *Journal of Bacteriology*, vol. 194, no. 3, pp. 740–741, 2012.

[90] S. R. Partridge and R. M. Hall, "The IS1111 family members IS4321 and IS5075 have sub-terminal inverted repeats and target the terminal inverted repeats of Tn21 family transposons," *Journal of Bacteriology*, vol. 185, no. 21, pp. 6371–6384, 2003.

[91] T. Tanabe, T. Funahashi, H. Nakao, S. I. Miyoshi, S. Shinoda, and S. Yamamoto, "Identification and characterization of genes required for biosynthesis and transport of the siderophore vibrioferrin in *Vibrio parahaemolyticus*," *Journal of Bacteriology*, vol. 185, no. 23, pp. 6938–6949, 2003.

[92] Q. Wang, Q. Liu, Y. Ma, L. Zhou, and Y. Zhang, "Isolation, sequencing and characterization of cluster genes involved in the biosynthesis and utilization of the siderophore of marine fish pathogen *Vibrio alginolyticus*," *Archives of Microbiology*, vol. 188, no. 4, pp. 433–439, 2007.

[93] M. D. Furones, C. J. Rodgers, and C. B. Munn, "*Yersinia ruckeri*, the causal agent of enteric redmouth disease (ERM) in fish," *Annual Review of Fish Diseases*, vol. 3, no. C, pp. 105–125, 1993.

[94] R. R. Rucker, "Redmouth disease of rainbow trout (*Salmo gairdneri*)," *Bulletin de l'Office International des Epizooties*, vol. 65, no. 5, pp. 825–830, 1966.

[95] E. Tobback, A. Decostere, K. Hermans, F. Haesebrouck, and K. Chiers, "*Yersinia ruckeri* infections in salmonid fish," *Journal of Fish Diseases*, vol. 30, no. 5, pp. 257–268, 2007.

[96] M. T. Horne and A. C. Barnes, "Enteric redmouth disease (*Y. ruckeri*)," in *Fish Diseases and Disorders: Viral, Bacterial and Fungal Infections*, P. T. K. Woo and D. W. Bruno, Eds., vol. 3, pp. 455–477, CABI Publishing, Oxfordshire, UK, 1999.

[97] A. J. Ross, R. R. Rucker, and W. H. Ewing, "Description of a bacterium associated with redmouth disease of rainbow trout

(*Salmo gairdneri*)," *Canadian Journal of Microbiology*, vol. 12, no. 4, pp. 763–770, 1966.

[98] R. L. Davies, "Virulence and serum-resistance in different clonal groups and serotypes of *Yersinia ruckeri*," *Veterinary Microbiology*, vol. 29, no. 3-4, pp. 289–297, 1991.

[99] R. M. V. Stevenson and D. W. Airdrie, "Serological variation among *Yersinia ruckeri* strains," *Journal of Fish Diseases*, vol. 7, pp. 247–254, 1984.

[100] J. L. Romalde, B. Magarinos, J. L. Barja, and A. E. Toranzo, "Antigenic and molecular characterization of *Yersinia ruckeri* proposal for a new intraspecies classification," *Systematic and Applied Microbiology*, vol. 16, no. 3, pp. 411–419, 1993.

[101] R. M. Stevenson, "Immunization with bacterial antigens: yersiniosis," *Developments in Biological Standardization*, vol. 90, pp. 117–124, 1997.

[102] R. L. Davies, "*Yersinia ruckeri* produces four iron-regulated proteins but does not produce detectable siderophores," *Journal of Fish Diseases*, vol. 14, pp. 563–570, 1991.

[103] M. Kotetishvili, A. Kreger, G. Wauters, J. G. Morris, A. Sulakvelidze, and O. C. Stine, "Multilocus sequence typing for studying genetic relationships among *Yersinia* species," *Journal of Clinical Microbiology*, vol. 43, no. 6, pp. 2674–2684, 2005.

[104] P. E. Chen, C. Cook, A. C. Stewart et al., "Genomic characterization of the *Yersinia* genus," *Genome Biology*, vol. 11, no. 1, article R1, 2010.

[105] J. E. Sanders and J. L. Fryer, "*Renibacterium salmoninarum* gen. nov., sp. nov., the causative agent of bacterial kidney disease in salmonid fishes," *International Journal of Systematic Bacteriology*, vol. 30, no. 2, pp. 496–502, 1980.

[106] D. L. Mitchum and L. E. Sherman, "Transmission of bacterial kidney disease from wild to stocked hatchery trout," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 38, pp. 547–551, 1981.

[107] G. L. Bullock, *Bacterial Kidney Disease of Salmonid Fishes Caused by Renibacterium salmoninarum*, no. 60, U.S. Fish and Wildlife Service, Fish Disease Leaflet, Washington, DC, USA, 1980.

[108] K. E. Wolf and C. E. Dunbar, "Test of 34 therapeutic agents for control of kidney disease in trout," *Transactions of American Fisheries Society*, vol. 88, pp. 117–124, 1959.

[109] S. K. Gutenberger, J. R. Duimstra, J. S. Rohovec, and J. L. Fryer, "Intracellular survival of *Renibacterium salmoninarum* in trout mononuclear phagocytes," *Disease of Aquatic Organisms*, vol. 28, no. 2, pp. 93–106, 1997.

[110] P. S. Sudheesh, S. Crane, K. D. Cain, and M. S. Strom, "Sortase inhibitor phenyl vinyl sulfone inhibits *Renibacterium salmoninarum* adherence and invasion of host cells," *Diseases of Aquatic Organisms*, vol. 78, no. 2, pp. 115–127, 2007.

[111] E. Stackebrandt, U. Wehmeyer, H. Nader, and F. Fiedler, "Phylogenetic relationship of the fish pathogenic *Renibacterium salmoninarum* to *Arthrobacter*, *Micrococcus* and related taxa," *FEMS Microbiology Letters*, vol. 50, no. 2-3, pp. 117–120, 1988.

[112] K. Salonius, C. Siderakis, A. M. MacKinnon, and S. G. Griffiths, "Use of *Arthrobacter davidanieli* as a live vaccine against *Renibacterium salmoninarum* and *Piscirickettsia salmonis* in salmonids," *Developments in Biologicals*, vol. 121, pp. 189–197, 2005.

[113] E. F. Mongodin, N. Shapir, S. C. Daugherty et al., "Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1," *PLoS Genetics*, vol. 2, no. 12, article e214, 2006.

[114] R. G. Getchell, J. S. Rohovec, and J. L. Fryer, "Comparison of *Renibacterium salmoninarum* isolates by antigenic analysis," *Fish Pathology*, vol. 20, pp. 149–159, 1985.

[115] F. Fiedler and R. Draxl, "Biochemical and immunochemical properties of the cell surface of *Renibacterium salmoninarum*," *Journal of Bacteriology*, vol. 168, no. 2, pp. 799–804, 1986.

[116] T. Kitao, "Streptococcal infections," in *Bacterial Disease of Fish*, V. Inglis, R. J. Roberts, and N. R. Bromage, Eds., pp. 196–210, Blackwell Scientific, Oxford, UK, 1993.

[117] A. Domenech, J. F. Fernandez-Garayzabal, C. Pasqual et al., "Streptococcosis in cultured turbot, *Scophthalmus maximus* (L.), associated with *Streptococcus parauberis*," *Journal of Fish Diseases*, vol. 19, pp. 33–38, 1996.

[118] J. L. Romalde, C. Ravelo, I. Valdés et al., "*Streptococcus phocae*, an emerging pathogen for salmonid culture," *Veterinary Microbiology*, vol. 130, no. 1-2, pp. 198–207, 2008.

[119] D. Vendrell, J. L. Balcázar, I. Ruiz-Zarzuela, I. de Blas, O. Gironés, and J. L. Múzquiz, "*Lactococcus garvieae* in fish: a review," *Comparative Immunology, Microbiology and Infectious Diseases*, vol. 29, no. 4, pp. 177–198, 2006.

[120] M. D. Collins, J. A. E. Farrow, B. A. Phillips, and O. Kandler, "*Streptococcus garvieae* sp. nov. and *Streptococcus plantarum* sp. nov.," *Journal of General Microbiology*, vol. 129, no. 11, pp. 3427–3431, 1983.

[121] A. M. Williams, J. L. Fryer, and M. D. Collins, "*Lactococcus piscium* sp. nov. a new *Lactococcus* species from salmonid fish," *FEMS Microbiology Letters*, vol. 68, no. 1-2, pp. 109–113, 1990.

[122] A. Domenech, J. Prieta, J. F. Fernandez-Garayzabal et al., "Phenotypic and phylogenetic evidence for a close relationship between *Lactococcus garvieae* and *Enterococcus seriolicida*," *Microbiologia SEM*, vol. 9, no. 1, pp. 63–68, 1993.

[123] A. Eldar, C. Ghittino, L. Asanta et al., "*Enterococcus seriolicida* is a junior synonym of *Lactococcus garvieae*, a causative agent of septicemia and meningoencephalitis in fish," *Current Microbiology*, vol. 32, no. 2, pp. 85–88, 1996.

[124] S. Wallbanks, A. J. Martinez-Murcia, J. L. Fryer, B. A. Phillips, and M. D. Collins, "16S rRNA sequence determination for members of the genus *Carnobacterium* and related lactic acid bacteria and description of *Vagococcus salmoninarum* sp. nov.," *International Journal of Systematic Bacteriology*, vol. 40, no. 3, pp. 224–230, 1990.

[125] E. S. Bromage and L. Owens, "Infection of barramundi *Lates calcarifer* with *Streptococcus iniae*: effects of different routes of exposure," *Diseases of Aquatic Organisms*, vol. 52, no. 3, pp. 199–205, 2002.

[126] J. A. Elliott, M. D. Collins, N. E. Pigott, and R. R. Facklam, "Differentiation of *Lactococcus lactis* and *Lactococcus garvieae* from humans by comparison of whole-cell protein patterns," *Journal of Clinical Microbiology*, vol. 29, no. 12, pp. 2731–2734, 1991.

[127] R. P. Perera, S. K. Johnson, M. D. Collins, and D. H. Lewis, "*Streptococcus iniae* associated with mortality of *Tilapia nilotica* X *T.aurea* hybrids," *Journal of Aquatic Animal Health*, vol. 6, no. 4, pp. 335–340, 1994.

[128] R. Kusuda, K. Kawai, F. Salati, C. R. Banner, and J. L. Fryer, "*Enterococcus seriolicida* sp. nov., a fish pathogen," *International Journal of Systematic Bacteriology*, vol. 41, no. 3, pp. 406–409, 1991.

[129] M. Aguado-Urda, G. H. López-Campos, M. M. Blanco et al., "Genome sequence of *Lactococcus garvieae* 21881, isolated in a case of human septicemia," *Journal of Bacteriology*, vol. 193, no. 15, pp. 4033–4034, 2011.

[130] P. Reimundo, M. Pignatelli, L. D. Alcaraz, G. D'Auria, A. Moya, and J. A. Guijarro, "Genome sequence of *Lactococcus garvieae* UNIUD074, isolated in Italy from a lactococcosis outbreak," *Journal of Bacteriology*, vol. 193, no. 14, pp. 3684–3685, 2011.

[131] H. Morita, H. Toh, K. Oshima et al., "Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*," *PLoS One*, vol. 6, no. 8, Article ID e23184, 2011.

[132] S. W. Nho, J. I. Hikima, I. S. Cha et al., "Complete genome sequence and immunoproteomic analyses of the bacterial fish pathogen *Streptococcus parauberis*," *Journal of Bacteriology*, vol. 193, no. 13, pp. 3356–3366, 2011.

[133] J. L. Watts, "Etiological agents of bovine mastitis," *Veterinary Microbiology*, vol. 16, no. 1, pp. 41–66, 1988.

[134] A. M. Williams and M. D. Collins, "Molecular taxonomic studies on *Streptococcus uberis* types I and II. Description of *Streptococcus parauberis* sp. nov.," *Journal of Applied Bacteriology*, vol. 68, no. 5, pp. 485–490, 1990.

[135] R. W. Bentley, J. A. Leigh, and M. D. Collins, "Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences," *International Journal of Systematic Bacteriology*, vol. 41, no. 4, pp. 487–494, 1991.

[136] J. B. Locke, R. K. Aziz, M. R. Vicknair, V. Nizet, and J. T. Buchanan, "*Streptococcus iniae* M-like protein contributes to virulence in fish and is a target for live attenuated vaccine development," *PLoS One*, vol. 3, no. 7, Article ID e2824, 2008.

[137] M. Kawanishi, T. Yoshida, M. Kijima et al., "Characterization of *Lactococcus garvieae* isolated from radish and broccoli sprouts that exhibited a KG+ phenotype, lack of virulence and absence of a capsule," *Letters in Applied Microbiology*, vol. 44, no. 5, pp. 481–487, 2007.

[138] T. J. Mitchell, "The pathogenesis of streptococcal infections: from tooth decay to meningitis," *Nature Reviews. Microbiology*, vol. 1, no. 3, pp. 219–230, 2003.

[139] L. A. Marraffini, A. C. Dedent, and O. Schneewind, "Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria," *Microbiology and Molecular Biology Reviews*, vol. 70, no. 1, pp. 192–221, 2006.

[140] J. T. Buchanan, J. A. Stannard, X. Lauth et al., "*Streptococcus iniae*phosphoglucomutase is a virulence factor and a target for vaccine development," *Infection and Immunity*, vol. 73, no. 10, pp. 6935–6944, 2005.

[141] G. N. Frerichs, "Acid-fast fish pathogens," in *Bacterial Diseases of Fish*, V. Inglis, R. J. Roberts, and N. R. Bromage, Eds., pp. 217–233, Blackwell Scientific, Oxford, UK, 1993.

[142] E. J. Noga, *Fish disease diagnosis and treatment*, Mosby, St. Louis, Mo, USA, 1996.

[143] L. G. Wayne, G. P. Kubica et al., "Genus *Mycobacterium* Lehmann and Neumann 1896, 363AL," in *Bergey's Manual of Systematic Bacteriology*, P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt, Eds., vol. 2, pp. 1436–1457, Williams & Wilkins, Baltimore, Md, USA, 1986.

[144] C. O. Thoen and T. A. Schliesser, "Mycobacterial infections in cold-blooded animals," in *The Mycobacteria*, G. P. Kubica and L. G. Wayne, Eds., part B, p. 1471, 1st edition, 1984.

[145] G. E. Sanders and L. E. Swaim, "Atypical piscine Mycobacteriosis in Japanese medaka (*Oryzias latipes*)," *Comparative Medicine*, vol. 51, no. 2, pp. 171–175, 2001.

[146] M. K. Stoskopf, *Fish Medicine*, WB Saunders, Philadelphia, Pa, USA, 1993.

[147] D. T. Smith and H. P. Willett, "Other *Mycobacterium* species," in *Zinsser Microbiology*, K. Joklik, H. P. Willet, and D. B. Amos, Eds., pp. 674–698, Appleton-Century-Crofts, New York, NY, USA, 17th edition, 1980.

[148] D. Huminer, S. D. Pitlik, C. Block et al., "Aquarium-borne *Mycobacterium marinum* skin infection. Report of a case and review of the literature," *Archives of Dermatology*, vol. 122, no. 6, pp. 698–703, 1986.

[149] M. Ucko, A. Colorni, H. Kvitt, A. Diamant, A. Zlotkin, and W. R. Knibb, "Strain variation in *Mycobacterium marinum* fish isolates," *Applied and Environmental Microbiology*, vol. 68, no. 11, pp. 5281–5287, 2002.

[150] M. J. Yip, J. L. Porter, J. A. M. Fyfe et al., "Evolution of *Mycobacterium ulcerans* and other mycolactone-producing mycobacteria from a common *Mycobacterium marinum* progenitor," *Journal of Bacteriology*, vol. 189, no. 5, pp. 2021–2029, 2007.

[151] M. Ucko and A. Colorni, "*Mycobacterium marinum* infections in fish and humans in Israel," *Journal of Clinical Microbiology*, vol. 43, no. 2, pp. 892–895, 2005.

[152] H. E. Volkman, H. Clay, D. Beery, J. C. W. Chang, D. R. Sherman, and L. Ramakrishnan, "Tuberculous granuloma formation is enhanced by a *Mycobacterium* virulence determinant," *PLoS Biology*, vol. 2, no. 11, article e367, 2004.

[153] L. E. Swaim, L. E. Connolly, H. E. Volkman, O. Humbert, D. E. Born, and L. Ramakrishnan, "*Mycobacterium marinum* infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity," *Infection and Immunity*, vol. 74, no. 11, pp. 6108–6117, 2006.

[154] T. P. Stinear, T. Seemann, P. F. Harrison et al., "Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*," *Genome Research*, vol. 18, no. 5, pp. 729–741, 2008.

[155] J. McFadden, "Recombination in mycobacteria," *Molecular Microbiology*, vol. 21, no. 2, pp. 205–211, 1996.

[156] N. R. Salama, B. Shepherd, and S. Falkow, "Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*," *Journal of Bacteriology*, vol. 186, no. 23, pp. 7926–7935, 2004.

[157] P. Cossart and R. Jonquières, "Sortase, a universal target for therapeutic agents against Gram-positive bacteria?" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 10, pp. 5013–5015, 2000.

[158] M. Maher, R. Palmer, F. Gannon, and T. Smith, "Relationship of a novel bacterial fish pathogen to *Streptobacillus monilifomris* and the fusobacteria group, based on 16S ribosomal RNA analysis," *Systematic and Applied Microbiology*, vol. 18, no. 1, pp. 79–84, 1995.

[159] J. L. Fryer and M. J. Mauel, "The rickettsia: an emerging group of pathogens in fish," *Emerging Infectious Diseases*, vol. 3, no. 2, pp. 137–144, 1997.

[160] T. H. Birkbeck, L. A. Laidler, A. N. Grant, and D. I. Cox, "*Pasteurella skyensis* sp. nov., isolated from Atlantic salmon (*Salmo salar* L.)," *International Journal of Systematic and Evolutionary Microbiology*, vol. 52, no. 3, pp. 699–704, 2002.

[161] C. Michel, J. F. Bernardet, P. Daniel, S. Chilmonczyk, M. Urdaci, and P. de Kinkelin, "Clinical and aetiological aspects of a summer enteritic syndrome associated with the sporulating segmented filamentous bacterium "*Candidatus Arthromitus*" in farmed rainbow trout, *Oncorhynchus mykiss* (Walbaum)," *Journal of Fish Diseases*, vol. 25, no. 9, pp. 533–543, 2002.

[162] M. Riley, T. Abe, M. B. Arnaud et al., "*Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005," *Nucleic Acids Research*, vol. 34, no. 1, pp. 1–9, 2006.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 860797, 8 pages
doi:10.1155/2012/860797

Research Article

Whole-Genome Profiling of a Novel Mutagenesis Technique Using Proofreading-Deficient DNA Polymerase δ

Yuh Shiwa,[1] Sanae Fukushima-Tanaka,[2] Ken Kasahara,[2] Takayuki Horiuchi,[2] and Hirofumi Yoshikawa[1,3]

[1] *Genome Research Center, NODAI Research Institute, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo 156-8502, Japan*
[2] *Neo-Morgan Laboratory Inc., 907 Nogawa, Miyamae-ku, Kawasaki, Kanagawa 216-0001, Japan*
[3] *Department of Bioscience, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo 156-8502, Japan*

Correspondence should be addressed to Hirofumi Yoshikawa, hiyoshik@nodai.ac.jp

Received 9 February 2012; Accepted 20 March 2012

Academic Editor: Hiromi Nishida

A novel mutagenesis technique using error-prone DNA polymerase δ (*polδ*), the disparity mutagenesis model of evolution, has been successfully employed to generate novel microorganism strains with desired traits. However, little else is known about the spectra of mutagenic effects caused by disparity mutagenesis. We evaluated and compared the performance of the *polδMKII* mutator, which expresses the proofreading-deficient and low-fidelity *polδ*, in *Saccharomyces cerevisiae* haploid strain with that of the commonly used chemical mutagen ethyl methanesulfonate (EMS). This mutator strain possesses exogenous mutant *polδ* supplied from a plasmid, tthereby leaving the genomic one intact. We measured the mutation rate achieved by each mutagen and performed high-throughput next generation sequencing to analyze the genome-wide mutation spectra produced by the 2 mutagenesis methods. The mutation frequency of the mutator was approximately 7 times higher than that of EMS. Our analysis confirmed the strong G/C to A/T transition bias of EMS, whereas we found that the mutator mainly produces transversions, giving rise to more diverse amino acid substitution patterns. Our present study demonstrated that the *polδMKII* mutator is a useful and efficient method for rapid strain improvement based on *in vivo* mutagenesis.

1. Introduction

Random mutagenesis is a powerful tool for generating enzymes, proteins, metabolic pathways, or even entire genomes with desired or improved properties [1]. Due to the technical simplicity and applicability to almost any organism, chemical or radiation mutagenesis is frequently used for the generation of genetic variability in a microorganism. However, these methods tend to be inefficient because they can cause substantial cell damage when performed *in vivo* [2].

A novel mutagenesis technique using error-prone DNA polymerase δ (*polδ*), based on the disparity mutagenesis model of evolution [3] has been successfully employed to generate novel microorganism strains with desired traits [4–11]. In the disparity model, mutations occur preferentially on the lagging strand, due to the more complex, discontinuous DNA replication that takes place there. Computer simulation shows that the disparity model accumulates more mutations than the parity model, in which mutations occur stochastically and evenly in both strands [3]. In addition, the disparity model produces greater diversity because some offspring will have mutant DNA while some offspring will have nonmutated, wild-type DNA.

Several studies have shown that the disparity mutagenesis method often achieved more satisfactory results (i.e., higher mutation rate and quick attainment of the desired phenotype) than conventional methods such as the chemical mutagen, ethyl methanesulfonate (EMS) [5, 10], which is known to produce mainly G/C to A/T transitions [12]. However, little else is known about the spectra of mutagenic effects caused by disparity mutagenesis.

polδ is involved in the synthesis of the lagging strand of DNA [13]. Several mutants, including the proofreading-deficient *pol3-01* strain and several low-fidelity mutants, have been shown to elevate the mutation rate [14–18]. To generate the strains with the greatest mutagenicity, Neo-Morgan Laboratory (Kanagawa, Japan) has developed the plasmid YCplac33/*polδMKII*, expressing the *polδ* mutant allele with 2 mutations: one mutation to inactivate the proofreading activity (D321A and E323A) [15] and another mutation to decrease the fidelity of replication (L612M) [14, 17, 18].

With the recent advent of next-generation sequencing technologies, an accurate characterization of the mutant genome, relative to the parental reference strain, is now achievable. In fact, Flibotte et al. have analyzed the mutation spectra induced by various mutagens, such as EMS, ENU, and UV/TMP, in the whole genome of *Caenorhabditis elegans* [12]. Another group has also used these sequencing technologies to analyze the genetic variations between a parental and EMS-mutagenized strain of yeast [19].

In this study, we evaluate the performance of the *polδMKII* mutator, which expresses the proofreading-deficient and low-fidelity *polδ* in *S. cerevisiae* haploid strain, compared with the commonly used chemical mutagen EMS. This mutator strain possesses exogenous mutant *polδ* supplied from a plasmid, thereby leaving the genomic one intact. We measured the mutation rate of this mutator strain and found that the mutation frequency of *polδMKII* was approximately 7 times higher than that of EMS. We also performed high-throughput next generation sequencing with Illumina GAII to analyze the genome-wide mutation spectra produced by the 2 different mutagenesis methods and found that the mutator strain exhibited more pleiotropy and gave rise to more diverse amino acid substitution patterns. Our present study has demonstrated that a proofreading-deficient and low-fidelity *polδMKII* mutator is a useful and efficient method for rapid strain improvement based on *in vivo* mutagenesis. This mutator is also useful for studying the acceleration of evolution.

2. Materials and Methods

2.1. Plasmid. Plasmid YCplac33/*polδMKII* was constructed as follows: a 4.8 kb DNA fragment containing the *S. cerevisiae* BY2961 *pol3* gene, plus the UTR 1 kb upstream and 0.5 kb downstream, (*Matα ura3-52, his3-Δ300, trp1-Δ901, leu2-3, 112 lys2-801, ade2-2*) was inserted into the *Sal*I-*Eco*RI site of YCplac33, and 3 amino acid substitutions, D321A, E323A, and L612M, were introduced into the *pol3* gene using site-directed mutagenesis [20]. YCplac33 is low-copy number plasmid and is stably maintained in *S. cerevisiae* [20].

2.2. Mutator Mutagenesis. YCplac33/*polδMKII* vector (and YCplac33 empty vector as nonmutator control) was introduced into *S. cerevisiae* BY2961 strain cells using the LiCl method, and the transformants (mutator strains) were selected on synthetic complete (SC)-agar plates without uracil. Five mutator strains were picked and independently cultivated in 1 mL SC medium at 30°C for 24 h (about 30 generations) in order to introduce mutations into their chromosomes. To determine the mutation frequencies of the 5 mutator strains, aliquots were spread on SC-agar plates containing L-canavanine sulfate salt (0.06 mg/mL) (Sigma, St. Louis, MO, USA) to identify *CAN1* mutants, and incubated until resistant colonies were formed. The mutation frequencies were calculated as the number of drug-resistant colonies divided by the number of colonies on SC-agar plate without drug. Forward mutation rates at *CAN1* were determined by fluctuation analysis using these 5 independent cultures [21]. In order to fix mutations, another aliquot of the mutator culture was spread on SC-agar plates containing 5-fluoroorotic acid monohydrate (Wako) to obtain demutatorized cells curing from YCplac33/*polδMKII* vector. The genomic DNA was prepared from the demutatorized cells using the procedure described in the following section.

2.3. EMS Mutagenesis. *S. cerevisiae* BY2961 strain cells were suspended in 0.1 M phosphate-buffered saline (PBS) (pH 7.0) containing 1.5, 2.0, 2.5, or 3.0% ethyl methanesulfonate (EMS) and were incubated at 30°C for 1 h to introduce chromosomal mutations. The cells were washed 3 times with 5% sodium thiosulfate, suspended in sterilized water, and spread on SC-agar plates containing L-canavanine sulfate salt (0.06 mg/mL) (Sigma) to identify *CAN1* mutants. The mutation frequencies were calculated as described above. Another aliquot of the EMS-treated cell suspension was spread on a YPD-agar plate to isolate single clones. The genomic DNA was prepared from 5 single clones derived from the cells treated with 1.5% EMS using the procedure described in the following section.

2.4. Library Preparation for Illumina Sequencing. The genomic DNA from *S. cerevisiae* was extracted using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA). Each sequenced sample was prepared according to the Illumina protocols. Briefly, 3 μg of genomic DNA was fragmented to an average length of 200 bp by using the Covaris S2 system (Covaris, Woburn, MA, USA). The fragmented DNA was repaired, a single "A" nucleotide was ligated to the 3′ end, Illumina Index PE adapters (Illumina, San Diego, CA, USA) were ligated to the fragments, and the sample was size selected for a 300 bp product using E-Gel SizeSelect 2% (Invitrogen, Grand Island, NY, USA). The size-selected product was amplified by 18 cycles of PCR with the primers InPE1.0, InPE2.0, and the Index primer containing 6-nt barcodes (Illumina). The final product was validated using the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA).

2.5. Sequencing and Data Analysis. The 11 barcoded libraries (the parental strain BY2961, 5 colonies from the mutator strain, and 5 colonies from the EMS-treated strain) were used for cluster generation in several multiplexed flow cell lanes in the Illumina Genome Analyzer II system. Ninety-one cycles of multiplexed paired-end sequencing was performed, running phi X 174 genomic DNA as a control in a separate lane of the flow cell. After the sequencing reactions were

complete, Illumina analysis pipeline (CASAVA 1.6.0) was used to carry out image analysis, base calling, and quality score calibration. Reads were sorted by barcode and exported in the FASTQ format. The quality of each sequencing library was assessed by evaluating the quality score chart and the nucleotide distribution plot using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).

Once the raw sequence data were curated, the reads of each sample were aligned to the S288c reference genome (http://www.yeastgenome.org/) using the BWA software (Ver. 0.5.1) with default parameters [22]. To avoid false positives and mutations from repetitive regions, we removed repetitive reads from the alignment files. We then used the SAMtools software (Ver. 0.1.9) [23] to produce the lists of mutations. To identify mutations that were produced by mutagenesis, we applied the following filtering criteria to the lists of mutations:

(a) the coverage at the mismatch positions should be at least 10;

(b) the variant is not present in the sequenced parental strain;

(c) indels meet a SNP quality threshold of 50 and substitutions meet a SNP quality threshold of 20 (SAMtools assigns SNP quality, which is the Phred-scaled probability that the consensus is identical to the reference);

(d) samples meet a mapping quality of 30 (SAMtools assigns Mapping quality, which is the Phred-scaled probability that the read alignment is wrong);

(e) the percentage of reads showing the variant allele exceeds 90%.

A variant must pass this filter to be considered a mutation. Alignments of all mutations were inspected by Integrative Genomics Viewer (IGV) [24]. The lists of mutations were then annotated using COVA (comparison of variants and functional annotation) (http://sourceforge.net/projects/cova). COVA was specifically designed to annotate the large number of identified mutants using the Genbank annotation files.

3. Results

3.1. Determination of Mutation Frequencies. In this study, we evaluated the performance of the *polδMKII* mutator, compared with that of the commonly used chemical mutagen, EMS. To assess EMS efficiency, *S. cerevisiae* BY2961 cells were treated with different concentrations of EMS. The lethality and mutation frequencies of the canavanine resistant colonies are shown in Table 1. At an EMS concentration of 1.5%, the mutation frequency was approximately 18-fold higher than that in the control (untreated) strain. Above 2.0% EMS, the survival rate decreased with no increase in mutation frequency. Based on this result, we decided to use cells treated with 1.5% EMS for whole-genome sequencing.

To assess the effectiveness of the mutator, we transformed the haploid BY2961 strain with a yeast expression plasmid, YCplac33/*polδMKII*, expressing the *polδ* mutant allele containing both the mutation to inactivate the proofreading activity (D321A and E323A) and the mutation to decrease the fidelity of replication (L612M). The mutator strain harboring the YCplac33/*polδMKII* plasmid will be referred to from here on as "mutator." We determined the mutation frequency by resistance to canavanine. As summarized in Table 2, the mutation frequency of the mutator was approximately 132-fold higher than in the cells containing the empty vector. The forward mutation rate at the *CAN1* (arginine permease) locus was calculated to be 7.9×10^{-6}/cell division. These results show that the plasmid-generated mutated *polδ* protein effectively competes with the endogenous wild-type *polδ* protein that is produced from the chromosome, and the semidominant negative expression of mutated *polδ* was effective in introducing mutations. These results also demonstrate that the mutation frequency of the mutator was approximately 7 times higher than that of EMS.

Table 1: Relationship between mutation frequency and survival after EMS treatment.

EMS concentration (%)	Mutation frequency of canavanine resistant ($\times 10^{-7}$)	Fold elevation*	Survival (%)
0.0	2	1	100
1.5	35	18	51
2.0	36	19	30
2.5	33	17	21
3.0	37	19	12

* Fold elevation is relative to untreated cells.

Table 2: Frequency of drug-resistant mutants in the mutator strains.

Plasmids	Mutation frequency of canavanine resistant ($\times 10^{-7}$)	Fold elevation*
YCplac33	3.70	1
YCplac33/*polδMKII*	486.7 ± 145.0#	132

* Fold elevation is relative to empty vector.
#Mean ± standard deviation of 3 SC plates.

3.2. Whole-Genome Sequencing. To analyze the genome-wide mutation spectra of the 2 different mutagenesis methods, we implemented a parallel sequencing approach with the Illumina Solexa technology (GAII instrument). We sequenced the parental haploid strain BY2961, each of the 5 clones from the mutator strains, and each of the 5 clones from the EMS-treated strains under nonselective conditions. Sequencing reads were aligned to the S288c reference genome using the BWA software [22]. To avoid false positives due to mutations from repetitive regions, reads mapped to multiple locations were discarded, and only uniquely mapped reads were used for subsequent analysis.

Table 3: Sequencing and mapping statistics.

Sample name	Number of mapped unique reads	% mapped reads	% genome covered* by unique reads	Average coverage by unique reads
BY2961	11,155,487	96.13	94.97	87.9×
EMS1	5,406,681	96.94	94.81	42.2×
EMS2	6,240,554	97.26	94.85	48.7×
EMS3	5,275,583	98.12	94.81	41.2×
EMS4	4,502,271	97.17	94.80	35.2×
EMS5	4,113,345	96.27	94.83	32.1×
Mutator 1	9,612,541	93.93	94.90	75.8×
Mutator 2	5,111,531	92.39	94.79	39.9×
Mutator 3	5,649,822	96.11	94.95	44.1×
Mutator 4	4,226,405	98.79	94.85	33.0×
Mutator 5	9,855,938	97.36	95.10	77.6×

* Coverage is defined as the percentage of bases in the genome that have at least 1 uniquely mapped read at that position.

In the current study, the average genomic coverage ranged from 32× to 87× (Table 3). On average, 94.18% of the S288c reference genome was covered with at least 1 uniquely mapped read at each base. Subsequently, we analyzed the data for 2 kinds of mutational events: single nucleotide variants (SNVs) and small insertions and deletions (Indels). Illumina sequencing found 6,766 genetic differences between our parental strain BY2961 and the S288c. Mutations induced by these mutagens were identified by subtracting the parental mutations. Sequence-processing details can be found in Section 2.

3.3. The Mutation Spectra of Mutator and EMS. We compared the average number of mutations between mutator strains and EMS-mutagenized strains (Figure 1). Mutator produced fewer SNVs than EMS (7.2 versus 55.8 per strain, resp., $P < 0.05$). Mutator and EMS produced few deletions (1.6 versus 2.8 per strain, resp.), as well as few insertions (0.2 versus 0.6 per strain, resp.). An average of 1.14×10^7 nucleotide sites fulfilled our criteria of read depth (≥10), with an average base-substitutional mutation rate estimate of EMS: 4.87 (SE = 1.34) $\times 10^{-6}$ per site, Mutator: 2.09 (0.55) $\times$ 10^{-8} per site per cell division (about 30 generations). The rate we calculated for the mutator is 100-fold higher than the previously reported spontaneous mutation rate, 3.3 (0.8) $\times$ 10^{-10}, based on 454 analyses of 4 mutation-accumulation (MA)-lines [26]. The 2 mutagens generate mutations that are distributed similarly across the various gene features although the mutator did produce more SNVs within exons than did EMS (Figure 2).

The mutation spectra are shown in Figure 3(a). In the genome-wide profile, we found that the mutator primarily induced transversions (72%) while EMS primarily induced transitions (97%), well in accord with the known mutagenic specificity of EMS [12]. Similarly, the mutator primarily induced transversions (69%) in the nonsynonymous substitutions in exons (Figure 3(b)), similar to what has been seen in *pol3-01* study using *URA3* reporter gene [16]. EMS treatment was also in agreement with the genome-wide spectra, induced transitions with a prevalence of 98%.

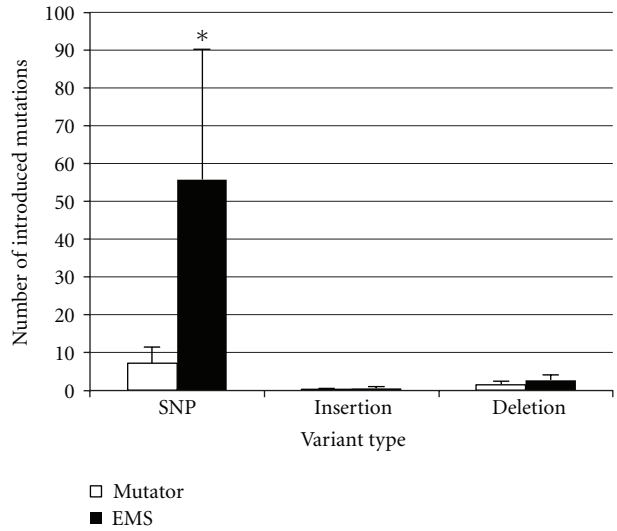


Figure 1: Average number of introduced mutations. By subtracting parental mutations from each mutagenized strain, we determined the number of mutations that were introduced by each mutagen. Bars represent mean ± standard error for 5 clones. $^*P < 0.05$ versus mutator in a two-sample t-test.

3.4. Amino Acid Substitution Patterns. The mutation spectra of a given mutagenesis method influences the repertoire of changed amino acids at the protein level, and we were able to evaluate the amino acid substitution patterns generated by our 2 protocols (Table 4). Initially, we classified mutations into those that preserved the corresponding amino acid,

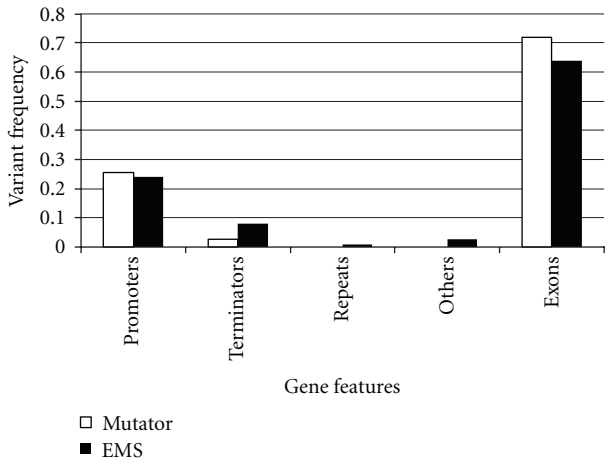


Figure 2: Relative frequency of SNVs affecting various gene features. The mutator and EMS generated mutations that were distributed similarly across the various gene features. The data for individual strains were combined according to the mutagen used. Promoters indicate the region 1 kb upstream of each gene. Terminators indicate the region 200 bp downstream of each gene.

changed the amino acid, or generated a stop codon. A clear difference was seen between mutator and EMS. Of the total mutations, the mutator changed the amino acid in approximately 85%, whereas EMS changed the amino acid in approximately 61%. The mutator also generated more stop codons than EMS (7% versus 2%, resp.). While mutator generated more changes to the first or second nucleotide of the codon, EMS generated changes in all 3 positions in approximately equal proportions.

Amino acid changes were classified into conservative and nonconservative substitutions, where a conservative substitution changed the encoded amino acid to a similar amino acid according to the criteria of the BLOSUM62 matrix [25]. Of the amino acid changes, mutator produced more nonconservative substitutions than EMS (83% and 53%). For the comparison of random mutagenesis methods, Wong et al. [27] proposed a useful structure indicator that takes into account Gly and Pro substitutions as well as stop codons. In our study, the mutator produced an equivalent number of Gly/Pro and stop codon substitutions, whereas EMS generated only stop codon substitutions.

4. Discussion

In this study, we evaluated the performance of a novel mutagenesis technique using error-prone proofreading-deficient and low-fidelity DNA polymerase δ by determining the mutation rate of the strain harboring the enzyme. We also analyzed the spectra of mutations across the entire *S. cerevisiae* genome and then assessed the diversity of mutation types at the amino acid level.

Proofreading-deficient *polδ* mutants, such as *pol3-01* strain, and several low-fidelity *polδ* mutants, such as L612M, have been shown to present a mutator phenotype and to elevate the mutation rate [14–18]. We generated a BY2961 strain expressing a *polδMKII* mutator, *polδ* mutant allele containing a combination of mutations to inactivate the proofreading activity (D321A and E323A) and to decrease the fidelity of replication (L612M). This mutant allele acts as a strong mutator, as evidenced by the high frequency of spontaneous mutations (131-fold over control, compared to 18-fold for EMS strains). Vencatesan et al. reported the forward *CAN1* mutation rates of *polδ* mutants as 1.5×10^{-6} in L612M, and 5.6×10^{-6} in *pol3-01* [18]. These mutant strains were constructed by integrating the *pol3-01* or *pol3*-L612M allele into the chromosomal *POL3* gene by targeted integration, thereby disrupting the endogenous *POL3* gene. In contrast, our mutator plasmid expressing the *polδ* mutant allele produced a mutation rate of 7.9×10^{-6}, which shows a high mutation rate as well as chromosomal integration. The use of the *polδMKII* mutator plasmid allows the continued expression of the endogenous wild-type *POL3* and provides for an efficient restoration of the wild-type mutation rate by curing the yeast strains of the mutator plasmid. Once the desired trait(s) has been selected, curing the cells from the mutator plasmid can stabilize the newly obtained phenotype.

In general, all random mutagenesis methods developed to date are biased toward transition mutations, although efforts have been made to overcome this [28]. While transition bias was observed in EMS, we actually observed transversion bias with the mutator (Figure 3(a)). Because of this, the mutator yielded a broader spectrum of nucleotide changes across the entire genome. The mutator was also biased toward transversions in the nonsynonymous substitutions (Figure 3(b)). For EMS, the spectrum of mutation events we observed is similar to what has been reported by others [12].

At the protein level, the amino acid substitution pattern differed between the mutator and EMS (Table 4). Mutations generated by the mutator resulted in amino acid substitutions more often than did mutations generated by EMS (85% versus 61%, resp.). Most of the substitutions made by the mutator were nonconservative, whereas only half of the substitutions made by EMS were nonconservative. In addition, the mutator generated more structure-disturbing amino acid changes (Gly/Pro). The transversion bias of non-synonymous substitutions by the mutator generates more diverse amino acid substitution patterns than does the transition bias of EMS.

Although the average base-substitution mutation rate of EMS was approximately 100 times higher than that of the mutator, the mutation frequency of the mutator was approximately 7 times higher than that of EMS. This gap between a higher apparent mutation frequency and fewer mutations may be explained by the higher proportion of amino acid changes and the diversity of amino acid substitutions by the mutator. This suggests one plausible explanation for the effectiveness of the disparity mutagenesis.

The disparity mutagenesis technique has been successfully applied to not only eukaryotic microorganisms such as *S. cerevisiae* [5, 7–9], *S. pombe* [9], and *Ashbya gossypii* [10], but also to prokaryotic microorganisms such as *Escherichia coli* [4] and *Bradyrhizobium japonicum* [6]. We believe that

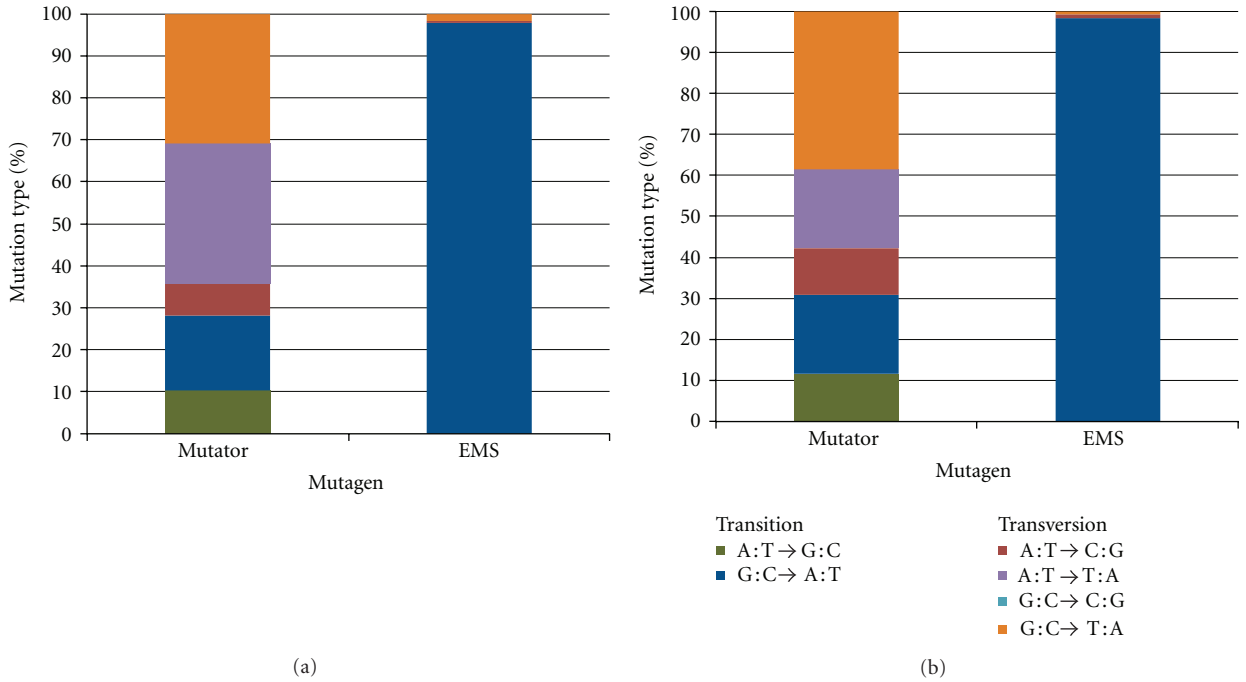


Figure 3: Relative frequency of transitions and transversions induced by *polδ* and EMS. The mutations spectra show the frequency of transitions and transversions generated by the mutator and EMS. The data for the individual strains were combined according to the mutagen used. The color key identifying the type of mutation is provided in the inset. Complementary mutations, such as A → C and T → G, are pooled. (a) Genome-wide profile; (b) Non-synonymous substitutions only.

Table 4: Mutations at protein level.

	Mutator		EMS	
	n	%	*n*	%
Total mutations	28	100	201	100
Preserved amino acids	2	7.1	74	36.8
Amino acid changes	24	85.7	123	61.2
Stop	2	7.1	4	2.0
Changes in codon letter	28	100	201	100
1st	11	39.3	64	31.8
2nd	13	46.4	65	32.3
3rd	4	14.3	72	35.8
Impact of amino acid change	24	100	123	100
Conservative[a]	4	16.7	57	46.3
Nonconservative	20	83.3	66	53.7
Stop and Gly/Pro codons	4	15.4	4	3.1
Stop	2	50.0	4	100.0
Gly/Pro	2	50.0	0	0.0

[a] Conservative and nonconservative amino acid substitutions were defined according to the BLOSUM62 matrix [25].

this novel mutagenesis technique has the potential to be applied to a wide variety of microorganisms.

Our present study has demonstrated that a proofreading-deficient and low-fidelity *polδMKII* mutator is a useful and efficient method for rapid strain improvement based on *in vivo* mutagenesis. It has been suggested that organisms may accelerate evolution by decreasing the fidelity of the proofreading activity of *polδ* in nature [29]; therefore, this mutator may also be useful for studying the acceleration of evolution.

Abbreviations

EMS: Ethyl methanesulfonate
SNV: Single nucleotide variant
Indel: Insertions and deletions
SC: Synthetic complete.

Data Access

The raw reads used in this study are available on the DDBJ Sequence Read Archive (DRA) under accession DRA000522.

Acknowledgments

Authors thank Dr. M. Furusawa for critical reading of the manuscript and for useful suggestions. This study was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (S0801025).

References

[1] N. E. Labrou, "Random mutagenesis methods for in vitro directed enzyme evolution," *Current Protein and Peptide Science*, vol. 11, no. 1, pp. 91–100, 2010.

[2] O. Selifonova, F. Valle, and V. Schellenberger, "Rapid evolution of novel traits in microorganisms," *Applied and Environmental Microbiology*, vol. 67, no. 8, pp. 3645–3649, 2001.

[3] M. Furusawa and H. Doi, "Promotion of evolution: disparity in the frequency of strand-specific misreading between the lagging and leading DNA strands enhances disproportionate accumulation of mutations," *Journal of Theoretical Biology*, vol. 157, no. 1, pp. 127–133, 1992.

[4] K. Tanabe, T. Kondo, Y. Onodera, and M. Furusawa, "A conspicuous adaptability to antibiotics in the *Escherichia coli* mutator strain, dnaQ49," *FEMS Microbiology Letters*, vol. 176, no. 1, pp. 191–196, 1999.

[5] C. Shimoda, A. Itadani, A. Sugino, and M. Furusawa, "Isolation of thermotolerant mutants by using proofreading-deficient DNA polymerase δ as an effective mutator in *Saccharomyces cerevisiae*," *Genes and Genetic Systems*, vol. 81, no. 6, pp. 391–397, 2006.

[6] M. Itakura, K. Tabata, S. Eda et al., "Generation of *Bradyrhizobium japonicum* mutants with increased N_2O reductase activity by selection after introduction of a mutated dnaQ gene," *Applied and Environmental Microbiology*, vol. 74, no. 23, pp. 7258–7264, 2008.

[7] H. Abe, Y. Fujita, Y. Chiba, Y. Jigami, and K. I. Nakayama, "Upregulation of genes involved in gluconeogenesis and the glyoxylate cycle suppressed the drug sensitivity of an N-glycan-deficient *Saccharomyces cerevisiae* mutant," *Bioscience, Biotechnology and Biochemistry*, vol. 73, no. 6, pp. 1398–1403, 2009.

[8] H. Abe, Y. Fujita, Y. Takaoka et al., "Ethanol-tolerant *Saccharomyces cerevisiae* strains isolated under selective conditions by over-expression of a proofreading-deficient DNA polymerase δ," *Journal of Bioscience and Bioengineering*, vol. 108, no. 3, pp. 199–204, 2009.

[9] H. Abe, Y. Takaoka, Y. Chiba et al., "Development of valuable yeast strains using a novel mutagenesis technique for the effective production of therapeutic glycoproteins," *Glycobiology*, vol. 19, no. 4, pp. 428–436, 2009.

[10] E. Y. Park, Y. Ito, M. Nariyama et al., "The improvement of riboflavin production in *Ashbya gossypii* via disparity mutagenesis and DNA microarray analysis," *Applied Microbiology and Biotechnology*, vol. 91, no. 5, pp. 1315–1326, 2011.

[11] T. Matsuzawa, Y. Fujita, N. Tanaka, H. Tohda, A. Itadani, and K. Takegawa, "New insights into galactose metabolism by *Schizosaccharomyces pombe*: isolation and characterization of a galactose-assimilating mutant," *Journal of Bioscience and Bioengineering*, vol. 111, no. 2, pp. 158–166, 2011.

[12] S. Flibotte, M. L. Edgley, I. Chaudhry et al., "Whole-genome profiling of mutagenesis in *Caenorhabditis elegans*," *Genetics*, vol. 185, no. 2, pp. 431–441, 2010.

[13] T. Fukui, K. Yamauchi, T. Muroya et al., "Distinct roles of DNA polymerases delta and epsilon at the replication fork in *Xenopus* egg extracts," *Genes to Cells*, vol. 9, no. 3, pp. 179–191, 2004.

[14] L. Li, K. M. Murphy, U. Kanevets, and L. J. Reha-Krantz, "Sensitivity to phosphonoacetic acid: a new phenotype to probe DNA polymerase δ in *Saccharomyces cerevisiae*," *Genetics*, vol. 170, no. 2, pp. 569–580, 2005.

[15] A. Morrison, A. L. Johnson, L. H. Johnston, and A. Sugino, "Pathway correcting DNA replication errors in *Saccharomyces cerevisiae*," *The EMBO Journal*, vol. 12, no. 4, pp. 1467–1473, 1993.

[16] A. Morrison and A. Sugino, "The $3' \rightarrow 5'$ exonucleases of both DNA polymerases δ and ε participate in correcting errors of DNA replication in *Saccharomyces cerevisiae*," *Molecular and General Genetics*, vol. 242, no. 3, pp. 289–296, 1994.

[17] S. A. N. McElhinny, C. M. Stith, P. M. J. Burgers, and T. A. Kunkel, "Inefficient proofreading and biased error rates during inaccurate DNA synthesis by a mutant derivative of *Saccharomyces cerevisiae* DNA polymerase," *The Journal of Biological Chemistry*, vol. 282, no. 4, pp. 2324–2332, 2007.

[18] R. N. Venkatesan, J. J. Hsu, N. A. Lawrence, B. D. Preston, and L. A. Loeb, "Mutator phenotypes caused by substitution at a conserved motif A residue in eukaryotic DNA polymerase δ," *The Journal of Biological Chemistry*, vol. 281, no. 7, pp. 4486–4494, 2006.

[19] B. Timmermann, S. Jarolim, H. Russmayer et al., "A new dominant peroxiredoxin allele identified by whole-genome re-sequencing of random mutagenized yeast causes oxidant-resistance and premature aging," *Aging*, vol. 2, no. 8, pp. 475–486, 2010.

[20] R. D. Gietz and A. Sugino, "New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites," *Gene*, vol. 74, no. 2, pp. 527–534, 1988.

[21] D. E. Lea and C. A. Coulson, "The distribution of the numbers of mutants in bacterial populations," *Journal of Genetics*, vol. 49, no. 3, pp. 264–285, 1949.

[22] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.

[23] H. Li, B. Handsaker, A. Wysoker et al., "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.

[24] J. T. Robinson, H. Thorvaldsdóttir, W. Winckler et al., "Integrative genomics viewer," *Nature Biotechnology*, vol. 29, no. 1, pp. 24–26, 2011.

[25] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 22, pp. 10915–10919, 1992.

[26] M. Lynch, W. Sung, K. Morris et al., "A genome-wide view of the spectrum of spontaneous mutations in yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 27, pp. 9272–9277, 2008.

[27] T. S. Wong, D. Roccatano, M. Zacharias, and U. Schwaneberg, "A statistical analysis of random mutagenesis methods used for directed protein evolution," *Journal of Molecular Biology*, vol. 355, no. 4, pp. 858–871, 2006.

[28] T. S. Rasila, M. I. Pajunen, and H. Savilahti, "Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment," *Analytical Biochemistry*, vol. 388, no. 1, pp. 71–80, 2009.

[29] M. Furusawa, "Implications of double-stranded DNA structure for development, cancer and evolution," *Open Journal of Genetics*, vol. 01, no. 03, pp. 78–87, 2011.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 324549, 5 pages
doi:10.1155/2012/324549

Research Article

Dispensabilities of Carbonic Anhydrase in Proteobacteria

Kenji Ueda,[1] Hiromi Nishida,[2] and Teruhiko Beppu[1]

[1] *Life Science Research Center, College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa 252-0880, Japan*
[2] *Agricultural Bioinformatics Research Unit, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*

Correspondence should be addressed to Kenji Ueda, ueda.kenji@nihon-u.ac.jp

Received 10 February 2012; Accepted 14 March 2012

Academic Editor: Hideaki Nojiri

Carbonic anhydrase (CA) (E.C. 4.2.1.1) is a ubiquitous enzyme catalysing interconversion between CO_2 and bicarbonate. The irregular distribution of the phylogenetically distinct classes of CA in procaryotic genome suggests its complex evolutionary history in procaryotes. Genetic evidence regarding the dispensability of CA under high-CO_2 air in some model organisms indicates that CA-deficient microorganisms can persist in the natural environment by choosing high-CO_2 niches. In this study, we studied the distribution of CA in the genome of Proteobacteria. While a large majority of the genome-sequenced Proteobacteria retained a CA gene(s), intracellular bacterial genera such as *Buchnera* and *Rickettsia* contained CA-defective strains. Comparison between CA-retaining and CA- deficient genomes showed the absence of whole coding sequence in some strains and the presence of frameshifted coding sequence in other strains. The evidence suggests that CA is inactivated and lost in some proteobacteria during the course of evolution based on its dispensability.

Carbonic anhydrase (CA) (EC 4.2.1.1) is a ubiquitous enzyme catalysing interconversion between CO_2 and bicarbonate (HCO_3^-) [1, 2]. CA is fundamental to various biological functions including photosynthesis, respiration, and CO_2 transport. To date, the existence of 3 major classes (alpha, beta, and gamma) of this enzyme has been known. Interestingly, no significant structural similarities are observed among these classes. Based on this feature, CA is recognised as an excellent example of convergent evolution [1, 2]. Most of the mammalian and plant CA specifically belong to alpha and beta class, respectively. On the other hand, the distribution of CA in procaryotes is irregular; some retain multiple classes of CA or multiple enzymes from the same class, and others do not retain any class of CA. Hence, it is likely that the evolution of CA function in procaryotes has a complex historical background [1].

Recently, a significant insight into the role of procaryotic CA has been provided by genetic studies in some model organisms such as *Ralstonia eutropha* [3] *Escherichia coli* [4], and *Saccharomyces cerevisiae* [5]. Knockout mutants for CA of these microorganisms are unable to grow under ambient air but normally grow under an atmosphere with high levels (1–5%) of CO_2.

This phenomenon is explained by the necessity of bicarbonate in the reaction catalysed by several housekeeping enzymes such as phosphoenolpyruvate carboxylase, carbamoyl phosphate synthase, and acetyl-CoA carboxylase [4, 5]. CA-positive microorganisms can generate bicarbonate from environmental CO_2 by the catalytic reaction of CA and supply it to these enzymes, but CA-negative ones cannot. Hence, the former can grow even under ambient air containing a low level of CO_2 (0.035%), but the latter cannot initiate growth unless they are supplied with a sufficient concentration of bicarbonate. The latter organisms, however, can grow under a high-CO_2 atmosphere since it generates a high concentration of bicarbonate to maintain natural equilibrium. This in turn indicates that CA is not essential for microbial growth under high-CO_2 environments, such as in soil, seawater, intestine, and some other syntrophic and commensal situations. Our previous study showed that an *E. coli* CA mutant was able to grow even under ambient air when it was cocultured with *Bacillus subtilis* [6].

The above-mentioned knowledge makes us speculate that the study of CA distribution in microbial genome will provide an insight into the history of adaptation of microorganisms to environment. Recently, we described that

Table 1: CA-deficient strains of genome-sequenced Proteobacteria.

Class	Genus*	Species/strain
Gammaproteobacteria	*Buchnera* (7)	*Buchnera aphidicola* APS
		Buchnera aphidicola Sg
		Buchnera aphidicola Bp
		Buchnera aphidicola Cc
		Buchnera aphidicola 5A
		Buchnera aphidicola Tuc7
		Buchnera aphidicola (*Cinara tujafilina*)
	Wigglesworthia (1)	*Wigglesworthia glossinidia*
	Blochmannia (3)	*Candidatus Blochmannia floridanus*
		Candidatus Blochmannia pennsylvanicus
		Candidatus Blochmannia vafer
	Riesia (1)	*Candidatus Riesia pediculicola*
	Moranella (1)	*Candidatus Moranella endobia*
	Actinobacillus (4)	*Actinobacillus pleuropneumoniae* JL03 (serotype 3)
	Thioalkalimicrobium (1)	*Thioalkalimicrobium cyclicum*
	Acidithiobacillus (1)	*Acidithiobacillus caldus*
	Baumannia (1)	*Baumannia cicadellinicola*
	Carsonella (1)	*Candidatus Carsonella ruddii*
Betaproteobacteria	*Zinderia* (1)	*Candidatus Zinderia insecticola CARI*
Deltaproteobacteria	*Desulfohalobium* (1)	*Desulfohalobium retbaense*
	Desulfococcus (1)	*Candidatus Desulfococcus oleovorans*
	Desulfatibacillum (1)	*Desulfatibacillum alkenivorans*
	Syntrophobacter (1)	*Syntrophobacter fumaroxidans*
	Hippea (1)	*Hippea maritima*
Alphaproteobacteria	*Rickettsia* (15)	*Rickettsia prowazekii*
		Rickettsia typhi
		Rickettsia canadensis
		Rickettsia conorii
		Rickettsia akari
		Rickettsia rickettsii Sheila Smith
		Rickettsia rickettsii Iowa
		Rickettsia massiliae
		Rickettsia heilongjiangensis
		Rickettsia japonica
		Rickettsia bellii RML369-C
	Orientia (2)	*Orientia tsutsugamushi* Boryong
		Orientia tsutsugamushi Ikeda
	Bartonella (6)	*Bartonella quintana*
	Hodgkinia (1)	Candidatus *Hodgkinia cicadicola*

* The number of genome-sequenced species/strains of each genus is shown in parentheses.

Symbiobacterium thermophilum, a unique syntrophic bacterium that effectively grows in coculture with a cognate *Geobacillus stearothermophilus* [7], lost CA in the course of evolution [8]. Our studies have shown that *S. thermophilum* grows on high CO_2 supply from environment and that this could be the reason for the absence of CA from its genome [6]. The phylogeny of CA distributed in Clostridia to which *S. thermophilum* belongs indicated that the common ancestor of this group of bacteria retained a CA gene and that *S. thermophilum* lost CA in the course of its adaptation to high CO_2 environments [8].

To deepen our insight into the correlation between CA deficiency and adaptation to high CO_2 environments, we studied the distribution of CA in the phylum Proteobacteria. Proteobacteria consists of five distinctive classes (alpha, beta, gamma, delta, and epsilon) and unclassified classes including the genus *Magnetococcus* (http://www.ncbi.nlm.nih.gov/genome/). To date (February 1, 2012), complete genome sequence information is available with regard to 649 strains of 249 genera (supplementary Table S1 in Supplementary Material available online at doi: 10.1155/2012/324549). Our search for the presence of CA by using the pathway database available at GenomeNet (http://www.genome.jp/) and BLAST searches (protein-protein searches based on BLOSUM62 scoring matrix) using known protein sequences annotated to be CA (corresponding to the protein encoded by the intact CA coding sequences shown in Figure 1) as queries showed that 39 strains of 20 genera (Table 1) of the

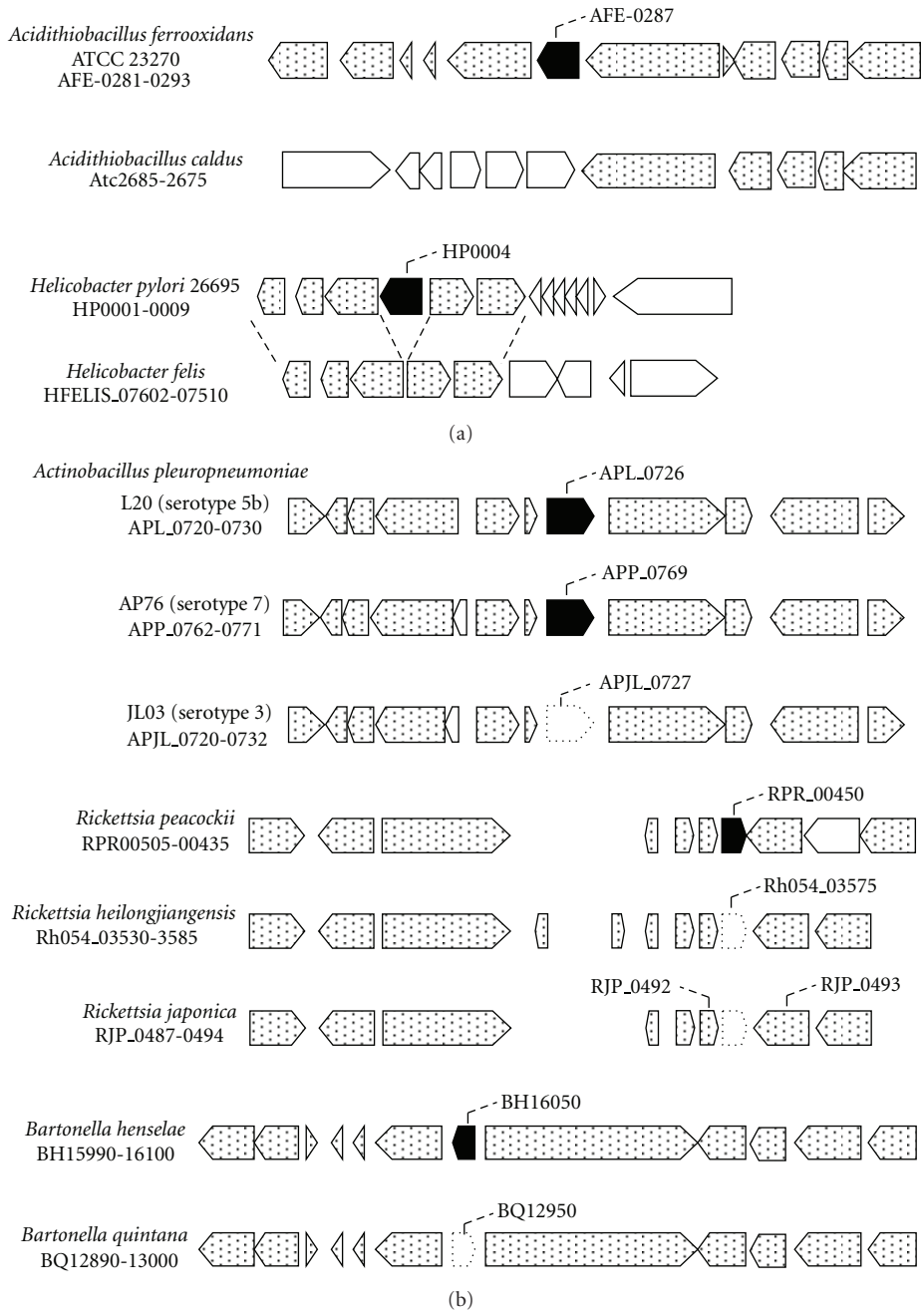


Figure 1: Comparison of the conserved gene cluster containing CA gene between CA-retaining and CA-deficient strains affiliating with the same genus of proteobacteria. Strains lacking whole CA coding sequences (a) and retaining frameshifted CA coding sequences (b) are compared with those retaining intact CA genes of the same genus. The coding sequence for CA and other conserved genes are shown in solid and dotted bars, respectively.

genome-sequenced Proteobacteria do not retain any gene encoding CA.

Among the CA-deficient 20 genera, 4 genera (*Buchnera*, *Blochmannia*, *Rickettsia*, and *Orientia*) contained multiple CA-deficient strains (Table 1). These were obligate intracellular bacteria. It is known that endosymbionts lack genes involved in primary metabolism. For example, *Buchnera* sp. lacks amino acid biosynthesis genes, which are compensated for by the activity of the host organism [9]. Such genetic defects in symbionts genome have probably occurred after establishing a tight, symbiotic relationship with the host organism. Presumably, the intracellular environments contain a high level of CO_2; hence the catalytic function of CA is not necessary for the bacteria habituating in such environments. On the other hand, some intracellular bacteria such as *Wolbachia* retain a putative CA gene (supplementary Table S1). This suggests that intracellular environment does not always compensate for CA deficiency.

13 out of the 20 CA-deficient genera contained only a single genome-sequenced strain (supplementary Table S1). They include intracellular *Candidatus* bacteria and lithoautotrophic and sulfate-reducing bacteria. It is not yet known whether the defect is a common feature of the genus or not, but it is possible that the CA deficiency is widespread among those intracellular bacteria as in the abovementioned genera.

Contrasting to the genuswide deficiency, the other 3 genera (*Actinobacillus*, *Acidithiobacillus*, and *Bartonella*) harboured strain-specific defect of CA (supplementary Table S1). In *Acidithiobacillus*, all strains except for *Acidithiobacillus caldus* retained the CA gene in the conserved gene cluster (the corresponding region of *Acidithiobacillus ferrooxidans* ATCC 23270 is shown in Figure 1(a)). Contrasting to this, *A. caldus* partially retained the conserved genes. While the genes upstream of CA were conserved, those downstream of CA including CA gene were not (Figure 1(a)). This makes us think of the possibility that the CA deficiency in *A. caldus* is due not to simple deletion but to a genetic rearrangement that has occurred in a relatively large scale.

Lack of CA gene in a conserved gene cluster was also observed with respect to the two species of *Helicobacter*, *Helicobacter felis* and *Helicobacter bizzozeronii*. All the genome-sequenced *Helicobacter* strains except for the two species contained the conserved gene cluster consisting of 6 coding sequences including CA gene (the corresponding region of *Helicobacter pylori* 26695 is shown in Figure 1(a)). Contrasting to this, the genome of the two *Helicobacter* spp. retained the conserved cluster lacking the coding region for CA (the corresponding region of *H. felis* is shown in Figure 1(a)). *H. felis* and *H. bizzozeronii* retained a CA gene in a different locus (corresponding to HFELIS_06160 and HBZC1_14670, resp.).

The other case of strain-specific CA deficiency was based on mutations in the coding sequence. Frameshift mutations inactivating CA gene were identified with respect to the four strains, *Actinobacillus pleuropneumoniae* JL03, *Rickettsia heilongjiangensis*, *Rickettsia japonica,* and *Bartonella quintana* (Figure 1(b)). These organisms retained a frame-shifted coding sequence exactly at the position corresponding to the locus where the intact CA ortholog is located in related strains (Figure 1(b)). The coding region of *A. pleuropneumoniae* JL03 and *R. japonica* contained a single-base deletion in the middle part (supplementary Figures S1 and S2). The coding region of *R. heilongjiangensis* lacked 95 bp corresponding to the N-terminal part of CA (supplementary Figure S2). *B. quintana* contained multiple mutations including two single-base deletions, one 8-base insertion, two single-base insertions, and one non-sense (ochre) mutation (supplementary Figure S3).

It is most likely that the abovementioned mutations inactivating the CA gene have been introduced into the ancestral intact coding sequence during the course of evolution. The diverged mode of mutation may reflect the process of how dispensable genes are lost from the bacterial genome. The existence of the strains carrying the inactivated coding sequence strongly suggests that the CA gene is not necessary for their persistency. It is not yet known how these mutant strains compensate for their CA deficiency, but we may reasonably speculate that it is correlated with the environmental CO_2 content.

The CA-deficient genera described in this paper are usually handled under a microaerobic or anaerobic atmosphere containing 1–5% CO_2 [10]. Hence, the conventional isolation method for these organisms has made possible isolation of strains requiring high CO_2. On the other hand, the standard isolation procedure for aerobic proteobacteria using ambient air prevents isolation of CO_2-requiring strains. This makes us think of the possibility that the very high proportion of CA-positive strains (610 out of 649 strains) (supplementary Table S1) is due to the limitation of isolation condition and is not appropriately reflecting the true distribution of CA in Proteobacteria.

The evolution of microbial genome reflects the history of environmental change. We expect that comprehensive analyses regarding the distribution of specific adaptive functions in microbial genome will provide deep insights into the constitution of the ecosystem.

Acknowledgment

This study was supported by the High-Tech Research Center Project of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] L. S. Smith and J. G. Ferry, "Prokaryotic carbonic anhydrases," *FEMS Microbiology Reviews*, vol. 24, no. 4, pp. 335–366, 2000.

[2] B. C. Tripp, K. Smith, and J. G. Ferry, "Carbonic anhydrase: new insights for an ancient enzyme," *The Journal of Biological Chemistry*, vol. 276, no. 52, pp. 48615–48618, 2001.

[3] B. Kusian, D. Sultemeyer, and B. Bowien, "Carbonic anhydrase is essential for growth of *Ralstonia eutropha* at ambient CO_2 concentrations," *Journal of Bacteriology*, vol. 184, no. 18, pp. 5018–5026, 2002.

[4] C. Merlin, M. Masters, S. McAteer, and A. Coulson, "Why is carbonic anhydrase essential to *Escherichia coli*?" *Journal of Bacteriology*, vol. 185, no. 21, pp. 6415–6424, 2003.

[5] J. Aguilera, J. P. van Dijken, J. H. de Winde, and J. T. Pronk, "Carbonic anhydrase (Nce103p): an essential biosynthetic enzyme for growth of *Saccharomyces cerevisiae* at atmospheric carbon dioxide pressure," *Biochemical Journal*, vol. 391, no. 2, pp. 311–316, 2005.

[6] T. O. Watsuji, T. Kato, K. Ueda, and T. Beppu, "CO_2 supply induces the growth of *Symbiobacterium thermophilum*, a syntrophic bacterium," *Bioscience, Biotechnology, and Biochemistry*, vol. 70, no. 3, pp. 753–756, 2006.

[7] K. Ueda and T. Beppu, "Lessons from studies of *Symbiobacterium thermophilum*, a unique syntrophic bacterium," *Bioscience, Biotechnology and Biochemistry*, vol. 71, no. 5, pp. 1115–1121, 2007.

[8] H. Nishida, T. Beppu, and K. Ueda, "*Symbiobacterium* lost carbonic anhydrase in the course of evolution," *Journal of Molecular Evolution*, vol. 68, no. 1, pp. 90–96, 2009.

[9] S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa, "Genome sequence of the endocellular bacterial

symbiont of aphids *Buchnera* sp. APS," *Nature*, vol. 407, no. 6800, pp. 81–86, 2000.

[10] "The proteobacteria," in *Bergey's Manual of Systematic Bacteriology*, D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity, Eds., vol. 2, Springer, New York, NY, USA, 2nd edition, 2005.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 745931, 6 pages
doi:10.1155/2012/745931

Research Article

Evolution of Lysine Biosynthesis in the Phylum *Deinococcus-Thermus*

Hiromi Nishida[1] and Makoto Nishiyama[2]

[1] *Agricultural Bioinformatics Research Unit, Graduate School of Agricultural and Life Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*
[2] *Biotechnology Research Center, University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*

Correspondence should be addressed to Hiromi Nishida, hnishida@iu.a.u-tokyo.ac.jp

Received 28 January 2012; Accepted 17 February 2012

Academic Editor: Kenro Oshima

Thermus thermophilus biosynthesizes lysine through the α-aminoadipate (AAA) pathway: this observation was the first discovery of lysine biosynthesis through the AAA pathway in archaea and bacteria. Genes homologous to the *T. thermophilus* lysine biosynthetic genes are widely distributed in bacteria of the *Deinococcus-Thermus* phylum. Our phylogenetic analyses strongly suggest that a common ancestor of the *Deinococcus-Thermus* phylum had the ancestral genes for bacterial lysine biosynthesis through the AAA pathway. In addition, our findings suggest that the ancestor lacked genes for lysine biosynthesis through the diaminopimelate (DAP) pathway. Interestingly, *Deinococcus proteolyticus* does not have the genes for lysine biosynthesis through the AAA pathway but does have the genes for lysine biosynthesis through the DAP pathway. Phylogenetic analyses of *D. proteolyticus* lysine biosynthetic genes showed that the key gene cluster for the DAP pathway was transferred horizontally from a phylogenetically distant organism.

1. Introduction

The *Deinococcus-Thermus* phylum constitutes one of the major bacterial evolutionary lineages [1, 2]. At present, the genome sequence data of 6 genera (13 organisms) belonging to this phylum are available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [3].

Two pathways for lysine biosynthesis have been described, namely, the α-aminoadipate (AAA) pathway and the diaminopimelate (DAP) pathway [5]. The AAA pathway has two different types [6]. In *T. thermophilus*, a gene cluster was found for lysine biosynthesis not through the DAP pathway but through the AAA pathway [6–8]. Although *Deinococcus radiodurans* has genes homologous to the *T. thermophilus* lysine biosynthetic genes, these genes are scattered on the genome [9]. In addition, the *D. radiodurans* aspartate kinase that catalyzes the phosphorylation of L-aspartate (the first reaction in the DAP pathway) is structurally and phylogenetically very different from that of *T. thermophilus* [10]. Recent studies have shown that the genome signatures of these 2 bacteria are different [4], supporting the theory that *Deinococcus* species acquired genes from various other bacteria to survive different kinds of environmental stresses, whereas *Thermus* species have acquired genes from thermophilic bacteria to adapt to high-temperature environments [11].

The distribution of lysine biosynthetic genes in the *Deinococcus-Thermus* phylum has not been clearly described. In this study, we compared the distribution of the genes for lysine biosynthesis between 13 organisms (*D. deserti*, *D. geothermalis*, *D. maricopensis*, *D. proteolyticus*, *D. radiodurans*, *Marinithermus hydrothermalis*, *Meiothermus ruber*, *M. silvanus*, *Oceanithermus profundus*, *T. scotoductus*, *T. thermophilus* HB8, *T. thermophilus* HB27, and *Truepera radiovictrix*).

2. Methods

We analyzed the distribution of each of the following 10 enzymes related to lysine biosynthesis through the AAA pathway in the *Deinococcus-Thermus* phylum: α-aminoadipate aminotransferase, homoisocitrate dehydrogenase, LysW-γ-L-lysine aminotransferase, LysW-γ-L-lysine

Table 1: Genes for lysine biosynthesis through the α-aminoadipate pathway in the *Deinococcus-Thermus* phylum.

Organism	Enzyme 1	Enzyme 2	Enzyme 3	Enzyme 4	Enzyme 5	Enzyme 6	Enzyme 7	Enzyme 8	Enzyme 9	Enzyme 10
Thermus thermophilus HB27	*TTC0043*	*TTC1012*	*TTC1393*	*TTC1396*	*TTC1541**	*TTC1542**	*TTC1543**	*TTC1546**	*TTC1547**	*TTC1550**
Thermus thermophilus HB8	*TTHA0411*	*TTHA1378*	*TTHA1755*	*TTHA1757*	*TTHA1903**	*TTHA1904**	*TTHA1907**	*TTHA1910**	*TTHA1911**	*TTHA1914**
Thermus scotoductus	*TSC_c05810*	*TSC_c20650*	*TSC_c03550*	*TSC_c3520*	*TSC_c01940**	*TSC_c01930**	*TSC_c01920**	*TSC_c01890**	*TSC_c01880**	*TSC_c01850**
Meiothermus ruber	*Mrub_0871*	*Mrub_2738*		*Mrub_0027*	*Mrub_2721**	*Mrub_2723**	*Mrub_2724**	*Mrub_2727**	*Mrub_2728**	
Meiothermus silvanus	*Mesil_2567*	*Mesil_1337*	*Mesil_0348*	*Mesil_0347*	*Mesil_0435**	*Mesil_0436**	*Mesil_0438**	*Mesil_0441**	*Mesil_0442**	
Oceanithermus profundus		*Ocepr_1387*	*Ocepr_1797**	*Ocepr_1798**	*Ocepr_1796**	*Ocepr_1788**	*Ocepr_1784**	*Ocepr_1781**	*Ocepr_1780**	*Ocepr_1779**
Marinithermus hydrothermalis		*Marky_1533*	*Marky_0665**	*Marky_0663**	*Marky_0666**	*Marky_0667**	*Marky_0668**	*Marky_0671**	*Marky_0672**	*Marky_0673**
Deinococcus radiodurans		*DR_1674*	*DR_0794*	*DR_1413*	*DR_1420*	*DR_0963*	*DR_2194*	*DR_1614*	*DR_1610*	*DR_1238*
Deinococcus geothermalis	*Dgeo_2084*	*Dgeo_1458*	*Dgeo_1416*	*Dgeo_1391*	*Dgeo_0678*	*Dgeo_0685*	*Dgeo_1151**	*Dgeo_1154**	*Dgeo_1156**	*Dgeo_1257*
Deinococcus deserti		*Deide_09240*	*Deide_16910*	*Deide_17960*	*Deide_10430*	*Deide_10350*	*Deide_13430**	*Deide_13460**	*Deide_13470**	*Deide_13980*
Deinococcus maricopensis	*Deima_0046*	*Deima_1545*	*Deima_2454*	*Deima_2593*	*Deima_1346**	*Deima_1349**	*Deima_1350**	*Deima_1353**	*Deima_1355**	*Deima_1358**
Deinococcus proteolyticus	*Deipr_0213*									
Truepera radiovictrix		*Trad_2841*	*Trad_1401**	*Trad_1404**	*Trad_1399**	*Trad_1395**	*Trad_1392**	*Trad_1390**	*Trad_1389**	*Trad_1388**

Enzyme 1, α-aminoadipate aminotransferase.
Enzyme 2, Homoisocitrate dehydrogenase.
Enzyme 3, LysW-γ-L-lysine aminotransferase.
Enzyme 4, LysW-γ-L-lysine hydrolase.
Enzyme 5, LysW-γ-L-α-aminoadipate kinase.
Enzyme 6, LysW-γ-L-α-aminoadipyl-6-phosphate reductase.
Enzyme 7, α-aminoadipate-LysW ligase LysX.
Enzyme 8, LysU.
Enzyme 9, LysT.
Enzyme 10, Homocitrate synthase.
*More than 3 genes are clustered.

Table 2: Genes for lysine biosynthesis through the diaminopimelate pathway in the *Deinococcus-Thermus* phylum.

Organism	Aspartate kinase	Aspartate-semialdehyde dehydrogenase	Dihydrodipicolinate synthase	Dihydrodipicolinate reductase	LL-diaminopimelate aminotransferase	Diaminopimelate decarboxylase
Thermus thermophilus HB27	*TTC0166*	*TTC0177*	*TTC0591*			
Thermus thermophilus HB8	*TTHA0534*	*TTHA0545*	*TTHA0957*			
Thermus scotoductus	*TSC_c07050*	*TSC_c08140*	*TSC_c10420*			*TSC_c10870*
Meiothermus ruber	*Mrub_0976*	*Mrub_1641*	*Mrub_1335*			*Mrub_0798*
Meiothermus silvanus	*Mesil_1711*	*Mesil_2173*	*Mesil_2308*			*Mesil_0318*
Oceanithermus profundus	*Ocepr_1316*	*Ocepr_1018*				*Ocepr_2076*
Marinithermus hydrothermalis	*Marky_1492*	*Marky_1381*	*Marky_1261*			
Deinococcus radiodurans	*DR_1365*	*DR_2008*				*DR_1758*
Deinococcus geothermalis	*Dgeo_1127*	*Dgeo_1782*				*Dgeo_0790*
Deinococcus deserti	*Deide_11430*	*Deide_15740*	*Deide_1p00310*, *Deide_3p00120*, *Deide_3p01100*			*Deide_12830*, *Deide_21880*
Deinococcus maricopensis	*Deima_1822*	*Deima_2680*				*Deima_2660*
Deinococcus proteolyticus	*Deipr_0941*	*Deipr_0985*	*Deipr_1377**	*Deipr_1378**	*Deipr_1376**	*Deipr_0627*, *Deipr_1375**
Truepera radiovictrix	*Trad_0977*	*Trad_0289*	*Trad_1893*			*Trad_0134*

* More than 3 genes are clustered.

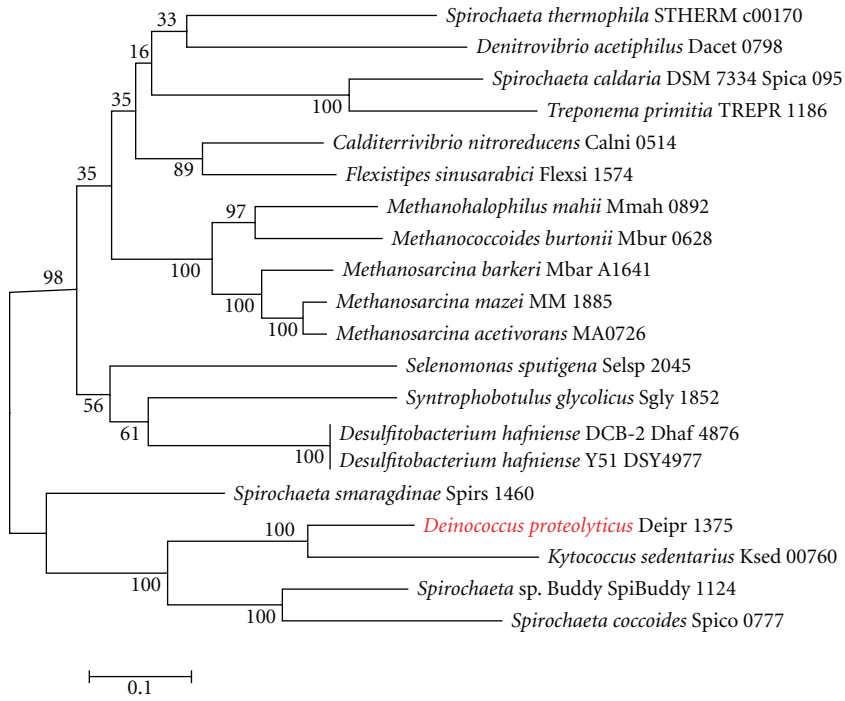


Figure 1: Phylogenetic relationship between *Deinococcus proteolyticus* diaminopimelate decarboxylase and related proteins. Multiple alignment was obtained using the top 20 amino acid sequences of the BLASTp search result for *D. proteolyticus* diaminopimelate decarboxylase (Deipro 1375), as based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The maximum-likelihood tree was constructed using MEGA software version 5 [12]. The WAG model was used as the amino acid substitution model. The nearest neighbor interchange was used for the maximum-likelihood heuristic method. The γ-distributed rate was considered, and the number of discrete γ categories was 3. Bootstrap analysis was performed with 100 replicates. Red indicates *D. proteolyticus*.

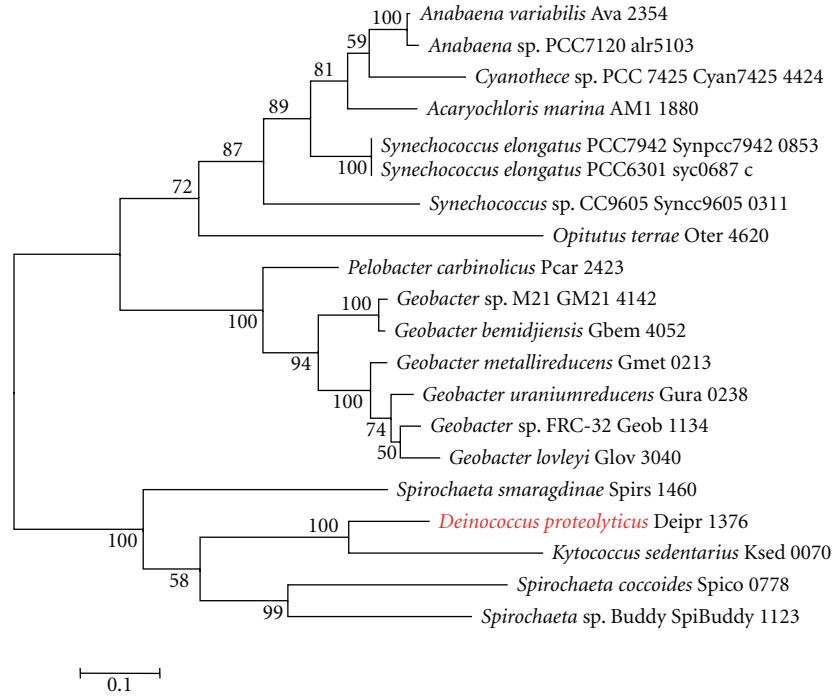


Figure 2: Phylogenetic relationship between *Deinococcus proteolyticus* LL-diaminopimelate aminotransferase and related proteins. Multiple alignment was obtained using the top 20 amino acid sequences of the BLASTp search result for *D. proteolyticus* LL-diaminopimelate aminotransferase (Deipro 1376), as based on the KEGG database. The maximum-likelihood tree was constructed using MEGA software version 5 [12]. The WAG model was used as the amino acid substitution model. The nearest neighbor interchange was used for the maximum-likelihood heuristic method. The γ-distributed rate was considered, and the number of discrete γ categories was 3. Bootstrap analysis was performed with 100 replicates. Red indicates *D. proteolyticus*.

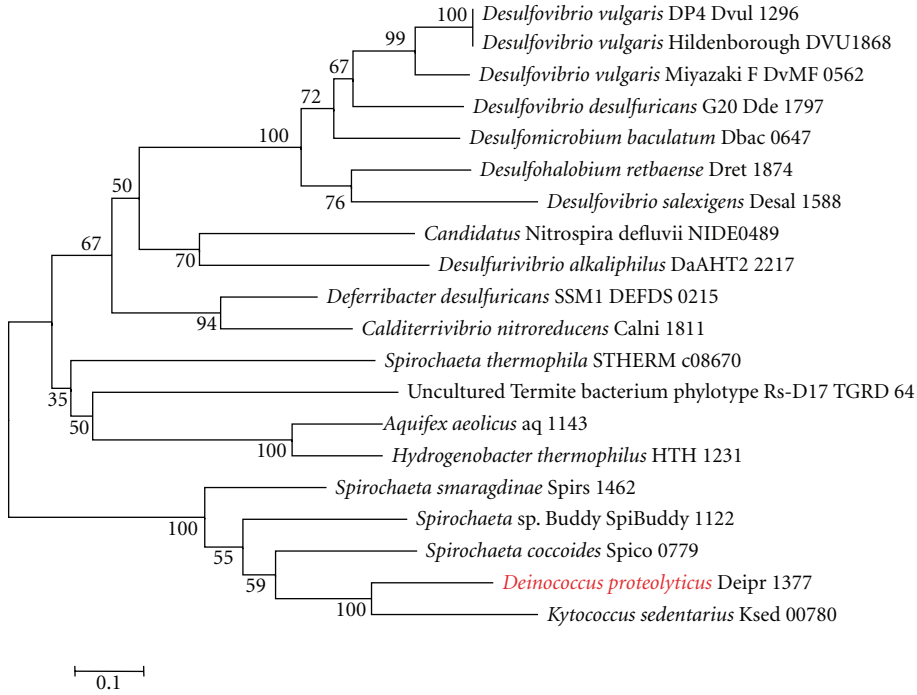


Figure 3: Phylogenetic relationship between *Deinococcus proteolyticus* dihydrodipicolinate synthase and related proteins. Multiple alignment was obtained using the top 20 amino acid sequences of the BLASTp search result for *D. proteolyticus* dihydrodipicolinate synthase (Deipro 1377), as based on the KEGG database. The maximum-likelihood tree was constructed using MEGA software version 5 [12]. The WAG model was used as the amino acid substitution model. The nearest neighbor interchange was used for the maximum-likelihood heuristic method. The γ-distributed rate was considered, and the number of discrete γ categories was 3. Bootstrap analysis was performed with 100 replicates. Red indicates *D. proteolyticus*.

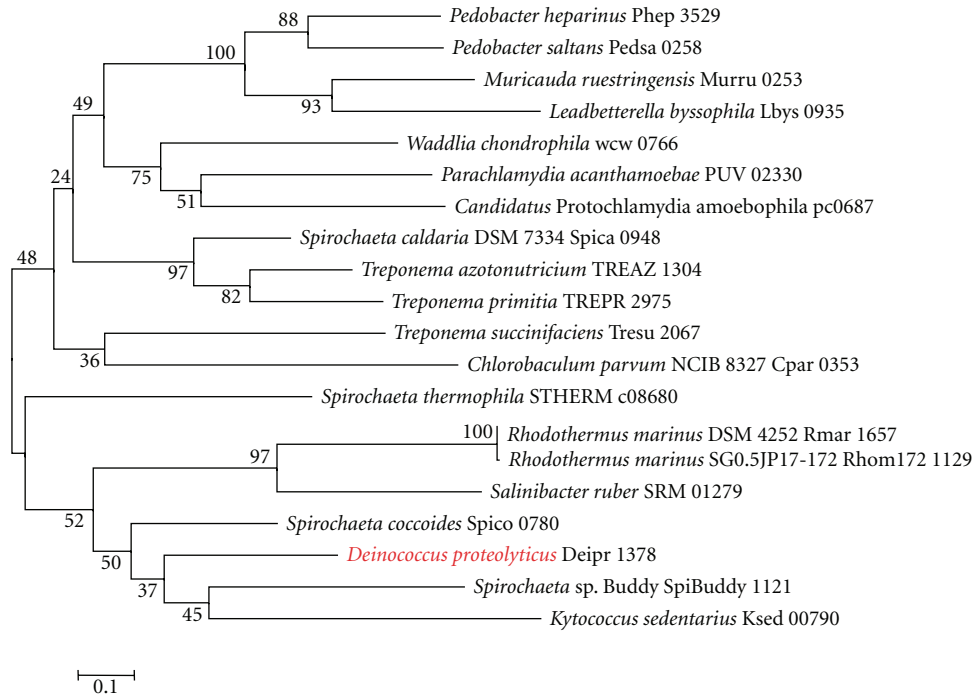


Figure 4: Phylogenetic relationship between *Deinococcus proteolyticus* dihydrodipicolinate reductase and related proteins. Multiple alignment was obtained using the top 20 amino acid sequences of the BLASTp search result for *D. proteolyticus* dihydrodipicolinate reductase (Deipro 1378), as based on the KEGG database. The maximum-likelihood tree was constructed using MEGA software version 5 [12]. The WAG model was used as the amino acid substitution model. The nearest neighbor interchange was used for the maximum-likelihood heuristic method. The γ-distributed rate was considered, and the number of discrete γ categories was 3. Bootstrap analysis was performed with 100 replicates. Red indicates *D. proteolyticus*.

hydrolase, LysW-γ-L-α-aminoadipate kinase, LysW-γ-L-α-aminoadipyl-6-phosphate reductase, α-aminoadipate-LysW ligase LysX, LysU, LysT, and homocitrate synthase. In addition, we analyzed the distribution of each of the following 6 enzymes related to lysine biosynthesis through the DAP pathway: aspartate kinase, aspartate-semialdehyde dehydrogenase, dihydrodipicolinate synthase, dihydrodipicolinate reductase, LL-diaminopimelate aminotransferase, and diaminopimelate decarboxylase.

Homologous genes were selected on the basis of BLASTp search results by using each *T. thermophilus* enzyme for lysine biosynthesis through the AAA pathway and each *D. proteolyticus* enzyme for lysine biosynthesis through the DAP pathway. Multiple alignments were obtained using 20 amino acid sequences, with the highest to the 20th highest score by the BLASTp result. Maximum-likelihood trees were constructed using MEGA software version 5 [12]. The WAG model [13] was used as the amino acid substitution model. The nearest neighbor interchange was used for the maximum-likelihood heuristic method. The γ-distributed rate was considered, and the number of discrete γ categories was 3. Bootstrap analysis was performed with 100 replicates.

3. Results and Discussion

Genes homologous to the *T. thermophilus* genes for lysine biosynthesis through the AAA pathway were found to be widely distributed in bacteria belonging to the *Deinococcus-Thermus* phylum, except for *D. proteolyticus* (Table 1). Among the 13 organisms examined, *Marinithermus*, *Oceanithermus*, and *Truepera* have the largest gene cluster, containing 8 lysine biosynthetic genes (Table 1). In each phylogenetic analysis of the 10 enzymes, lysine biosynthetic genes of the *Deinococcus-Thermus* phylum were found to have a common ancestor (See in Supplementary Material Figures S1−S10 available online at doi:10.1155/2012/745931). We hypothesize that a common ancestor of the *Deinococcus-Thermus* phylum biosynthesized lysine through the AAA pathway.

In contrast, the distribution of genes for lysine biosynthesis through the DAP pathway was found to be limited in the *Deinococcus-Thermus* phylum (Table 2). Thus, LL-diaminopimelate aminotransferase and dihydrodipicolinate reductase were identified in no bacteria other than *D. proteolyticus* (Table 2). This observation supports our hypothesis that a common ancestor of the *Deinococcus-Thermus* phylum biosynthesized lysine not through the DAP pathway, but through the AAA pathway.

Interestingly, *D. proteolyticus* was found to have the genes for lysine biosynthesis through the DAP pathway (Table 2). *D. proteolyticus* has 2 diaminopimelate decarboxylases, namely, Deipro 0627 and Deipro 1375 (Table 2), which are structurally different from each other. Because Deipro 1375 forms a gene cluster with other genes for lysine

biosynthesis through the DAP pathway, we used Deipro 1375 as a query sequence in the BLASTp search. Each phylogenetic tree based on diaminopimelate decarboxylase (Figure 1), LL-diaminopimelate aminotransferase (Figure 2), dihydrodipicolinate synthase (Figure 3), and dihydrodipicolinate reductase (Figure 4) showed that the *D. proteolyticus* enzyme is closely related to that of the genera *Kytococcus* (a member of Actinobacteria) and *Spirochaeta* (a member of Spirochaetes) (Figures 1–4). The 3 phyla Actinobacteria, *Deinococcus-Thermus*, and Spirochaetes do not form a monophyletic lineage in the phylogenetic tree, as based on genomewide comparative studies [14]. In addition, the 4 genes encoding diaminopimelate decarboxylase, LL-diaminopimelate aminotransferase, dihydrodipicolinate synthase, and dihydrodipicolinate reductase are clustered in each genus (Figures 1–4). Thus, these findings strongly suggested that a DNA fragment including the 4 *D. proteolyticus* genes was horizontally transferred from a phylogenetically distant organism. This horizontal transfer event may have induced the loss of the genes for lysine biosynthesis through the AAA pathway in *D. proteolyticus*.

References

[1] W. G. Weisburg, S. J. Giovannoni, and C. R. Woese, "The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction," *Systematic and Applied Microbiology*, vol. 11, pp. 128–134, 1989.

[2] E. Griffiths and R. S. Gupta, "Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the *Deinococcus-Thermus* phylum," *Journal of Bacteriology*, vol. 186, no. 10, pp. 3097–3107, 2004.

[3] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, "KEGG for integration and interpretation of large-scale molecular datasets," *Nucleic Acids Research*, vol. 40, pp. D109–D114, 2012.

[4] H. Nishida, R. Abe, T. Nagayama, and K. Yano, "Genome signature difference between *Deinococcus radiodurans* and *Thermus thermopilus*," *International Journal of Evolutionary Biology*, vol. 2012, Article ID 205274, 6 pages, 2012.

[5] A. M. Velasco, J. I. Leguina, and A. Lazcano, "Molecular evolution of the lysine biosynthetic pathways," *Journal of Molecular Evolution*, vol. 55, no. 4, pp. 445–459, 2002.

[6] H. Nishida, M. Nishiyama, N. Kobashi, T. Kosuge, T. Hoshino, and H. Yamane, "A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis," *Genome Research*, vol. 9, no. 12, pp. 1175–1183, 1999.

[7] T. Kosuge and T. Hoshino, "Lysine is synthesized through the α-aminoadipate pathway in *Thermus thermophilus*," *FEMS Microbiology Letters*, vol. 169, no. 2, pp. 361–367, 1998.

[8] N. Kobashi, M. Nishiyama, and M. Tanokura, "Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: lysine is synthesized via α-aminoadipic acid not via diaminopimelic acid," *Journal of Bacteriology*, vol. 181, no. 6, pp. 1713–1718, 1999.

[9] H. Nishida, "Distribution of genes for lysine biosynthesis through the aminoadipate pathway among prokaryotic genomes," *Bioinformatics*, vol. 17, no. 2, pp. 189–191, 2001.

[10] H. Nishida and I. Narumi, "Phylogenetic and disruption analyses of aspartate kinase of *Deinococcus radiodurans*," *Bioscience, Biotechnology and Biochemistry*, vol. 71, no. 4, pp. 1015–1020, 2007.

[11] M. V. Omelchenko, Y. I. Wolf, E. K. Gaidamakova et al., "Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance," *BMC Evolutionary Biology*, vol. 5, article 57, 2005.

[12] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.

[13] S. Whelan and N. Goldman, "A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach," *Molecular Biology and Evolution*, vol. 18, no. 5, pp. 691–699, 2001.

[14] H. Nishida, T. Beppu, and K. Ueda, "Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae," *Genomics*, vol. 98, no. 5, pp. 370–375, 2011.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 874153, 17 pages
doi:10.1155/2012/874153

Research Article

Vehicles, Replicators, and Intercellular Movement of Genetic Information: Evolutionary Dissection of a Bacterial Cell

Matti Jalasvuori[1,2]

[1] *Department of Biological and Environmental Science, Center of Excellence in Biological Interactions, University of Jyväskylä, 40014 Jyväskylä, Finland*
[2] *Division of Evolution, Ecology and Genetics, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia*

Correspondence should be addressed to Matti Jalasvuori, matti.jalasvuori@jyu.fi

Received 9 December 2011; Revised 6 February 2012; Accepted 8 February 2012

Academic Editor: Hiromi Nishida

Prokaryotic biosphere is vastly diverse in many respects. Any given bacterial cell may harbor in different combinations viruses, plasmids, transposons, and other genetic elements along with their chromosome(s). These agents interact in complex environments in various ways causing multitude of phenotypic effects on their hosting cells. In this discussion I perform a dissection for a bacterial cell in order to simplify the diversity into components that may help approach the ocean of details in evolving microbial worlds. The cell itself is separated from all the genetic replicators that use the cell vehicle for preservation and propagation. I introduce a classification that groups different replicators according to their horizontal movement potential between cells and according to their effects on the fitness of their present host cells. The classification is used to discuss and improve the means by which we approach general evolutionary tendencies in microbial communities. Moreover, the classification is utilized as a tool to help formulating evolutionary hypotheses and to discuss emerging bacterial pathogens as well as to promote understanding on the average phenotypes of different replicators in general. It is also discussed that any given biosphere comprising prokaryotic cell vehicles and genetic replicators may naturally evolve to have horizontally moving replicators of various types.

1. Introduction

Viruses that infect prokaryotic cells are known to be enormously diverse in terms of genetic information [1, 2]. Most novel viral isolates are likely to have at least some genes that have no homologues among any of the previously known genes, including those in the genomes of related viruses [3]. Yet, there has been a dispute whether or not new genes may actually emerge in viruses [3]. Viruses are dependent on cellular resources such as nucleotides, amino acids, and lipids for producing more viruses; therefore it seems justified to ask whether they also use cellular genes for their genetic information. Yet, when viral genes are compared to other genes in databases, it often appears that they have no cellular counterparts [2]. Where then do these viral genes come from? Have they been acquired from a cellular host that we simply have not sequenced before? Or alternatively, are the cellular genes perhaps just evolving rapidly in viral genomes so that their common ancestry with the host genes can no longer be derived? Or perhaps, is it indeed possible that new genes actually emerge in viruses themselves?

Forterre and Prangishvili from Pasteur Institute argued that the core of the dispute appears to be in the notion that viruses are often considered to be just their protein-encapsulated extracellular forms [4] that are only stealing cellular resources (including genes) for their own purposes [3, 5, 6]. Take any textbook on viruses and majority of the pictures representing viruses are of the various types of viral shells composed of proteins (and sometimes lipids) that enclose the viral genome. But these infectious virus particles, or virions, are inert in all respects unless they encounter a susceptible host cell [7]. And due to this inertness of virions it is difficult to understand how a virus could ever come up with completely new genes.

The answer is, naturally, that viruses cannot produce new genes during their extracellular state, and thus any potential

event for the emergence of a new viral gene must still occur within a cell during the replication cycle of a virus [5]. But if the gene emerges in the genome of a virus, then would it rather be the virus, and not the cell, that was the originator of that gene? Or, to put it differently, was it not the virus that benefited from the emergence of new genetic information? The actual process that causes the genetic information to acquire the status of a gene would still be due to similar processes as the origin of genes within chromosomes (these being different types of genetic changes, such as point mutations, insertions, deletions, gene duplications, etc.), but these changes would be selected due to their improvements on the fitness of the virus. This reasoning has made Forterre to propose a model where viruses are seen essentially as a cellular life form that can also have an extracellular state [7, 8]. Virus is not strictly equivalent to the protein-enclosed viral genome. Rather, the extracellular form of a virus should be denoted as a virion, and this virion should not be mistaken for a virus. Viruses, in a complete sense, are organisms that live within cells (i.e., ribosome-encoding organisms) and can transform other cells into virus-cell organisms by producing more virions. In other words, viruses can utilize an extracellular encapsulated form to transfer its genetic information from one cell to another. Forterre coined a term *virocell*, which refers to the stage of viral life during which the virus is within a cell [7]. The virocell organism is indeed both a (capsid encoding) virus and a (ribosome encoding) chromosome, and the actual phenotype of the virocell is encoded by both of these genetic entities. The virocells are entirely capable of coming up with novel genetic information just as cells are, and thus approaching viruses from this perspective should clear any controversies about the emergence of new genetic information in viruses.

Forterre's line of reasoning along with my own studies on various different genetic elements (including characterization of temperate and virulent viruses [9, 10]; determination of common ancestor between plasmids, viruses and chromosomal elements [11]; conduction of evolution experiments with bacteria, viruses, and plasmids [12, 13]; as well as more theoretical work on horizontal movement of genetic information [14, 15]) has served as an inspiration for this paper. Indeed, it could be argued in more general terms what it means that prokaryotic cells can be (and often are) chimeras of various types of genetically reproducing elements. Virocell concept clears effectively many of the confusions between viruses and virions and their relationship with cells. Nonetheless, virocell is only a special case among all the possible types of prokaryotic organisms. Bacterial and archaeal cells can also contain conjugative plasmids, various types of transposons, defective prophages, and many other independent replicators that are distinct from the ribosome encoding prokaryotic chromosome. Together these replicators can produce organisms in all possible combinations. In order for the arguments about virocells to be consistent with the other potential chimeras of genetic replicators, the cell itself must be considered as a separate entity from all the genetic replicators (including chromosomes) that exploit the cell structure for replication. In the following chapters I will perform an evolutionary dissection to a bacterial cell. This will lead into the separation of cell vehicles and replicators from each other and thus provide one potential way to approach the evolution of bacterial organisms.

2. Vehicles and Replicators

"*A vehicle is any unit, discrete enough to seem worth naming, which houses a collection of replicators and which works as a unit for the preservation and propagation of those replicators*", Richard Dawkins wrote in *Extended Phenotype*. Dawkins utilized the concepts of replicators and vehicles in an argument which stated that evolution ultimately operated on the level of genetic information and not on the level of populations of organisms, species, or even cells. Replicators refer to packages of genetic information that are responsible for any effective phenotype of the vehicle. Vehicle itself can be a cell, a multicellular organism, or, for example, the host organism of a parasite. "*A vehicle is not a replicator*", argued Dawkins in an attempt to underline that it is the replicator (like the chromosome of a parasite) and not the vehicle (like the parasitized cell) that evolves. This difference, however, may sometimes be seemingly trivial, which is why it has caused some dissonance among evolutionary biologists.

Nevertheless, Dawkins' work focused mostly on explaining evolutionary issues of eukaryotic organisms, but the replicator-centered evolution naturally operates also within and between prokaryotic cells. Indeed, there is a vast diversity of different forms of genetic replicators that use prokaryotic cell vehicles for their preservation and propagation. Any particular prokaryote that lives in this biosphere, being that a bacterium on your forehead or an archaeon in the bottom of Pacific Ocean, harbors a chromosome but may also host a collection of other replicators, including plasmids, transposons, and viruses. Some of the replicators, like conjugative plasmids and viruses, are able to actively move between available vehicles in its environment, thus making these replicators less dependent on the survival of any particular lineage of cell vehicles. Therefore they are not an inherent part of any particular bacterium and may thus be considered as distinct forms of genetically replicating entities that utilize cells for their propagation and survival (similarly with the viruses in Forterre's virocell concept).

The continuous struggle for existence within and between prokaryotic vehicles modifies the phenotypes of the replicators. A lot of theoretical and experimental work has been done in order to clarify the functions and the evolutionary trajectories of viruses, bacterial cells, and plasmids in different ecological contexts and under various selection pressures. However, in this discussion I take a step away from any particular type of a replicator or an organism and explore from a general perspective whether the lateral movement potential (or lack of it) of the replicators could help illuminate some evolutionary aspects of the prokaryotic biosphere. This discussion attempts to provide an intuitive view on the selfish genes and various types of replicators in bacterial and archaeal cells. It is my intention to keep the text simple and readable regardless of the reader's expertise on bacteria, viruses, plasmids, or, for that matter,

evolutionary theory. Moreover, given the vast amount of details in microbial world, I hope that the readers realize that certain corners had to be cut in various places in order to keep the text within realistic length.

Furthermore, in an attempt to maintain the simplicity, the following nomenclature and definitions are used throughout this paper. A *cell vehicle* denotes a prokaryotic cell with membranes, resources, and everything else but excludes any genetic material. *Cell-vehicle lineage* indicates a single vehicle and its direct descendant that emerge by cell division. A *replicator* is any discrete enough collection of genetic material (that seems worth naming), which utilizes the cell vehicle for its preservation and propagation. Replicators are replicated as distinct units forming a coherent collection of genetic material that can be separated with reasonable effort from other replicators. Replicators may be replicated as a part of the replication of other replicators, as integrative viruses are replicated along with host-chromosome multiplication, but essentially these two replicators can be denoted as two distinct entities given that the integrative virus can replicate its genetic information also separately from the replication of the chromosome. The mean by which the genetic information of a replicator is replicated is not relevant. However, I prefer to not make a too strict definition for a replicator as it is likely to lead to unproductive hair-splitting arguments. Yet, it must be noted that replicators do not include ribosomes or other nucleic acids containing molecules that essentially have an enzymatic function but that are not used as template for their own replication. *Vertical relationship* or vertical inheritance of a replicator indicates that this genetic replicator preserves itself within a dividing lineage of cell vehicles. *Horizontal movement potential* denotes that the replicator is able to introduce itself into a cell-vehicle lineage where the replicator was previously absent. Any feature that is encoded or induced by a replicator is denoted as a *phenotype*. Figure 1 links these terms with their biological counterparts.

3. Laterally Moving Replicators

Prokaryotic world contains a number of different types of replicators that have potential for lateral movement between cell-vehicle lineages. Here I briefly introduce the basic types of laterally moving replicators.

3.1. Conjugative Plasmids. Conjugative plasmids are extrachromosomal assemblies of genetic material that replicate independently within their host vehicles [16, 17]. Conjugative plasmids may encode complex toxin-antitoxin systems and other effectors that ensure that the dividing cell vehicles harbor copies of the plasmid [18]. Conjugative plasmids also encode proteins that facilitate the transfer of the conjugative plasmid from one cell vehicle to another [19]. Conjugative plasmids can spread between distantly related cell vehicles, but one copy of the plasmid is always maintained within the donating cell. Conjugative plasmids have no extracellular stage and are thus dependent on the host cell at all times.

3.2. Integrative and Conjugative Elements (ICEs). Similarly with conjugative plasmids, ICEs can force the host cell vehicle to form a cell-to-cell contact with other cells in the present environment and use this contact for transporting the genetic element from one cell to another [16, 20, 21]. ICEs can spread between distantly related cell vehicles and replicate therein. ICEs integrate into the chromosome during their life cycle and differ from conjugative plasmids in this respect. This integration may often lead to the transfer of some chromosomal genes from one host to another.

Conjugative plasmids and ICEs are known for their antibiotic resistance genes [22]. Arguably the lateral movement of conjugative plasmids and ICEs is responsible for majority of novel drug-resistant bacterial phenotypes in hospitals and other clinically important environments [16]. Conjugative plasmids and ICEs contain variety of different types of genes including those encoding for virulence factors. However, detailed analysis of this genetic variability and their exact functions and/or roles in certain ecological contexts are beyond the scope of this paper.

3.3. Temperate Viruses. Viruses are replicators that enclose their genetic material within a protective protein capsid [3]. This capsid can leave the host cell and introduce the genetic material into a new cell vehicle far away from the initial host. Thus viruses (unlike plasmids) can be transiently independent from the survival of any particular host cell vehicle. The extracellular state of viruses is known as virion, and it should not be mistaken for a virus [7, 8]. Differences between virions and viruses were discussed in the introduction.

The assembly of viral particles often leads to the destruction of the cell vehicle. However, temperate viruses are able to exist peacefully within their host cell as a so-called provirus [23]. During the provirus state no viral particles are produced. Yet, this lysogenic cycle can be interrupted, which then leads into reigniting the virus particle production. Viruses can become integrated into the host chromosome or exist as extrachromosomal genetic elements during the provirus state [24–26]. A lysogenized cell vehicle is (usually) resistant to infections of other related viruses.

3.4. Virulent Viruses. Virulent viruses are incapable of lysogenic life cycle as they do not maintain regulation machinery that would allow them to retain from virus particle production. Virulent viruses destroy the infected cell vehicle at the end of their replication cycle. However, some virulent viruses can sometimes halt their replication cycle when the host cell is going into dormant state [27].

3.5. Passive Movement of Other Replicators. Prokaryotic cell vehicles can harbor other replicators that can occasionally move horizontally between cell-vehicle lineages, but they do not actively encode functions that would facilitate horizontal movement. These replicators include genetic elements like nonconjugative plasmids, and transposons. Plasmids, transposons, and even complete chromosomes can become transferred from one vehicle to another through the same

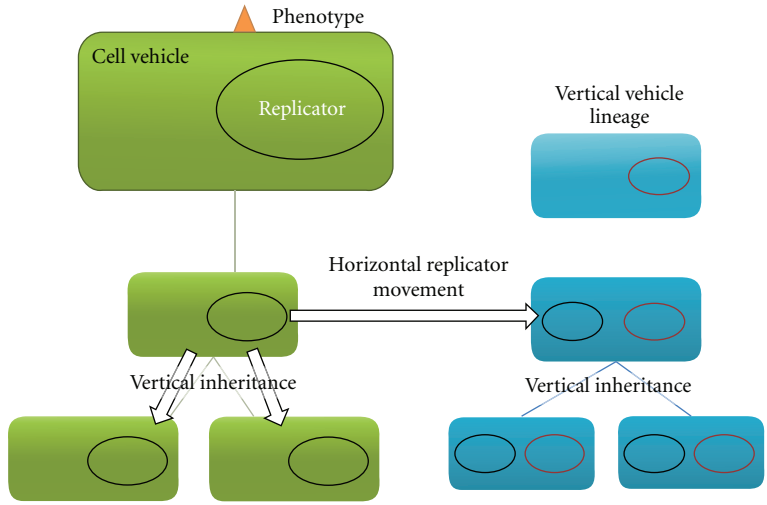


Figure 1: The basic terminology used throughout the paper and their biological counterparts.

conjugation channels that conjugative plasmids and ICEs use. Some plasmids or plasmid-like elements can spread from one cell to another within virus capsid [28] or cell-to-cell connecting nanotubes [29, 30]. Moreover, the natural competence of certain cell vehicles allows the uptake of foreign genetic material from the environment [31, 32], which can lead into horizontal movement of replicators between unrelated vehicle lineages. I will not perform thorough analysis of these various types of ways by which genetic information may become transferred between cell vehicles, but it is important to note that such events occur in natural systems.

4. Replicator Dependency on Vertical Survival of Cell Vehicles

Vertical lineage of a cell vehicle indicates a single prokaryotic cell vehicle and all of its direct descendants that emerge via cell division. If a replicator is exclusively dependent on the survival of a certain cell-vehicle lineage, the replicator would inevitably die along with the lineage. Chromosomes exemplify such a replicator. Yet, certain genes of chromosomes may become horizontally transferred even if the cell-vehicle lineage in general would go extinct (e.g., by transposon-induced transfer and recombination). However, for the clarity of this paper, all such (relatively) random potentials that are not general (enough) features of the replicators are being ignored.

Virulent viruses represent a class of replicators that are not bounded by the vertical survival of the cell vehicle. They even cause the demise of the particular cell vehicle as a part of their replication cycle. Yet, we must realize the limits of such definitions as these are just depictions of the average behaviors of biological entities within reasonable time frames. Naturally even virulent viruses are dependent on the survival of the particular vehicle they are infecting until new virus particles are completely assembled. They are also dependent on the existence of susceptible vehicles in the environment. Nevertheless, it can be argued that, due to their survival strategy, virulent viruses are not dependent on any particular vehicle.

All other replicator types, like plasmids and temperate viruses, are intermediates between virulent viruses and chromosomes in respect to their dependencies on the vertical survival of their current vehicles. This relationship between replicators and vehicles is, naturally, reciprocal as the cell vehicle is not able to survive in absence of the chromosome whereas it fails to survive in presence of a virulent virus. Interestingly, however, this seemingly trivial notion allows us to position the different replicators on a scale where their dependency on cell vehicle appears to (negatively) correlate with their vehicle survival affecting phenotypes (see Section 10). In other words, it is possible that the average phenotype of any replicator matches its position on a chart where lateral movement is on one axis and the vehicle-benefitting phenotype is on the other. Of course, this is just a rough approximation and only an artificial depiction of the result of natural selection repeatedly acting on the replicators. Yet, it can provide a tool to describe the average behavior of prokaryotic replicators. Before addressing this aspect in greater detail, we need to analyze and classify the replicators in a more definitive manner.

5. Classification of Replicators

Most (if not all) of the different types of replicators that utilize prokaryotic cell vehicles for preservation and propagation can be classified according to their horizontal movement potential between individual cell-vehicle lineages and according to their vertical dependencies on cell vehicles. I will attempt to argue that certain phenotypic traits usually associate with replicators of the same class. Subsequently I will discuss the reasons behind this by analyzing few hypothetical scenarios where natural selection might favor the association of these phenotypic traits with horizontally

moving replicators rather than with strictly vertically evolving chromosomes. The classification is presented in Table 1.

My attempt was to retain the classification as simple as possible while maintaining the essential insights that may be derivable from it. Yet, it must be noted that strict boundaries cannot be drawn between different classes because this classification seeks to group together highly different and usually unrelated biological entities. Indeed, there are numerous cases where replicators have changed their present classes and have done this rapidly in evolutionary terms. For example, many chromosome-integrating proviruses (*Class IV*) are known to have become defective viruses by conjoining with a *Class I* replicator (chromosome) and thus becoming only a vertically inherited element [23]. Conjugative plasmids (*Class III*) are known to have become conjugation-defective plasmids (turning into *Class II* replicators) [13], and it has been noted that homologous genetic elements can belong to multiple different classes [10, 11]. In other words, the classes do not represent any permanent characteristics of the replicators. Therefore it seems appropriate to ask whether assignment of replicators into any of the classes is able to catch any practical attributes of an evolving biosphere (and thus justify its formulation). In the remainder of the paper I will attempt to address this question from few different perspectives. For example, I will argue (with some examples) that by changing a *class* the replicator starts to evolve towards other replicators within that group. This suggests that repeated rounds of selection on the replicator can have a general trend in shaping the replicator into a typical member of its *class*. Nevertheless, the complexity of the actual natural systems and the shortcomings of such classifications in relation to this complexity are discussed to some extent.

Finally, I want to emphasize that this classification only attempts to provide a tool to improve our means for understanding and discussing the evolution of prokaryotes and their genetic elements. Some might find the classification trivial or obvious, but I believe that it can help some of us simplify the vastly diverse prokaryotic world into evolutionarily useful components. In any case, there are various types of genetic elements in this biosphere that express different phenotypes and vary in their potential for horizontal movement between cells. It seems very likely that the phenotype is associated at least for some parts with the movement potential. Assuming the opposite (that the horizontal movement potential is not related with the phenotype) seems impossible as you may consider the vehicle-terminating virulent viruses as an example of an expressed phenotype (the only mean by which a virulent virus may survive is due to its horizontal movement between vehicles). Therefore, whether we find it practical or not, it is possible to group these features to some extent (regardless of the usefulness of the presented classification). Naturally some extensions to the presented classification (like, e.g., inclusion of the notion of plasmid incompatibility with each other and with some chromosomes [33]) can be introduced, if found necessary. However, I tried to avoid any unnecessary complexity in order to keep the classification intuitively comprehensive.

6. Phenotypic Traits of Replicators

In this section I will go through the usual phenotypic traits of each replicator *classes*. However, it must be noted there are many replicators that have minor or major exceptions to the general traits within each *class*. In other words, replicators in general form a highly diverse group of genetic entities that utilize cell vehicles for replication and preservation in various environments and in various ecological contexts. Yet, general approximations may be done to some extent.

6.1. Class I: Prokaryotic Chromosomes. Chromosomes are the main genetic replicator in cell vehicles. It segregates into both daughter cells during division. It is often considered that any prokaryotic cell is "equal" to its chromosome. Indeed, when studies attempt to identify the genus of a bacterium, the ribosomal genes or some other highly essential chromosomal genes are selected for sequencing. By determining the divergence of sequence of that gene in comparison to other homologous genes in other cell vehicles, it is possible to assign the taxonomic position of the bacterium. This indicates that many chromosomal genes are absolutely essential for the survival of the cell vehicle, and therefore they can be reliably used to determine the evolutionary histories of both the chromosomes and their corresponding cell vehicles (even if I here treat chromosomes and vehicles as distinct and separate components of a cell organism).

The survival of the chromosome replicator is tightly interlocked with the survival of its current cell vehicle. Natural selection favors any phenotypic change in the chromosome that improves the reproductive success and survival of the cell vehicle. In other words, a favorable mutation (or other genetic change) in a chromosome should not decrease the fitness (or increase the reproductive cost) of the cell vehicle. However, evolutionary process within actual populations of prokaryotes is very complex process (even if other replicator types are not involved), and selection may operate on levels above individual cell vehicles. Yet, for the purposes of this discussion, the correlation of the fitness of the chromosomal replicator with the fitness of the cell vehicle is satisfying enough.

6.2. Class II: Plasmids and Transposons. Plasmids are circular or linear DNA molecules that replicate independently to chromosomes within cell vehicles. However, plasmids always require certain genetic products of chromosomes (being those ribosomes, DNA polymerases, or something else). The sizes of their genome vary from a few kilobases to hundreds of thousands of bases.

Plasmids rely on few different strategies to ensure their survival within the dividing host vehicles. They can encode molecular mechanisms that separate the plasmids along with the chromosomes. Some plasmids contain genes for a toxin-antitoxin system. Plasmid encodes both a stable toxin and unstable antitoxin. The stable toxin will destroy host vehicle, if the vehicle does not contain a copy of the antitoxin-producing plasmid. The plasmids that have

Table 1: Classification of replicators.

Class	Example replicators	Vertical dependency	Horizontal movement potential	Description of average phenotypes
I	Prokaryotic chromosomes	*Completely dependent*	*No potential*	Encodes the main functional units of all cell vehicles. Required for the binary fission of the cell vehicle.
II	Plasmids, transposons	*Highly dependent*	*Passive*	Low reproductive cost to host cell vehicle. Can encode opportunistically useful phenotypic traits.
III	Conjugative plasmids, integrative and conjugative elements (ICEs)	*Moderately dependent* (always requires a cell vehicle)	*Active* without an extracellular stage	Moderate or low reproductive cost to host cell vehicle. Usually encode opportunistically useful phenotypic traits.
IV	Temperate viruses	*Somewhat dependent* (can survive even if the cell-vehicle terminates)	*Active* with an extracellular stage	Moderate or low reproductive cost to host cell vehicle. Sometimes encode opportunistically useful phenotypic traits.
V	Virulent viruses	*Not dependent*	*Active* with an extracellular stage	Insurmountable reproductive cost that terminates the host cell vehicle. Does not encode cell-vehicle benefitting traits.

either segregation or toxin-antitoxin system (or both) usually control the copy number of plasmids within cell vehicles [34]. These plasmids are large in their size, and thus each copy of the plasmid is a burden to the general reproductive rate of the cell vehicle. Similarly with temperate viruses, plasmids are able to prevent other vehicles harboring similar plasmids to conjugate with their present vehicle [35].

Smaller plasmids may not encode sophisticated segregation mechanisms, but instead they can exist in high numbers within cell vehicles (tens to hundreds of copies) and are stably maintained due to the high probability that the dividing cell will contain a copy of the plasmid in both daughter cells.

Several studies have shown that the presence of plasmids in cell vehicles increases the reproductive cost of the cell. In other words, when cells without and with plasmids are grown in similar conditions, cells without plasmids are able to reproduce more rapidly. Moreover, cell vehicles themselves are generally not dependent on their plasmids. From this perspective it is obvious that the plasmid has to ensure its survival within the vehicle. Should the plasmids decrease the cost of reproduction of the cell vehicle, then selection would favor plasmid-containing cells over plasmid-free cells even without any encoded survival mechanisms.

However and despite the general burden of plasmid, they can sometimes greatly improve the reproductive success of the cell vehicle. Antibiotic resistance genes are often part of plasmid replicators [16, 36]. Other plasmids have genes that help the cell vehicle utilize rare resources when nutrients are scarce. Plasmids can also encode toxins that help the cell vehicle destroy surrounding cells, like human tissues, and thus utilize the resources from these cells for their own benefit [37]. The reasons behind the existence of these genes in *Class II* (and *III*) replicators are discussed later.

6.3. Class III: Conjugative Plasmids and Other Conjugative Elements. Conjugative plasmids are extrachromosomal genetic elements similar to *Class II* plasmids. However, their existence within a cell vehicle changes the vehicle phenotype by such that the cell can form conjugation channel between its current vehicle and another vehicle in the surrounding environment. Through this channel *Class III* plasmid transfers itself into another cell vehicle. Conjugations put a reproductive cost on the hosting cell vehicle, and thus plasmids can regulate its repression as well as inhibit superconjugation with vehicles that already contain a copy of *Class III* plasmid [35]. Conjugative elements can respond to the stress of the host vehicle, like the presence of antibiotics in the environment, and ignite transfer of the element to other vehicles [38].

Conjugative plasmids use similar and homologous mechanisms for their stable maintenance within vertical cell-vehicle lineages with nonconjugative (*Class II*) counterparts. *Class III* replicators often contain antibiotic resistance genes, and studies suggest that *Class III* replicators are the main cause behind the emergence of clinically relevant bacteria resistant to antibiotics [16].

6.4. Class IV: Temperate Viruses. Temperate viruses can produce virions, that is, the infectious virus particles, and therefore exist in a "dormant" state in the extracellular environment. However, they can also vertically coexist within cell-vehicle lineages along with *Class I*, *II,* and *III* replicators. Temperate viruses may integrate into the host chromosome and replicate as a part of *Class I* replicator during cell division. This integration, however, does not abolish the ability to move horizontally between vehicle lineages.

The genomes of temperate viruses may contain genes that are beneficial to the reproduction of their host vehicles (under certain conditions). Presence of a provirus can transform an avirulent bacterium into a virulent one by providing genes for different types of toxins [39]. These toxins can, for example, allow the bacterium to destroy host tissues. Proviruses may also change the host-vehicle phenotype so that it cannot be recognized by eukaryotic immune systems [40].

Class IV viruses are able to detect the malfunction, damage or stress of their host cell vehicles. Proviruses react to these signals by igniting the production of virus particles

[41]. In other words, temperate viruses can predict the upcoming interruption of the vertical cell-vehicle lineage and readily progress into expressing their horizontally moving phenotype. As temperate phages are not dependent on the survival of the host vehicle, they often destroy the doomed vehicle themselves as a part of their lytic life cycle.

6.5. Class V: Virulent Viruses. Virulent viruses also produce virions and thus spend part of their life cycle in the extracellular environment as inert particles. Virulent viruses exclusively destroy the host vehicle as part of their life cycle. Virulent viruses generally do not contain genes that would benefit the vertical survival of the host cell-vehicle. The genetic content of *Class V* replicators appears to aim to effectively utilize the resources of cell vehicles in order to produce multiple horizontally moving virus particles. This, however, does not mean that virulent viruses are simple. Many lytic viruses, like T4, can independently encode essential functions such as some transfer RNA genes, and, indeed, T4 is one of the most complex bacteriophages described to date [42].

7. How Replicators Benefit from the Horizontal Movement between Vehicle Lineages?

Why should a replicator change or move to another vehicle lineage? It is not always obvious why the horizontal movement can be beneficial for a replicator. Indeed, without acknowledging the horizontal movement potential, it appears difficult to understand why bacterial cells or independent replicators have certain types of genes or phenotypes. By realizing that bacterial cells themselves are not always the actual units that are targeted by natural selection can help adopting a truthful image of the microbial world. In this section I consider few simple hypothetical cases that exemplify the effects of horizontal movement on the evolution of replicators and on bacterial organisms.

However, it must be pointed out that this section does not aim to provide any general models or prove any concepts, but instead it is an attempt to intuitively promote the way by which we see the replicators as dynamic components of cell-vehicle populations. The following scenarios are artificial, but their simplicity may help grasping the essence behind the evolution of horizontal movement potential.

7.1. Benefit of Being a Plasmid (Class II and Class III). Imagine a world consisting of hundred independent cell-vehicle lineages. Each of these lineages contains only a single cell that reproduces as fast as it dies, keeping the number of each cell-vehicle type effectively at one. All the lineages replicate and die at identical rates in ultimate resources, and thus the proportions of each cell-vehicle type remains the same. In practice, there is no evolution in this system. By definition this means that the genetic composition of the population is not changing in respect to time.

However, assuming that one of the hundred lineages contains its genetic information in two independent replicators: a chromosome and a reproductively costless plasmid, given that the plasmid has a potential for horizontal movement between vehicle lineages, then the separation of these replicators into two distinct entities already brings evolution to the system.

In the beginning the plasmid is present only in one percent of the cells in the world. Yet, sometimes after cell death the plasmid is released into the environment. From the environment it has a tiny chance to become introduced into a new cell-vehicle lineage. Each new transformed lineage increases the proportion of the plasmid by one percent and further contributes to the plasmid spread rate. Eventually the plasmid would be present in all of the cell vehicles, and therefore, in comparison to the initial chromosomal partner of the plasmid, the plasmid will be hundred times more successful in terms of prevalence among vehicle lineages. The simple existence of a replicator in an extrachromosomal form with tiny chance for horizontal movement has given it the potential to become by far the most abundant replicator in the system. This simple mind exercise can provide us with a glimpse of the underlying forces of natural selection that operates in actual biological systems. But why should natural selection favor the maintenance of the extrachromosomal form of the plasmid? Why not integrate with the chromosome after entering the cell? If some of the plasmids had permanently integrated to the chromosomes, they would have ceased transforming new cell vehicles into plasmid-containing lineages after the death of the bacterial organism. Thus, as long as there are plasmid-free vehicles available in the system, some of the plasmids may retain their extrachromosomal status as it facilitates the spread (as depicted in Figure 2).

Now consider how the introduction of reproductive cost on the plasmid replication would change the system. Or what if the plasmid somehow evolved a more effectively spreading phenotype and sometimes the plasmid could be lost due to segregation infidelity? Or if the plasmid contained genes that can sometimes increase the reproduction rate of the hosting cell vehicle while they put a general fitness cost on the host? Some of these questions are discussed below. Yet, such complexity is the reality of the ecological dynamics of plasmids in natural environments, and thus these mind games can only provide a platform from which to dive into the real world.

Nevertheless, it appears reasonable to assume that under certain conditions evolution may favor extrachromosomal genetic elements, such as plasmids, that can occasionally join previously plasmid-free cell vehicles. Plasmids also benefit from being as little reproductive cost to their host cells as possible. However, we immediately notice that the faster the plasmid can spread among plasmid-free cells, the faster it takes over the cell-vehicle populations. If there were hundred million cell-vehicle lineages instead of a hundred, even tiniest changes in the rate of spread would hugely affect the reproductive success of the plasmid (given some restricted time window for observing the success). Many studies have tackled the details of the interplay between the spread rates and reproductive costs of plasmids [19, 35, 43]. Theoretical work suggests that certain parameter values generally allow the stable maintenance of plasmids in a (sub) population of

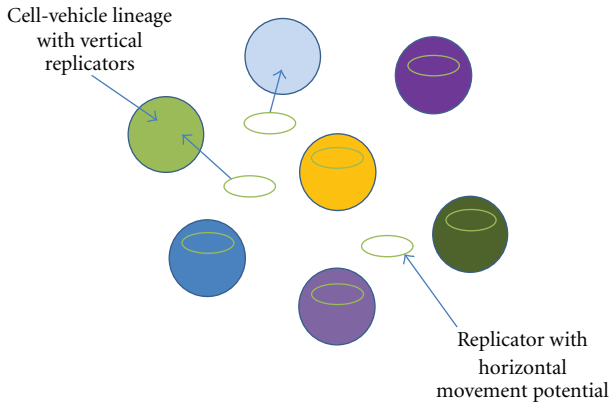


Figure 2: Replicators with horizontal movement potential can become common in various cell-vehicle lineages and therefore free of the survival of any particular lineage.

cell vehicles [44, 45]. Nevertheless, the rapid spread leads us to conjugative plasmids, which can actively force their host cell-vehicles to conjunct with plasmid-free cell vehicles in an attempt to transfer the plasmid.

Conjugative plasmids (generally) spread faster than nonconjugative plasmids, and thus, if the two plasmid types were equal in other respects, conjugative plasmids would apparently be evolutionarily more favorable plasmid type. However, the formation of conjugation channels between cell vehicles does not come without a reproductive cost. Indeed, evolutionary research of bacteria often focuses on studying such tradeoffs where one phenotype (e.g., conjugative) is favorable in certain conditions and the other phenotype (e.g., nonconjugative) in alternative conditions. In principle, the conjugative phenotype is practically useless if all cells in the population already harbor a copy of the conjugative plasmid and similarly highly useful when there are plenty of plasmid-free cell vehicles around [43]. Conjugative plasmids always require a cell-to-cell contact for plasmid transfer, indicating that only one (or few) cell(s) at the time can receive the plasmid.

However, as a mind exercise, consider a high copy number nonconjugative plasmid, which can release several plasmid replicators to the environment upon the death of the host vehicle. In principle, each of these replicators has a potential to become introduced into a new cell vehicle, and under ideal conditions high copy number plasmids could spread very fast in a plasmid-free population of cell vehicles. Yet, the naked DNA molecule is fragile in an extracellular environment and the uptake of the molecule requires a competence for plasmid intake from the cell vehicle. In other words, the plasmid will not survive long in the environment and it cannot force the cells to internalize the DNA molecule. Therefore it must be favorable from the perspective of the chromosome or other in-vehicle replicator (as they would encode the competent phenotype of the cell-vehicle) to introduce the new DNA molecule into the cell vehicle. Genes for antibiotic resistances and other beneficial functions can, under certain conditions, significantly increase the fitness of any cell-vehicle lineages. For this reason, opportunistic genes do not need to only improve the survival of their present vehicles but may sometimes also indirectly improve the probability by which the plasmid can spread horizontally to a new cell vehicle lineage and survive within that lineage thereafter. Natural competence, or the uptake of genetic material into the cell-vehicle from its vicinity, is as the name indicates a natural trait of many bacteria [46]. However, there are also many reasons why natural competence can backfire, and, supposedly, for this reason it is not prevalent trait among bacteria.

Nevertheless, plasmids may evolve mechanisms that allow them to hitchhike through conjugation channels build by other plasmids. This allows them to utilize the horizontal transfer potential without the burden of maintaining genetic machinery for it. Plasmids may also favor evolution towards higher copy numbers within a single cell vehicle in order for the highest copy-number plasmid to have the highest chance for getting transferred into new host vehicle. Yet, the increased cost of maintaining most copies can become compensated on population level by the lower reproductive cost that the lower copy-number plasmid put on individual vehicles [47]. As these different aspects hopefully demonstrate, the actual evolution of the phenotypes of plasmids is a complex subject in which several aspects must be considered. It is not immediately obvious which traits are favorable, and thus I want to retain here the more distant perspective on plasmids and other genetic elements.

7.2. Benefit of Being a Virus (Class IV and V Replicators). In previous section it was considered how the release of high-copy-number plasmids into the environment could provide these replicators a high spread rate among vehicles, if the vehicles in the same environment are willing to take in these replicators. However, viruses are able to overcome this barrier of willing uptake by having the extracellular phenotype that forces the intrusion of the replicator into a suitable host vehicle.

Viral life strategy is dependent on the existence of suitable vehicles in the environment. However, given a susceptible population of cell vehicles, viral strategy is the fastest way by which the replicator can spread in the population. For this reason, all cellular organisms are under constant pressure to avoid viral infections. This, in turn, has led to the everlasting evolutionary arms race between viruses and their hosts [48, 49]. Viruses can obviously effectively maintain their life strategy despite the cost that they put on their current host. However, the ubiquity of viruses cannot be understood without taking the cell vehicles and the vehicle phenotypes into account. Indeed, virions, the extracellular forms of viruses, are the most abundant biological entities on our planet [6]. Yet, as Forterre has argued, virions themselves cannot be considered as living organisms in the same respect as cells can. Ultimately, viruses survive because their hosts survive [50].

7.3. Benefit of Being a Chromosome (Class I Replicator). The existence of chromosomes in any cellular organism is so profound to our concept of cells that we might not even

come to think of them as one of the replicators that utilize the cell vehicle for its propagation and preservation. However, in order to distinct the vehicles and replicators from each other under natural selection, we must also address the benefit (and cost) of being a strictly vertically inherited replicator (e.g., a chromosome). To emphasize the reality behind the distinction of replicators from vehicles, it was recently shown that the genome of one bacterial cell vehicle can be replaced by a (closely related) chromosome from another cell vehicle or by an artificially synthesized chromosome [51, 52]. This indicates that the concept of bacterial cell vehicles and their chromosomes is compatible with experiments and therefore their separation is not just a theoretical notion. I discuss here one possible way to approach the evolution of replicators towards a strictly vertical phenotype.

As stated before, all replicators are dependent on cell vehicles for their propagation. The actual living systems have limited resources, and thus the number of cell vehicles rapidly advances to its maximum as the system can support only limited number of cells. This forces the population of cell vehicles to compete for resources. The vertical survival of the vehicle lineage depends on the competitive success of the vehicle. This indicates that for a replicator inhabiting the most successful vehicle *at the beginning of the competition* provides you with most descendants at the end of the experiment—unless, of course, the replicator can horizontally be transferred to other vehicles (as was argued above).

Now, for the sake of argument, let us play with this idea and consider a situation where all the genes within a cell vehicle are separate replicators (these being like very simple *Class II* plasmids). Each gene has a potential for being horizontally transferred between vehicles after cell destruction, but it also has a chance to become lost during cell-vehicle division (depicted in Figure 3). The reproductive success of the vehicle corresponds to the current combination of genes and other genetic information therein as they are responsible for the phenotype of the vehicle. Certain combinations are more successful than others, and therefore they have more descendants within certain timeframe. Some genes are essential for the survival and division of the vehicle, and thus loss of these replicators would terminate the vehicle lineage. Selection should focus on ensuring that the most essential genes are vertically stably maintained as any resources spent on an attempt to divide are wasted unless the essential genes are present in the new cell vehicle. Yet, maintenance of the faithful distribution of thousand individual molecules during a single cell division appears difficult to evolve or heavily costly (given that each of these molecules should have, e.g., an individual type of a segregation system or have regions for chromosome-like segregation), and selection should therefore intuitively progress towards the fusion of these genetic replicators into a single or as few molecules as possible (since this should help the robustness of the segregation during cell division). These replicators would be *Class I* replicators in the presented classification. This is very superficial analysis, yet, it might help grasp the idea that certain genetic functions need to be present within all vehicles at all times, and therefore they would be vertically inherited to all functional cell vehicles during vehicle division.

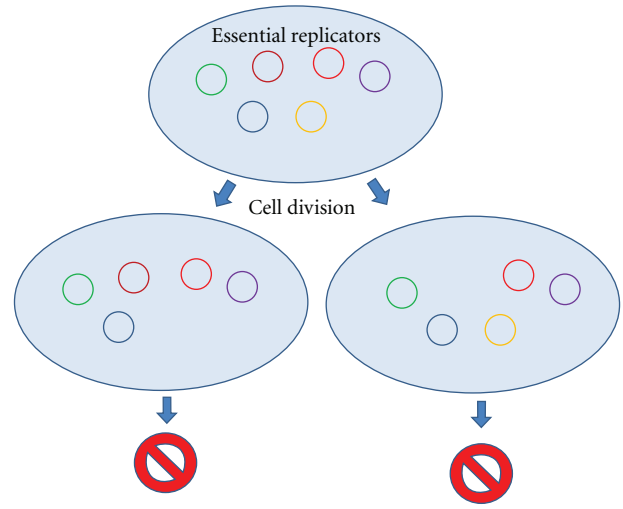


Figure 3: A cell vehicle, which contains its essential genetic information in multiple independent replicators, may be prone to lose some replicators during cell division and thus produce incompetent cells.

8. Replicators Evolving from One Class to Another

In Section 5 I briefly described few examples of replicators evolving into replicators of different *classes*. Now I will go through some examples where the ecological context favors the replicator to adopt the life strategy of replicators belonging to another *class*. Moreover, I will argue that the subsequent evolution of the replicator starts to favor phenotypes that resemble other replicators within its new *class*.

The general point for discussing this evolution is to illustrate that the classification can provide a framework for approaching complex evolutionary settings. Scientific classifications, however, may be harmful for profound understanding of systems, if we are unable to see beyond the classes themselves. Yet, I believe that a proper classification can give a simplifying touch on some of the acting forces of nature. It must be noted that the different classes of the presented classification do not have strict boundaries and replicators can readily change their classes. Still, the possibility to situate the replicators into these *classes* may reflect general evolutionary tendencies of complex microbial systems and thus prove practical in understanding microbial world.

8.1. Temperate Viruses Evolve into Virulent Viruses. Many bacteriophages are known to acquire mutations, which makes them unable to repress their lytic pathway [9, 24, 41]. These *Class IV* replicators lose their potential to exist vertically within a lineage of cell vehicles, and thus they transform into *Class V* replicators in the classification.

These virulent mutants (or so-called clear plaque mutants) enter bacterial cells, replicate their genomes, express their structural proteins, assemble new virions, and lyse the cell. This evolution of *Class IV* replicators into *Class V* replicators is commonly used in bacteriophage research as the "new" *Class V* replicators are devoid of vertical survival within lineages, and therefore their fitness correlates only with their potential for replicating in other vehicle lineages. This, in turn, often increases the production rate of virions [24], which therefore helps conducting experiments that require virus particles. In other words and from the viewpoint of the classification, *Class IV* replicators started to approach the typical phenotypes of *Class V* replicators due to their incapability for vertical existence within a vehicle lineage.

8.2. Conjugative Plasmids Evolve into Nonconjugative Plasmids. Dahlberg and Chao, 2003, cultivated bacterial cell vehicles containing certain conjugative plasmids for 1100 generations (about half a year) [53]. The system did not contain plasmid-free vehicles, and therefore there was essentially no selection for maintaining the horizontal transfer potential of the conjugative plasmid. Indeed, it was observed that some of the *Class III* replicators had lost their potential for conjugation or the rate of conjugation had decreased during the 1100 generations of their host vehicles. Moreover, the reproductive cost of the plasmid had decreased significantly, indicating that selection efficiently focused on improving the vertical survival of the element within its current vehicle lineage.

After invading the whole population of cell vehicles, horizontal movement had no benefits for *Class III* replicator whereas the vertical survival improved its reproductive success. Therefore, the phenotype of *Class III* replicator in this study started approaching that of *Class II* and *Class I* replicators.

8.3. Temperate Viruses Evolve into Chromosomal Elements. Defective bacteriophages are abundant in many bacterial chromosomes [23]. What good does permanent colonization of a certain vertical lineage of cell vehicles do for *Class IV* replicator? Why not maintain the potential for forming the extracellular viral particle and thus the horizontal transfer potential? Indeed, it has been shown that bacterial genomes harboring functional prophages can have advantage over relatives that lack the phage [54].

Given the modern genomics, natural selection operating repeatedly on microbial communities appears to sometimes favor bacterial chromosomes that have defective bacteriophages integrated into them [23]. Naturally, there must be some reason why it is more favorable for the chromosome to maintain a defective provirus rather than a functional one. One possible (and obvious) explanation considers the differences between functional and defective proviruses. A functional provirus can occasionally induce its lytic activity and thus destroy the host cell vehicle (and the chromosome). Those cells that maintain a prophage are immune to infections by other similar viruses as these defective viruses can encode mechanisms that prevent superinfection, that is, multiple infections, of a single cell. However, given that the key elements for producing virions become in some way dysfunctional, then the defective virus becomes unable to destroy the host cell vehicle. In a population of cell vehicles where all chromosomes host a same provirus, then the ones hosting a defective provirus may have an advantage over the others [54].

Moreover, defective proviruses appear to start evolve a strictly vertical life strategy. Studies have demonstrated that the cost for carrying a provirus abates the longer the cells are grown in presence of the virus. Some of the proviral genes belonging to defective proviruses are still expressed within cells, suggesting that the provirus phenotype is benefitting only its present cell vehicle [23, 55]. This illustrates how replicators change their classes and utilize its previous genetic information in support to its new life style.

9. Why Antibiotic Resistance Genes Are Often Associated with *Class II* and *III* Replicators?

Why do bacteria help other, sometimes very distantly related, bacteria in their environment by sharing their antibiotic resistance genes with them? If you think that bacteria are generally competing with other bacteria for available resources, then it appears controversial to realize that the same bacteria are helping their rivals against antibiotic-producing organisms. Should it not be evolutionarily favorable for bacteria to let other bacteria die to antibiotics and thus allow them become the sole survivors of the system? This, however, is not the case when we observe bacteria in environments that are abundant with antibiotics. Have the bacteria allied against us just for the heck of it?

In order to realize why bacteria appear to be cooperating against our attempts to utilize antibiotics as an antimicrobial therapy, we must note that antibiotic resistance genes are often part of independent replicators which are not dependent on any particular bacterial cell [13, 16, 20, 21]. This scenario illustrates how and why the presented dissection of bacterial cell can be useful in comprehending bacterial evolution in environments where their evolution might be the matter of life and death.

It is known that majority of antibiotic resistance genes among clinical isolates of bacteria are actually part of conjugative or nonconjugative plasmids or transposons rather than being an inherent feature of any particular chromosome [16]. The spread of plasmids is considered the most common mean by which bacterial strains transform into drug-resistant phenotypes not only in clinical environments but also within other natural environments [20, 21, 38, 56, 57]. Indeed, antibiotic resistance provides a good example of natural selection where certain genes may become a part of horizontally moving replicators rather than vertical ones.

Once again, I will present a hypothetical scenario (adapted from [58]) that may illuminate how natural selection results in rapidly spreading antibiotic-resistance genes within communities of competing bacteria (depicted in Figure 4). Imagine a system containing ten different bacterial species occupying their individual niches. Each of

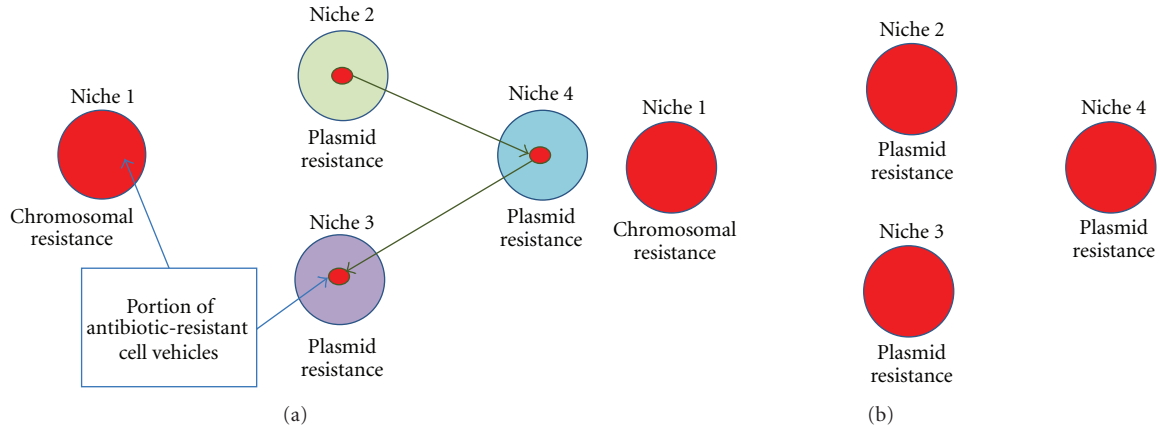


Figure 4: When plasmid- and chromosome-borne antibiotic resistances are compared, the plasmid-borne resistance can become more abundant after exposure to antibiotics.

the bacterial lineages is well adapted to their own niche, and none of the other nine lineages are able to invade these niches. One of the nine vehicle lineages contains an antibiotic-resistance gene in its chromosome whereas one of the lineages contains a conjugative plasmid which carries the same antibiotic resistance gene. The conjugative plasmid poses a reproductive cost to its host cell vehicle, but it moves seldom to other lineages. The plasmid does not become prevalent in any single lineage due to the cost, but a portion of cell-vehicles in each of the lineages ends up harboring the plasmid at all times.

Now, an antibiotics-producing organism enters the environment and subjects all bacterial cell vehicles in all of the ten ecological niches to antibiotic selection. The bacteria will either die or suffer a significant reproductive cost due to the antibiotics that disrupt or terminate the functionality of the cell vehicles. Only those vehicles that happen to contain the antibiotic-resistance gene go unaffected by the antibiotics. The selection results in the death of majority of cell vehicles in the system, leaving room for the remaining cells to repopulate each niche.

Which cell vehicles are likely to occupy the free niches? In this scenario we can imagine two possibilities: either it is the cell vehicle that contains the chromosome with the antibiotic resistance gene or it is one of the cell vehicles that harbor the conjugative plasmid. The fitness of the cell vehicle in any of the niches is likely to correlate with its evolutionary history. In other words, cell-vehicles that previously occupied a certain niche are supposedly best adapted to that niche despite of the presence the plasmid in those vehicles. For this reason the vehicle population containing the chromosomal resistance gene might be unable to conquer any of the niches that suffered from the antibiotic selection despite the fact that the chromosomal resistance lineage itself was not affected by the selection. The result would be that nine of the ten niches became occupied by cell vehicles in which the conjugative plasmid is prevalent due to the opportunistic antibiotic resistance gene, and only one of the ten niches contained the resistance gene in the chromosome.

Horizontally spreading replicators, like plasmids and conjugative plasmids, might not be able to become abundant in cell-vehicle lineages due to their cost to the vehicle reproduction. They can, however, be present in multiple lineages as a minority. This minority of plasmid harboring vehicles with opportunistic genes can provide sudden boost to the vehicle fitness (as described above) and therefore become dominant in the population [58].

10. Do the Replicator and Vehicle Dependencies on Each Other Reflect General Evolutionary Tendencies?

As was argued in Section 4, replicators depend on the vertical survival of vehicles to various degrees. Similarly vehicles fail to survive in absence of certain replicators whereas they fail to survive in presence of other replicators. Chromosomes, for example, are fully dependent on their present lineages while virulent viruses are independent from any particular lineage of vehicles. This allows us to plot these dependencies on an approximate scale where on one axis there is the dependency of the replicators on vehicle lineages and on the other axis there is the effect of the replicator on the survival of its present vehicle (Figure 5). I will attempt to demonstrate that this plot may be useful visualization for approaching the evolution of prokaryotic replicators.

First, we observe that the more dependent a replicator is on a certain vehicle lineage, the more dependent a vehicle lineage is on the replicator. Second, we see that the more harmful a replicator is to a lineage, the less it depends on the survival of any particular vehicle. This correlation may appear to be a trivial tautology, but I suggest that, when we know the replicators' position on one axis, we also know its position on the other. I intend to state here that natural selection may be "aware" of this plot and therefore replicators generally evolve towards the corresponding position on the two-dimensional chart. In other words, if a vehicle cannot survive without some particular replicator, then selection

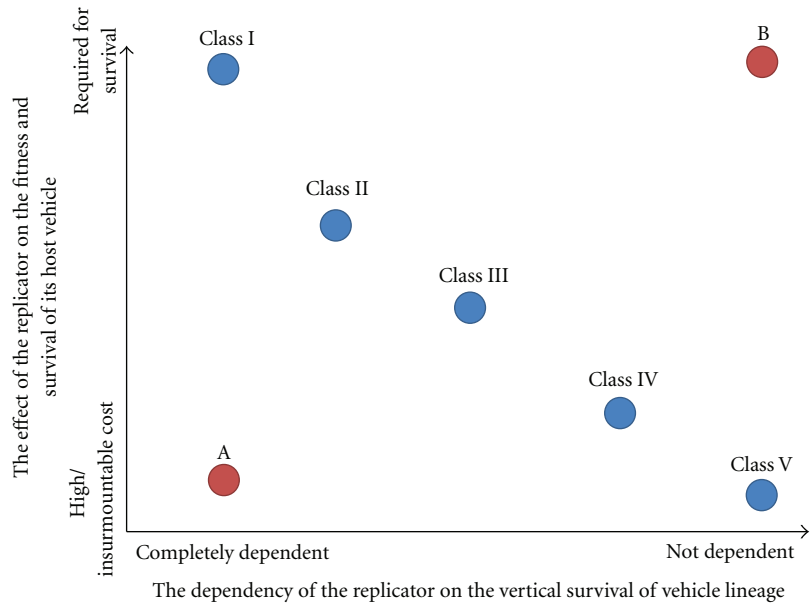


Figure 5: Positioning of the different *classes* of the classification into a two-dimensional plot where on one axis there is the horizontal movement potential of the class and on the other there is the effect of the replicator on its present cell vehicle.

favors changes that make it evolve towards a less horizontal form. Similarly, if a replicator is very costly in terms of reproductive success to its vehicle, then it survives best by being able to move horizontally between vehicle lineages or by evolving a less costly phenotype. While I will not attempt to prove this, I propose this as a hypothesis that may be used as a framework for predicting results of simulations or experiments and also for providing a general perspective on the evolution of prokaryotic biosphere.

For the sake of argument, I put letters A and B (representing imaginary replicators) on the plot (Figure 5) at positions that are free of *natural* replicators. Would it be possible that A and B actually existed in nature? I argue that the answer is negative. However, I want to emphasize that such replicators may, of course, exist transiently, but natural selection favors the change towards their correct positions on either of the axes or, alternatively, they will go extinct altogether. Therefore replicators A and B are not evolutionarily stable replicators with their present life strategies.

Replicator A decreases the reproductive fitness of its vehicle. Therefore any vehicle in the environment that lacks A is able to outreproduce vehicles containing A, leading into the extinction of A. However, A could also achieve potential to be transferred horizontally between vehicle lineages (by recombining with a conjugative plasmid, e.g.), which would make A less dependent on the survival of its current lineage. This means that A would be likely to move rightwards on the two-dimensional plot. The other possibility is that A could evolve into a less costly replicator, making it move upwards on the plot. You may consider a host-destroying virus that makes defective virions as an example of A. This virus should evolve either a phenotype that does not destroy the vehicle or it should form functional virions in order to survive.

The case of replicator B is a somewhat less obvious one. B is essential for its present vehicle, but it is not dependent on the vertical survival of any particular vehicle lineages. In other words, vehicles require B for survival, but B itself can freely move between vehicles. However, if we think of the situation, we realize that (by definition) all vehicles must contain a copy of B in order to survive and reproduce. Therefore B would be present in every single surviving vehicle lineage, and the horizontal movement potential would pose only an unnecessary reproductive cost for the current vehicle. From this perspective it appears logical that B will lose its horizontal movement potential as selection would favor the nonhorizontal and therefore less costly phenotype.

A and B depict two unnatural cases, but they provide an example how natural selection may be operating against these positions in the plot. However, the situation becomes increasingly more difficult when we consider any intermediates between A and B. In natural environments and ecological communities, the position of the replicator on y-axis is likely to constantly change depending on the current surrounding conditions of its present vehicle. In presence of antibiotic-producing organisms, the plasmid providing the resistance might be essential for the cell, but this essentiality ceases when the antibiotic producing organism disappears from the effective area of the vehicle. What is the position of such a replicator on y-axis? Similarly certain replicators might be relatively costly to their host vehicles, but they can sometimes give huge advantage to their vehicles due to seldom occurring conditions. However, the diversity and the complexity of natural environments is vast, and it is easy to get lost into the ocean of details. For this reason the plot should be seen as a tool, which may allow approaching

complex phenotypes of multiple different types of replicators from a very general perspective.

11. An Example Case of the Emergence of a Relevant Bacterial Organism through Accumulation of Multiple Replicators into a Single Vehicle

Vehicle concept and replicators can provide a general way to approach and explain changing behaviors of bacterial organisms. Microbial world is often seen to consist of just bacteria (and archaea) and viruses. These microbes are living on this planet in any suitable habitat, that is being anything from a rectum of an animal to a hydrothermal vent in the mid-Atlantic ridge. This view is not wrong, and indeed it is the one that we observe with our microscopes. Similarly, general books about microbes generally describe a variety of different viruses and prokaryotes with their taxonomic families and evolutionary relationships. However, these books often credit other horizontally moving replicators to lesser extent despite the fact that they may play a significant role in biological systems and that they are arguably distinct entities in respect to any particular bacterium. Moreover, the general view fails to distinguish the different roles of temperate and lytic viruses. Indeed, the chimerical reality of multiple intercellular and extracellular replicators is a fundamental part of bacterial life, and thus acknowledging this diversity can help us realize why and how certain microbial organisms arise.

What kind of an organism was the enterohemorrhagic *Escherichia coli* (EHEC) that was responsible for the outbreak in Germany in 2011 and that tragically killed tens and caused a severe disease in thousands? From the perspective of this paper, it is interesting that replicators of various classes of the presented classification played a role in the outbreak [59].

Mainstream media described EHEC as a common human bacterium that happened to cause a dreadful disease. To people who are unaware of the details of microbial world, the overall image must have been that *Escherichia coli* can sometimes become extremely harmful. How do some of these naturally commensal bacteria happen to turn into hazardous or even lethal pathogens? Naturally, the evolution of virulence is a complicated matter with a variety of affecting factors. Yet, for a realistic approach, we must understand that it can be independent genetic elements that are responsible for forcing the commensal bacterial organisms to turn into the causative agents of epidemics. Indeed, sometimes news articles about EHEC mentioned that bacteria naturally swap genes with each other and this exchange is behind the emergence of this new lethal version of the bacterium. However, it may still appear unclear how bacteria know to transfer these nasty genes into other bacteria and why do they do that. As in case of antibiotic resistances, for profound understanding we must realize that it is not any actual bacterium transferring these genes, but instead that a bacterium is an organism that consists of a cell vehicle along with a chromosome and possibly some other genetic replicators. And these other replicators are the ones that induce the phenotype that is responsible for transferring horizontally genes into other bacteria. And they do it because it is beneficial for their own survival and reproduction.

EHEC behind the Germany outbreak contained temperate viruses (*Class IV*) that provided the Shiga toxin genes responsible for the pathogenic phenotype of the bacterium [60]. In other words, EHEC would not have caused the epidemic if there were no *Class IV* replicator that used the same cell vehicle with the chromosome for its propagation and preservation. Moreover, EHEC strain contained a large conjugative plasmid (*Class III*) that provided the vehicle with antibiotic resistances and some other useful phenotypes. However, EHEC infections are usually not treated with antibiotics anyway as antibiotics may increase Shiga toxin production of the bacterium. Nevertheless, the plasmid may have given the vehicle a potential to survive in environments where it would have naturally succumbed. Overall, by realizing that bacterial cells are combinations of various independent genetic entities, we may understand how new diseases and super bugs emerge from previously harmless organisms.

12. Examples of Using the Classification in Formulating Hypotheses for Evolutionary Experiments

I want to demonstrate that the presented classification could be used to provide a framework for formulating some practical scientific hypotheses (e.g., predicting outcomes of evolution experiments). I give few simple examples that essentially ask whether or not we may approach evolving bacterial populations from the viewpoint of various replicators with differing potentials for horizontal movement and with differing effects on the survival of the cell vehicles.

Opportunistic genes that only sometimes but significantly improve the survival or reproductive rate of a cell vehicle are likely to become associated with horizontally moving replicators rather than Class I vertical replicators in natural communities of bacteria. In principle, this hypothesis could be tested by cultivating a diverse bacterial population in an environment where there are multiple niches available. One of the bacteria would contain the opportunistic gene (like antibiotic resistance) in its chromosome replicator, and one of them would have the gene in a horizontally moving replicator (like conjugative plasmid). The system would be let to grow for some time before and after introducing the antibiotic selection to the system. The prevalence of the opportunistic gene in horizontal replicator instead of the chromosomal replicator could be measured.

If an opportunistic gene associated with a horizontally moving replicator becomes mandatory for the survival of the cell vehicles in the environment, then the replicator associated with the opportunistic gene evolves towards a (more) vertical phenotype or the gene becomes part of one of the vertical replicators. In principle, the hypothesis could be tested by introducing a conjugative plasmid (containing an opportunistic gene, like antibiotic resistance) to a population of bacteria. Then, lethal doses of antibiotic selection would be stably maintained

in the system over several bacterial generations. After the selection, the cost of the plasmid to the cell-vehicle and its conjugation rate could be measured.

If selection focuses against a replicator on which the cell vehicle is not dependent, then the complete replicator can become eliminated. If selection focuses on an essential replicator, then the replicator is likely to only change its phenotype or the whole vehicle lineage becomes terminated. This was the actual hypothesis in a recently published experimental paper by me and my colleagues [13]. We tested what happens when plasmid-dependent phages were cultivated with bacteria harboring plasmids in presence and absence of the selection for the plasmid. In absence of the selection, the plasmid was shown to become lost. In presence of the selection, the plasmid (or the selected parts of it) survived but its phenotype changed.

13. Host Range and the Replicators in the Evolution of Biospheres

In this final section I will consider what the possibility to classify replicators according to their effects on the survival of host cell vehicle and their horizontal movement potential might implicate about the evolution of vehicle- and replicator-based biospheres. For those interested in pondering the development of hypothetical forms of life, this discussion can serve as a (testable) hypothesis about the general trends in the evolution of any living system in this universe.

Replicators that move between vehicles have varying host ranges. By the term host range is meant the portion of cell vehicles into which a replicator can transfer to and subsequently replicate in. A virulent virus can usually infect only a tiny fraction of closely related cells whereas conjugative plasmid can be transferred successfully to a much wider range of unrelated cells. The host range of virulent viruses is narrow whereas the host range of a plasmid is large. Naturally, this is not a coincidence.

Virulent viruses terminate the cell vehicles wherein they replicate. Therefore all the other replicators, especially chromosomes, become eliminated due to virus replication. Selection therefore favors those chromosomes among a population of cell vehicles that produce phenotypes which are unrecognizable by viruses. This has been confirmed in various studies that demonstrate the coevolutionary arms race between viruses and their hosts [61]. On the other hand, conjugative plasmids have been shown to be able to transfer and replicate in a variety of different types of cells. There is stronger selection pressure for chromosomal replicators to avoid viruses than to avoid plasmids. Sometimes avoidance of a plasmid can be lethal whereas avoidance of a virus is rarely harmful. In other words, evolutionary dynamics, in general, force the replicators with higher cost on the host cell vehicle to have narrower host range. Now, it can be asked whether this notion may provide any insights about evolving biosystems. There are already numerous papers about coevolutionary dynamics of viruses and cells [61], about virus-driven evolution [62], and about host ranges [63]. My intent is not to repeat them but instead to try applying a more general perspective on the issue.

Our biosphere is abundant with all the types of replicators of the proposed classification, and therefore we may not consider it relevant to think whether or not this is mere coincidence or a direction towards which *any given biosphere* progresses. But what if we take another independently emerged and evolved (although hypothetical) living system which contains vehicles and replicators? If we go through the replicators in that system, are we able to use this same classification for them as we are for replicators on Earth? Do all the *classes* have at least some representatives in the foreign biosphere? Or are there systems where, for example, only chromosomes or just chromosomes and plasmids thrive?

We need to note that the considered biosphere must be large enough in order for this question to be relevant. When we take a small sample of microbes in our world, we may find that some of the replicators, like conjugative plasmids, cannot be found. Therefore tiny cellular communities may not be able to support the full variety of replicators. But what is the case when we take, let us say, a planet full of microbial life? Can we say with relative certainty that we are going to find plasmids, conjugative elements, and viruses just because that is how natural selection in general tends to shape evolving biospheres that are abundant with single-celled organisms?

In order to approach this question, we may consider biospheres where replicators of some of the classes are absent and evaluate whether or not it is possible that some other replicators will inevitably evolve to represent the missing *class*. In Section 10, I argued that replicators may be evolving towards the *correct* position on the two-dimensional plot presented in Figure 5. Now, if one of the classes depicted in Figure 5 had no representatives in a given biosphere, like there were no *Class V* replicators at all, would some of the other replicators be likely to evolve to fill this free *niche*? I will not go through all the possible cases or scenarios but instead address few general ideas.

If a foreign biosphere completely lacked viruses (that can directly cause the demise of their hosting cell vehicles), what would likely to be different in comparison to our biosphere? Naturally, one can think of a huge number of things. However, perhaps one of the most relevant for our considerations is the notion that there would be no evolutionary arms race between viruses and hosts. Cellular populations would not need to maintain variation against constantly evolving virosphere, and, therefore, in absence of viral-induced selection for variance there might be a huge number of cells that maintain, for example, highly conserved surface components. This could indicate that if a virus emerged, it would be likely to be able to reproduce within a huge population of hosts. In other words, any crudest form of a virus would be likely to have a very wide host range and thus be highly successful in producing copies of itself. Therefore, the naivety of the biosphere due to the lack of previous exposure to viruses might render it highly vulnerable to viral invasion. Given a large biosphere and long-enough timeframe, viral strategy might be bound to emerge sooner or later. Experiments have shown that

bacterial populations unexposed to viral selection tend to be more homogenous in comparison to those with viral predators [62].

What if a system had viruses and chromosomes but was devoid of plasmids? Would plasmids be likely to emerge? To address this question, we may need to consider what the usual characteristics of plasmids in our biosphere are, and then ask whether these characteristics should also become associated with plasmid-like replicators (with higher horizontal movement potential than chromosomes) in any other biosphere. Indeed, plasmids often appear to harbor opportunistic genes, like those conferring antibiotic resistance. Such genes may also be likely to exist in foreign living systems, given that biospheres anywhere should inhabit environments where selection pressures are likely to change according to the current ecological and environmental conditions of the particular cell vehicles. If such opportunistic genes are present, then by reconsidering the mind exercise presented in Section 6 and Figure 4, we may find it logical that *Class II* or *Class III* like replicators may emerge due to local evolutionary dynamics. In other words, opportunistic genes may provide an evolutionarily favorable path for the appearance of smaller low-cost replicators that have increased potential for horizontal movement.

In more general terms, I suggest that it is possible that in large biospheres evolution may progress towards various types of replicators with varying potential for horizontal movement, perhaps even to fill all the slots in the presented classification. Naturally, this suggestion can and must be subjected to variety of different types of experimental tests. Nevertheless, in our biosphere all the different classes appear to be evolutionarily stable strategies as they are abundant and ancient. Therefore, given a sizable enough frame from which to observe evolving systems with cell vehicles and replicators, similar stability may be inevitable to emerge. However, it is still very much possible that these classes may be a feature solely of our type of microbial life. Either way, improved knowledge of the underlying issues would help us understand evolving systems nonetheless.

Finally, I want to emphasize that all of the replicator types we now observe in our biosphere may have emerged before the formation of the first consistently reproducing cell vehicle and chromosome. However, discussing the emergence of all the *classes* as a part of an evolving primordial community is far beyond the scope and length of this paper (although being previously discussed to some extent [14, 15, 64–66]). It may, nevertheless, be possible that the early evolutionary dynamics of emerging life anywhere in this universe may naturally generate replicators with varying potential for horizontal movement between cell vehicles. And as the life advances, the replicators remain as a permanent part of the system.

To conclude, horizontal movement and replicator phenotypes may be approached from a general perspective where we do not pay attention to exact details but rather observe the overall characteristics of replicators in an attempt to understand why and how evolving systems, such as prokaryotic biospheres, may appear to be constructed the way they are. At this time, however, it might be impossible to say whether or not this would be of any practical use or lead to meaningful insights.

Acknowledgments

This work was supported by the Academy of Finland. The author wants to thank the reviewer for helpful comments especially on the host ranges of replicators. WISE is thanked for interesting discussions and sessions (47UMa).

References

[1] E. Hambly and C. A. Suttle, "The viriosphere, diversity, and genetic exchange within phage communities," *Current Opinion in Microbiology*, vol. 8, no. 4, pp. 444–450, 2005.

[2] Y. Yin and D. Fischer, "Identification and investigation of ORFans in the viral world," *BMC Genomics*, vol. 9, article 24, 2008.

[3] P. Forterre and D. Prangishvili, "The origin of viruses," *Research in Microbiology*, vol. 160, no. 7, pp. 466–472, 2009.

[4] F. Jacob and E. L. Wollman, "Viruses and genes," *Scientific American*, vol. 204, pp. 93–107, 1961.

[5] J. M. Claverie, "Viruses take center stage in cellular evolution," *Genome Biology*, vol. 7, no. 6, article 110, 2006.

[6] P. Forterre and D. Prangishvili, "The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties," *Annals of the New York Academy of Sciences*, vol. 1178, pp. 65–77, 2009.

[7] P. Forterre, "Manipulation of cellular syntheses and the nature of viruses: the virocell concept," *Comptes Rendus Chimie*, vol. 14, pp. 392–399, 2011.

[8] P. Forterre, "Giant viruses: conflicts in revisiting the virus concept," *Intervirology*, vol. 53, no. 5, pp. 362–378, 2010.

[9] S. Sozhamannan, M. McKinstry, S. M. Lentz et al., "Molecular characterization of a variant of Bacillus anthracis-specific phage AP50 with improved bacteriolytic activity," *Applied and Environmental Microbiology*, vol. 74, no. 21, pp. 6792–6796, 2008.

[10] M. Jalasvuori, S. T. Jaatinen, S. Laurinavičius et al., "The closest relatives of icosahedral viruses of thermophilic bacteria are among viruses and plasmids of the halophilic archaea," *Journal of Virology*, vol. 83, no. 18, pp. 9388–9397, 2009.

[11] M. Jalasvuori, A. Pawlowski, and J. K. H. Bamford, "A unique group of virus-related, genome-integrating elements found solely in the bacterial family Thermaceae and the archaeal family Halobacteriaceae," *Journal of Bacteriology*, vol. 192, no. 12, pp. 3231–3234, 2010.

[12] V. P. Friman, T. Hiltunen, M. Jalasvuori et al., "High temperature and bacteriophages can indirectly select for bacterial pathogenicity in environmental reservoirs," *PLoS ONE*, vol. 6, no. 3, article e17651, 2011.

[13] M. Jalasvuori, V. P. Friman, A. Nieminen, J. K. H. Bamford, and A. Buckling, "Bacteriophage selection against a plasmid-encoded sex apparatus leads to the loss of antibioticresistance plasmids," *Biology Letters*, vol. 7, no. 6, pp. 902–905, 2011.

[14] M. Jalasvuori, M. P. Jalasvuori, and J. K. H. Bamford, "Dynamics of a laterally evolving community of ribozyme-like agents as studied with a rule-based computing system," *Origins of Life and Evolution of Biospheres*, vol. 40, no. 3, pp. 319–334, 2010.

[15] M. Jalasvuori and J. K. H. Bamford, "Structural co-evolution of viruses and cells in the primordial world," *Origins of Life and Evolution of Biospheres*, vol. 38, no. 2, pp. 165–181, 2008.

[16] P. M. Bennett, "Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria," *British Journal of Pharmacology*, vol. 153, no. 1, pp. S347–S357, 2008.

[17] H. Hradecka, D. Karasova, and I. Rychlik, "Characterization of *Salmonella enterica* serovar Typhimurium conjugative plasmids transferring resistance to antibiotics and their interaction with the virulence plasmid," *Journal of Antimicrobial Chemotherapy*, vol. 62, no. 5, pp. 938–941, 2008.

[18] M. Adamczyk and G. Jagura-Burdzy, "Spread and survival of promiscuous IncP-1 plasmids," *Acta Biochimica Polonica*, vol. 50, no. 2, pp. 425–453, 2003.

[19] F. Dionisio, I. Matic, M. Radman, O. R. Rodrigues, and F. Taddei, "Plasmids spread very fast in heterogeneous bacterial communities," *Genetics*, vol. 162, no. 4, pp. 1525–1532, 2002.

[20] V. Burrus, G. Pavlovic, B. Decaris, and G. Guédon, "Conjugative transposons: the tip of the iceberg," *Molecular Microbiology*, vol. 46, no. 3, pp. 601–610, 2002.

[21] M. I. Ansari, E. Grohmann, and A. Malik, "Conjugative plasmids in multi-resistant bacterial isolates from Indian soil," *Journal of Applied Microbiology*, vol. 104, no. 6, pp. 1774–1781, 2008.

[22] M. I. Ansari, E. Grohmann, and A. Malik, "Conjugative plasmids in multi-resistant bacterial isolates from Indian soil," *Journal of Applied Microbiology*, vol. 104, no. 6, pp. 1774–1781, 2008.

[23] S. Casjens, "Prophages and bacterial genomics: what have we learned so far?" *Molecular Microbiology*, vol. 49, no. 2, pp. 277–300, 2003.

[24] A. Gaidelyte, S. T. Jaatinen, R. Daugelavičius, J. K. H. Bamford, and D. H. Bamford, "The linear double-stranded DNA of phage Bam35 enters lysogenic host cells, but the late phage functions are suppressed," *Journal of Bacteriology*, vol. 187, no. 10, pp. 3521–3527, 2005.

[25] C. Canchaya, G. Fournous, S. Chibani-Chennoufi, M. L. Dillmann, and H. Brüssow, "Phage as agents of lateral gene transfer," *Current Opinion in Microbiology*, vol. 6, no. 4, pp. 417–424, 2003.

[26] H. Brüssow, C. Canchaya, and W. D. Hardt, "Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion," *Microbiology and Molecular Biology Reviews*, vol. 68, no. 3, pp. 560–602, 2004.

[27] W. J. J. Meijer, V. Castilla-Llorente, L. Villar, H. Murray, J. Errington, and M. Salas, "Molecular basis for the exploitation of spore formation as survival mechanism by virulent phage ϕ29," *EMBO Journal*, vol. 24, no. 20, pp. 3647–3657, 2005.

[28] Y. Wang, Z. Duan, H. Zhu et al., "A novel Sulfolobus non-conjugative extrachromosomal genetic element capable of integration into the host genome and spreading in the presence of a fusellovirus," *Virology*, vol. 363, no. 1, pp. 124–133, 2007.

[29] T. A. Ficht, "Bacterial exchange via nanotubes: lessons learned from the history of molecular biology," *Frontiers in Microbiology*, vol. 2, article 179, 2011.

[30] G. P. Dubey and S. Ben-Yehuda, "Intercellular nanotubes mediate bacterial communication," *Cell*, vol. 144, no. 4, pp. 590–600, 2011.

[31] M. G. Lorenz and W. Wackernagel, "Bacterial gene transfer by natural genetic transformation in the environment," *Microbiological Reviews*, vol. 58, no. 3, pp. 563–602, 1994.

[32] J. Davison, "Genetic exchange between bacteria in the environment," *Plasmid*, vol. 42, no. 2, pp. 73–91, 1999.

[33] L. A. Marraffini and E. J. Sontheimer, "CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA," *Science*, vol. 322, no. 5909, pp. 1843–1845, 2008.

[34] J. Paulsson and M. Ehrenberg, "Molecular clocks reduce plasmid loss rates: the R1 case," *Journal of Molecular Biology*, vol. 297, no. 1, pp. 179–192, 2000.

[35] M. P. Garcillán-Barcia and F. de la Cruz, "Why is entry exclusion an essential feature of conjugative plasmids?" *Plasmid*, vol. 60, no. 1, pp. 1–18, 2008.

[36] A. Saeed, H. Khatoon, and F. A. Ansari, "Multidrug resistant gram-negative bacteria in clinical isolates from Karachi," *Pakistan Journal of Pharmaceutical Sciences*, vol. 22, no. 1, pp. 44–48, 2009.

[37] M. Sengupta and S. Austin, "Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria," *Infection and Immunity*, vol. 79, no. 7, pp. 2502–2509, 2011.

[38] P. J. Hastings, S. M. Rosenberg, and A. Slack, "Antibiotic-induced lateral transfer of antibiotic resistance," *Trends in Microbiology*, vol. 12, no. 9, pp. 401–404, 2004.

[39] C. L. Gyles, "Shiga toxin-producing *Escherichia coli*: an overview," *Journal of animal science*, vol. 85, no. 13, pp. E45–62, 2007.

[40] M. R. King, R. P. Vimr, S. M. Steenbergen et al., "*Escherichia coli* K1-specific bacteriophage CUS-3 distribution and function in phase-variable capsular polysialic acid O acetylation," *Journal of Bacteriology*, vol. 189, no. 17, pp. 6447–6456, 2007.

[41] N. Fornelos, J. K. H. Bamford, and J. Mahillon, "Phage-borne factors and host LexA regulate the lytic switch in phage GIL01," *Journal of Bacteriology*, vol. 193, no. 21, pp. 6008–6019, 2011.

[42] V. M. Petrov, S. Ratnayaka, J. M. Nolan, E. S. Miller, and J. D. Karam, "Genomes of the T4-related bacteriophages as windows on microbial genome evolution," *Virology Journal*, vol. 7, article 292, 2010.

[43] J. smith, "Superinfection drives virulence evolution in experimental populations of bacteria and plasmids," *Evolution*, vol. 65, no. 3, pp. 831–841, 2011.

[44] B. R. Levin, M. Lipsitch, V. Perrot et al., "The population genetics of antibiotic resistance," *Clinical Infectious Diseases*, vol. 24, no. 1, supplement, pp. S9–S16, 1997.

[45] F. M. Stewart and B. R. Levin, "The population biology of bacterial plasmids: a priori conditions for the existence of conjugationally transmitted factors," *Genetics*, vol. 87, no. 2, pp. 209–228, 1977.

[46] D. Dubnau, "DNA uptake in bacteria," *Annual Review of Microbiology*, vol. 53, pp. 217–244, 1999.

[47] M. M. Watve, N. Dahanukar, and M. G. Watve, "Sociobiological control of plasmid copy number in bacteria," *PLoS ONE*, vol. 5, no. 2, Article ID e9328, 2010.

[48] J. S. Weitz, H. Hartman, and S. A. Levin, "Coevolutionary arms races between bacteria and bacteriophage," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9535–9540, 2005.

[49] A. Buckling and P. B. Rainey, "Antagonistic coevolution between a bacterium and a bacteriophage," *Proceedings of the Royal Society B*, vol. 269, no. 1494, pp. 931–936, 2002.

[50] F. M. Stewart and B. R. Levin, "The population biology of bacterial viruses: why be temperate," *Theoretical Population Biology*, vol. 26, no. 1, pp. 93–117, 1984.

[51] D. G. Gibson, J. I. Glass, C. Lartigue et al., "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science*, vol. 329, no. 5987, pp. 52–56, 2010.

[52] C. Laitigue, S. Vashee, M. A. Algire et al., "Creating bacterial strains from genomes that have been cloned and engineered in yeast," *Science*, vol. 325, no. 5948, pp. 1693–1696, 2009.

[53] C. Dahlberg and L. Chao, "Amelioration of the cost of conjugative plasmid carriage in *Eschericha coli* K12," *Genetics*, vol. 165, no. 4, pp. 1641–1649, 2003.

[54] L. Bossi, J. A. Fuentes, G. Mora, and N. Figueroa-Bossi, "Prophage contribution to bacterial population dynamics," *Journal of Bacteriology*, vol. 185, no. 21, pp. 6467–6471, 2003.

[55] A. Campbell, "Comparative molecular biology of lambdoid phages," *Annual Review of Microbiology*, vol. 48, pp. 193–222, 1994.

[56] A. Malik, E. K. Çelik, C. Bohn, U. Böckelmann, K. Knobel, and E. Grohmann, "Detection of conjugative plasmids and antibiotic resistance genes in anthropogenic soils from Germany and India," *FEMS Microbiology Letters*, vol. 279, no. 2, pp. 207–216, 2008.

[57] E. Kristiansson, J. Fick, A. Janzon et al., "Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements," *PLoS ONE*, vol. 6, no. 2, article e17038, 2011.

[58] C. T. Bergstrom, M. Lipsitch, and B. R. Levin, "Natural selection, infectious transfer and the existence conditions for bacterial plasmids," *Genetics*, vol. 155, no. 4, pp. 1505–1519, 2000.

[59] O. Bezuidt, R. Pierneef, K. Mncube, G. Lima-Mendez, and O. N. Reva, "Mainstreams of horizontal gene exchange in enterobacteria: consideration of the outbreak of enterohemorrhagic *E. coli* O104 : H4 in Germany in 2011," *PLoS ONE*, vol. 6, no. 10, article e25702, 2011.

[60] D. A. Rasko, D. R. Webster, J. W. Sahl et al., "Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany," *The New England Journal of Medicine*, vol. 365, no. 8, pp. 709–717, 2011.

[61] J. S. Weitz, H. Hatman, and S. A. Levin, "Coevolution arms races between bacteria and bacteriophage," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 9535–9540, 2011.

[62] S. Paterson, T. Vogwill, A. Buckling et al., "Antagonistic coevolution accelerates molecular evolution," *Nature*, vol. 464, no. 7286, pp. 275–278, 2010.

[63] C. O. Flores, J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz, "Statistical structure of host-phage interactions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 28, pp. E288–E297, 2011.

[64] M. Jalasvuori and J. K. H. Bamford, "Viruses and life: can there be one without the other?" *Journal of Cosmology*, vol. 10, pp. 3446–3454, 2010.

[65] M. Jalasvuori and J. K. H. Bamford, "Did the ancient crenarchaeal viruses from the dawn of life survive exceptionally well the eons of meteorite bombardment?" *Astrobiology*, vol. 9, no. 1, pp. 131–137, 2009.

[66] E. V. Koonin, T. G. Senkevich, and V. V. Dolja, "The ancient virus world and evolution of cells," *Biology Direct*, vol. 1, article 29, 2006.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 342482, 5 pages
doi:10.1155/2012/342482

Research Article

Comparative Analyses of Base Compositions, DNA Sizes, and Dinucleotide Frequency Profiles in Archaeal and Bacterial Chromosomes and Plasmids

Hiromi Nishida

Agricultural Bioinformatics Research Unit, Graduate School of Agricultural Sciences, University of Tokyo, Tokyo 113-8657, Japan

Correspondence should be addressed to Hiromi Nishida, hnishida@iu.a.u-tokyo.ac.jp

Received 25 November 2011; Revised 11 January 2012; Accepted 19 January 2012

Academic Editor: Hideaki Nojiri

In the present paper, I compared guanine-cytosine (GC) contents, DNA sizes, and dinucleotide frequency profiles in 109 archaeal chromosomes, 59 archaeal plasmids, 1379 bacterial chromosomes, and 854 bacterial plasmids. In more than 80% of archaeal and bacterial plasmids, the GC content was lower than that of the host chromosome. Furthermore, most of the differences in GC content found between a plasmid and its host chromosome were less than 10%, and the GC content in plasmids and host chromosomes was highly correlated (Pearson's correlation coefficient $r = 0.965$ in bacteria and 0.917 in archaea). These results support the hypothesis that horizontal gene transfers have occurred frequently via plasmid distribution during evolution. GC content and chromosome size were more highly correlated in bacteria ($r = 0.460$) than in archaea ($r = 0.195$). Interestingly, there was a tendency for archaea with plasmids to have higher GC content in the chromosome and plasmid than those without plasmids. Thus, the dinucleotide frequency profile of the archaeal plasmids has a bias toward high GC content.

1. Introduction

DNA base composition, specifically guanine-cytosine (GC) content, is a bacterial taxonomic marker. For example, actinobacteria have high, whereas clostridia have low GC-containing genomes [1]. In addition, assessing the dinucleotide frequency profile, a genome signature, of a genomic DNA sequence is a powerful tool to compare different chromosomes and plasmids [2–6]. In bacterial chromosomes, GC content and DNA size are correlated [7–10]. In bacterial phages, plasmids, and inserted sequences, the GC contents are lower than those of their host chromosomes [11].

Replication of and transcription from plasmid DNA are controlled mainly by factors encoded by the chromosome of the host organism. Therefore, it is hypothesized that the GC content and genome signature of a plasmid are similar to those of the chromosome of the host organism. In addition, it is believed that horizontal gene transfers have occurred frequently via plasmid distribution during evolution [12]. For example, a cell-cell communication system may be distributed among the genus *Streptomyces* using horizontal gene transfer via plasmids [13].

Prokaryotes consist of 2 evolutionarily distinct groups: archaea and bacteria [14]. Comparative genomics in bacteria is very advanced, while the whole genome sequence data of archaea is currently limited. Due to recent developments in DNA sequence technology, more than 100 archaeal genome sequences have been elucidated. In this study, I compared GC contents, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids.

2. Materials and Methods

In this study, 109 archaeal chromosomes, 59 archaeal plasmids, 1379 bacterial chromosomes, and 854 bacterial plasmids were used from the database OligoWeb, searching oligonucleotide frequencies (http://insilico.ehu.es/oligoweb/). According to the annotation of the database OligoWeb, chromosomes and plasmids were distinguished.

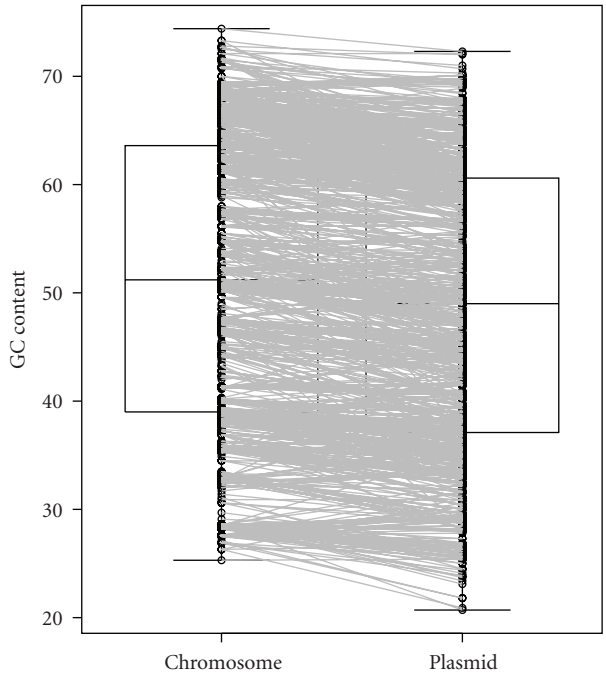


Figure 1: Boxplot of GC contents in bacterial plasmids and host chromosomes. Circles indicate the GC content (%) of each plasmid or chromosome, and lines link each plasmid to its host chromosome. The data set was shown in Supplementary Table S1 available online at doi:10.1155/2012/342482.

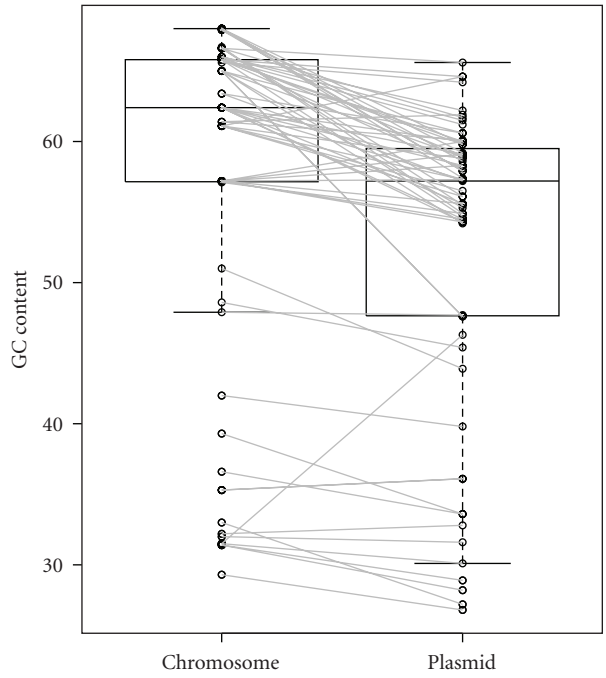


Figure 2: Boxplot of GC contents of archaeal plasmids and host chromosomes. Circles indicate the GC content (%) of each plasmid or chromosome, and lines link each plasmid to its host chromosome. The data set was shown in Supplementary Table S2.

Pearson's correlation coefficient calculation, statistical tests, and drawing plots were performed using the software R (http://www.r-project.org/).

3. Results

The 59 archaeal plasmids and 854 bacterial plasmids are distributed into 26 and 393 organisms, respectively. Some of the archaea and bacteria have 2 or 3 chromosomes. Therefore, in total, the 26 archaeal host organisms and 393 bacterial host organisms have 28 and 441 chromosomes, respectively. The GC contents of bacterial plasmids were found to be lower than those of the host chromosomes (Figure 1, Supplementary Table S1), which is consistent with a previous study [11]. In addition, the GC contents of archaeal plasmids were also lower than those of the host chromosomes (Figure 2, Supplementary Table S2). Furthermore, 777 (81.5%) of the 953 pairs of bacterial chromosome and plasmid, and 57 (85.1%) of the 67 pairs of archaeal chromosome and plasmid showed that the plasmid GC content is lower than that of its host chromosome (Figure 3). In addition, 746 (78.3%) of the 953 bacterial pairs and 47 (70.1%) of the 67 archaeal pairs showed less than 10% difference between GC content of the plasmid and its host chromosome (Figure 3).

The GC contents in plasmids and the host chromosomes were highly correlated in both bacteria and archaea (Pearson's correlation coefficient $r = 0.965$ and $r = 0.917$, respectively; Figures 4 and 5, resp.). Furthermore, in terms

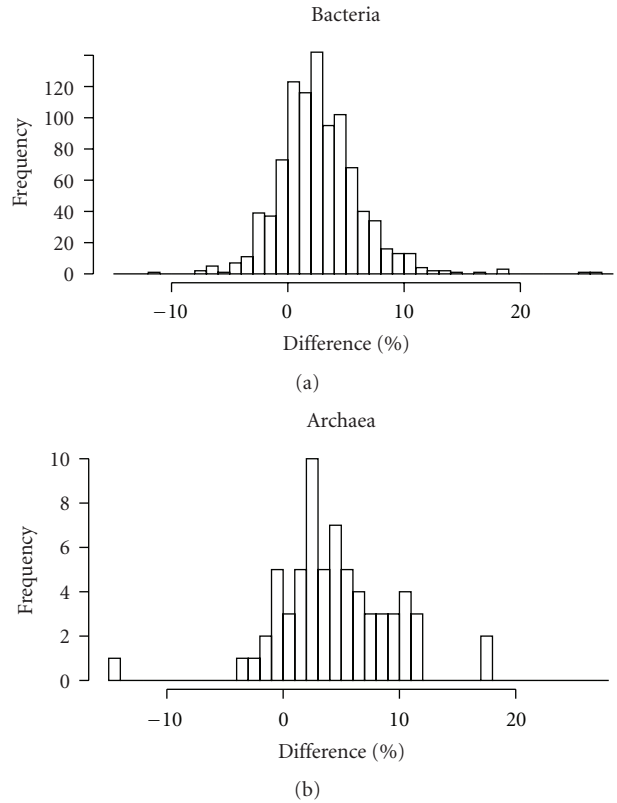


Figure 3: Histogram showing the difference between GC contents of plasmids and host chromosomes. Frequency means the number of pairs of chromosome and plasmid.

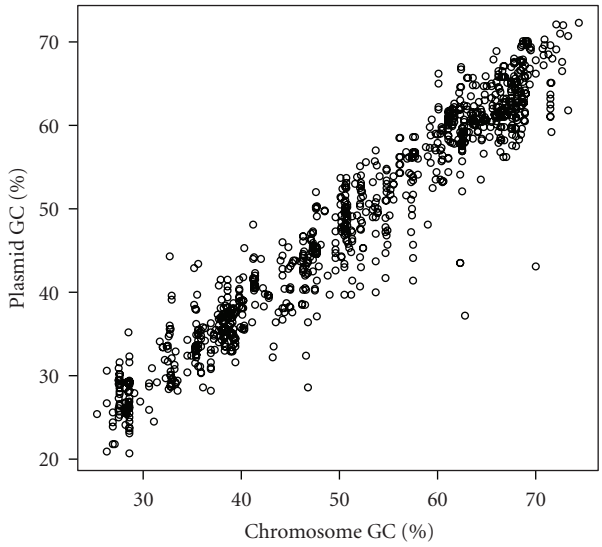


Figure 4: Scatter plot of GC contents of bacterial plasmids and host chromosomes. The Pearson's correlation coefficient is 0.965. The data set was shown in Supplementary Table S1.

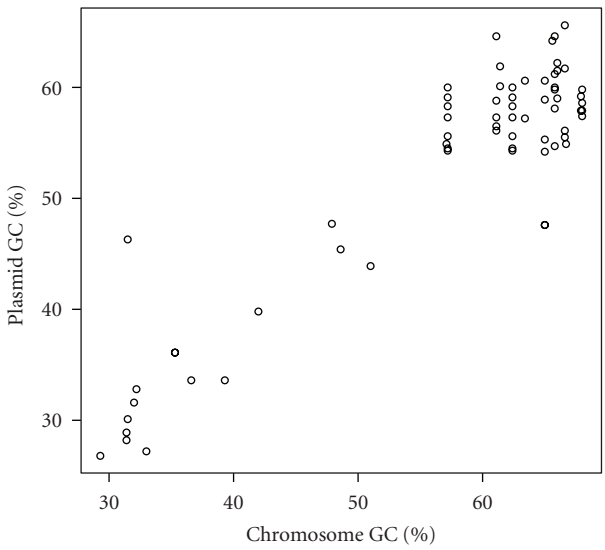


Figure 5: Scatter plot of GC contents of archaeal plasmids and host chromosomes. The Pearson's correlation coefficient is 0.917. The data set was shown in Supplementary Table S2.

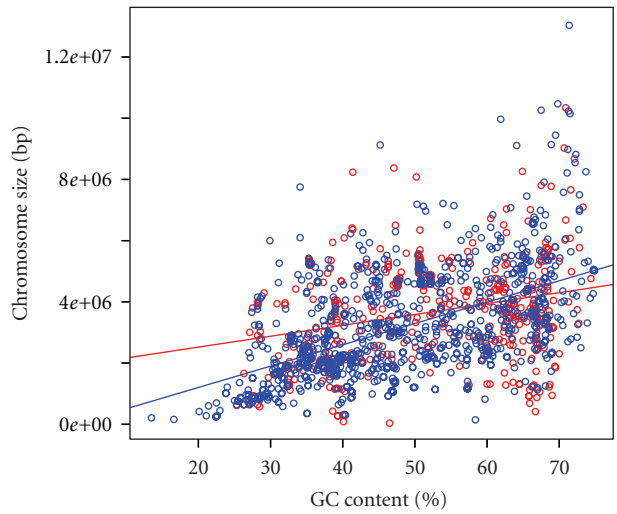


Figure 6: Scatter plot of GC contents and chromosome sizes in bacteria. Red and blue circles indicate chromosomes with and without plasmids, respectively. Red and blue lines indicate the regression lines. The data set was shown in Supplementary Table S3.

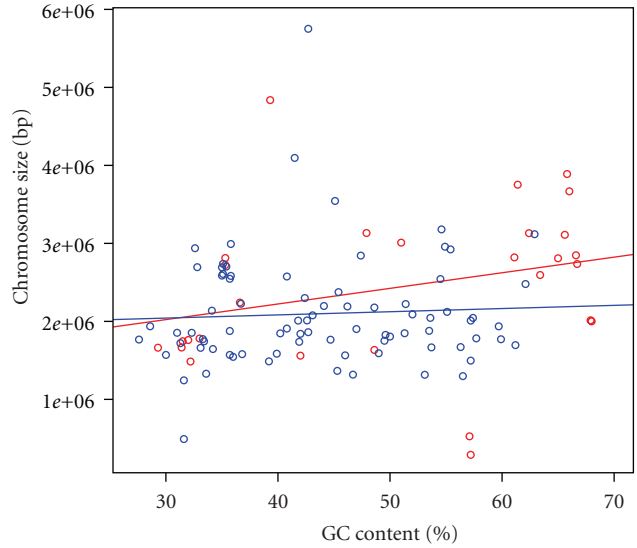


Figure 7: Scatter plot of GC contents and chromosome sizes in archaea. Red and blue circles indicate chromosomes with and without plasmids, respectively. Red and blue lines indicate the regression lines. The data set was shown in Supplementary Table S4.

of size, the GC content and chromosome size were more highly correlated in bacteria than archaea (Figures 6 and 7, Supplementary Tables S3 and S4). Pearson's correlation coefficients between GC content and chromosome size of archaea and bacteria were 0.195 and 0.460, respectively. In archaea, organisms with high GC content chromosome tend to have plasmid (Figures 2 and 7). Thus, the dinucleotide frequency profile of the archaeal plasmids has a bias toward high GC content (Figure 8).

4. Discussion

I hypothesize that GC content, a genomic signature, of a plasmid is related to host specificity and host range. Here, I showed that the GC content of a plasmid is lower than that of its host chromosome (Figures 1 and 2). However, in most cases, the difference in GC content between a plasmid and its host chromosome was less than 10% (Figure 3), strongly suggesting that host organisms cannot maintain and regulate plasmids with very different base compositions.

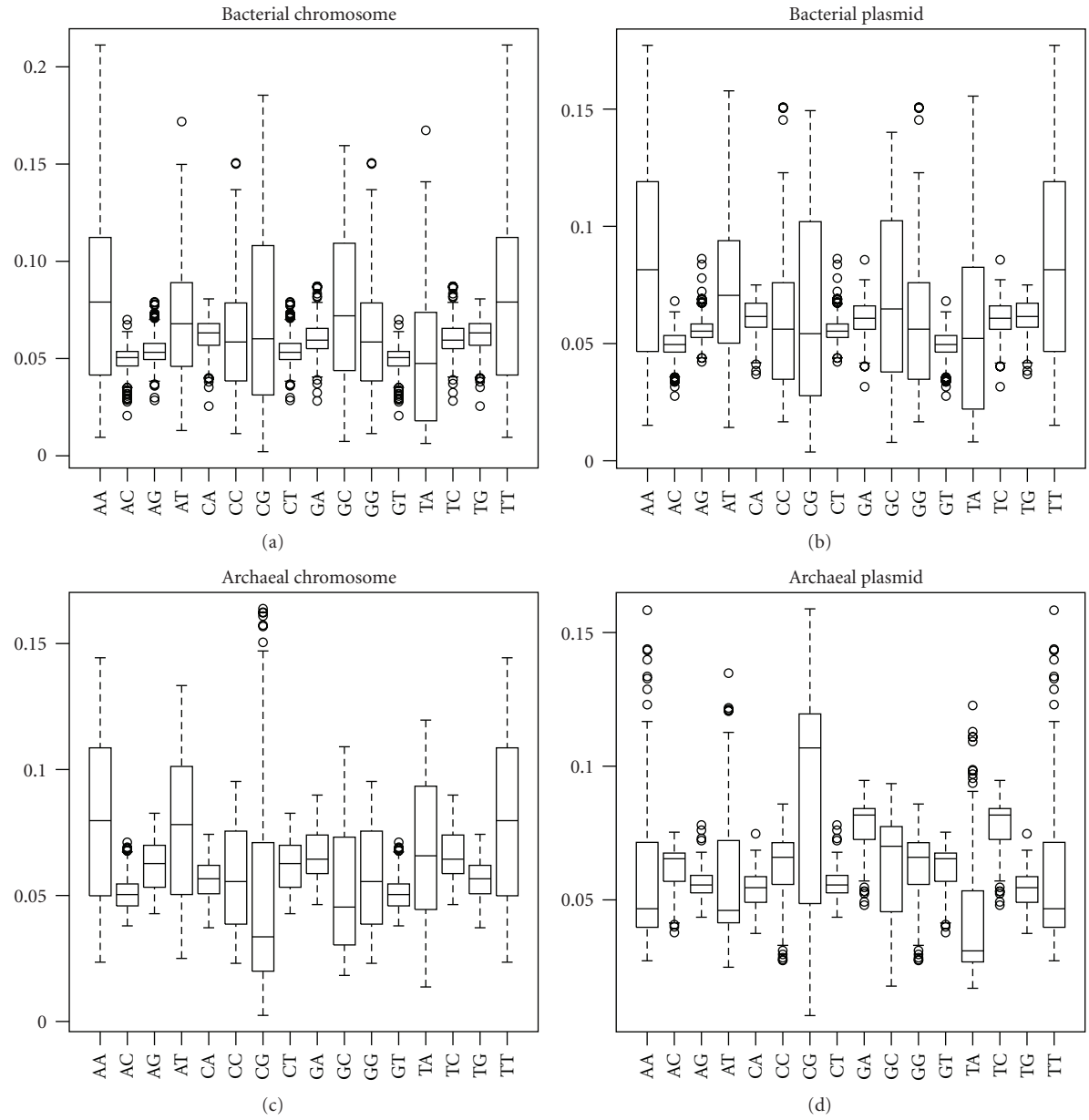


Figure 8: Boxplots of dinucleotide frequency profiles in chromosomes and plasmids of archaea and bacteria. Archaeal chromosomes, archaeal plasmids, bacterial chromosomes, and bacterial plasmids had frequency profiles of 109, 59, 1379, and 854, respectively.

On the other hand, some organisms had a great difference in GC content between their chromosomes and plasmids. For example, in bacteria, *Frankia* symbiont of *Datisca glomerata* has the greatest difference (GC content of the chromosome is 70%; that of the plasmid pFSYMDG02 is 43.1%), and *Desulfovibrio magneticus* RS-1 has the second greatest difference (GC content of the chromosome is 62.8%; that of the plasmid pDMC2 is 37.2%) (Supplementary Table S1). I am so interested in the regulation system for these plasmids.

In this analysis, there was a tendency for plasmid-containing archaea to have higher GC content in the host chromosome and plasmid than those without plasmids (Figures 2, 5, and 7). I have no idea why archaea with mid- and low-GC chromosome tend to lack plasmids. The GC content bias was not found in bacteria (Figures 1, 4, and 6). Thus, although the dinucleotide frequency profiles between the bacterial chromosomes and plasmids were similar, those between the archaeal chromosomes and plasmids were different (Figure 8).

GC content and chromosome size in bacteria are weakly correlated ($r = 0.460$), which is consistent with previous reports [7–10]. However, the GC content and chromosome size in archaea are less correlated ($r = 0.195$). Considering these results, the relationship between GC content and chromosome size may differ in archaea and bacteria. In

order to understand the high GC content bias of archaeal plasmids and elucidate the relationship between GC content and chromosome size in archaea, more archaeal genome sequence data are needed.

Acknowledgment

The author thanks Professor Teruhiko Beppu for his valuable comments.

References

[1] N. Sueoka, "Variation and heterogeneity of base composition of deoxyribonucleic acids: a compilation of old and new data," *Journal of Molecular Biology*, vol. 3, no. 1, pp. 31–40, 1961.

[2] A. Campbell, J. Mrázek, and S. Karlin, "Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 16, pp. 9184–9189, 1999.

[3] M. W. J. van Passel, A. Bart, A. C. M. Luyf, A. H. C. van Kampen, and A. van der Ende, "Compositional discordance between prokaryotic plasmids and host chromosomes," *BMC Genomics*, vol. 7, article 26, 2006.

[4] J. Mrázek, "Phylogenetic signals in DNA composition: limitations and prospects," *Molecular Biology and Evolution*, vol. 26, no. 5, pp. 1163–1169, 2009.

[5] H. Suzuki, H. Yano, C. J. Brown, and E. M. Top, "Predicting plasmid promiscuity based on genomic signature," *Journal of Bacteriology*, vol. 192, no. 22, pp. 6045–6055, 2010.

[6] H. Suzuki, M. Sota, C. J. Brown, and E. M. Top, "Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes," *Nucleic Acids Research*, vol. 36, no. 22, article e147, 2008.

[7] H. Musto, H. Naya, A. Zavala, H. Romero, F. Alvarez-Valín, and G. Bernardi, "Genomic GC level, optimal growth temperature, and genome size in prokaryotes," *Biochemical and Biophysical Research Communications*, vol. 347, no. 1, pp. 1–3, 2006.

[8] D. Mitchell, "GC content and genome length in Chargaff compliant genomes," *Biochemical and Biophysical Research Communications*, vol. 353, no. 1, pp. 207–210, 2007.

[9] F. B. Guo, H. Lin, and J. Huang, "A plot of G + C content against sequence length of 640 bacterial chromosomes shows the points are widely scattered in the upper triangular area," *Chromosome Research*, vol. 17, no. 3, pp. 359–364, 2009.

[10] S. D. Bentley and J. Parkhill, "Comparative genomic structure of prokaryotes," *Annual Review of Genetics*, vol. 38, pp. 771–792, 2004.

[11] E. P. C. Rocha and A. Danchin, "Base composition bias might result from competition for metabolic resources," *Trends in Genetics*, vol. 18, no. 6, pp. 291–294, 2002.

[12] J. Davison, "Genetic exchange between bacteria in the environment," *Plasmid*, vol. 42, no. 2, pp. 73–91, 1999.

[13] H. Nishida, Y. Ohnishi, T. Beppu, and S. Horinouchi, "Evolution of γ-butyrolactone synthases and receptors in Streptomyces," *Environmental Microbiology*, vol. 9, no. 8, pp. 1986–1994, 2007.

[14] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 12, pp. 4576–4579, 1990.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 254941, 4 pages
doi:10.1155/2012/254941

Research Article

Comparative Analyses of Homocitrate Synthase Genes of Ascomycetous Yeasts

Hiromi Nishida

Agricultural Bioinformatics Research Unit, Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo 113-8657, Japan

Correspondence should be addressed to Hiromi Nishida, hnishida@iu.a.u-tokyo.ac.jp

Received 8 December 2011; Accepted 12 January 2012

Academic Editor: Shinji Kondo

Most ascomycetous yeasts have 2 homocitrate synthases (HCSs). Among the fungal lysine biosynthesis-related genes, only the HCS gene was duplicated in the course of evolution. It was recently reported that HCS of *Saccharomyces cerevisiae* has an additional function in nuclear activities involving chromatin regulation related to DNA damage repair, which is not related to lysine biosynthesis. Thus, it is possible that the bifunctionality is associated with HCS gene duplication. Phylogenetic analysis showed that duplication has occurred multiple times during evolution of the ascomycetous yeasts. It is likely that the HCS gene duplication in *S. cerevisiae* occurred in the course of *Saccharomyces* evolution. Although the nucleosome position profiles of the two *S. cerevisiae* HCS genes were similar in the coding regions, they were different in the promoter regions, suggesting that they are subject to different regulatory controls. *S. cerevisiae* has maintained HCS activity for lysine biosynthesis and has obtained bifunctionality.

1. Introduction

Organisms synthesize lysine from 2-oxoglutarate through α-aminoadipate or from aspartic acid through diaminopimelate [1]. Animals cannot synthesize lysine. Fungi synthesize lysine through α-aminoadipate [2–4]. The other eukaryotes synthesize lysine through diaminopimelate. Archaea and bacteria were also believed to synthesize lysine through diaminopimelate until it was reported that the extremely thermophilic bacterium *Thermus thermophilus* synthesizes lysine through α-aminoadipate [5–8].

During lysine biosynthesis in the budding yeast *Saccharomyces cerevisiae*, α-aminoadipate is synthesized from 2-oxoglutarate and acetyl-CoA by the enzymes Lys20 or Lys21 (homocitrate synthase [HCS]), Lys4 (homoaconitase), Lys12 (homoisocitrate dehydrogenase), and α-aminoadipate aminotransferase [9]. Lysine is synthesized from α-aminoadipate by the enzymes Lys2 (aminoadipate reductase), Lys5 (phosphopantetheinyl transferase which posttranslationally modifies Lys2), Lys9 (saccharopine dehydrogenase, glutamate forming), and Lys1 (saccharopine dehydrogenase, lysine forming) [1, 4].

It has been unclear why *S. cerevisiae* has 2 HCSs (Lys20 and Lys21). For example, homocitrate is mainly synthesized through Lys21 during growth on ethanol, while under fermentative metabolism, Lys20 and Lys21 play redundant roles [11]. It was recently reported that Lys20 of *S. cerevisiae* functions in nuclear activities involving chromatin regulation that are distinct from its previously established role in lysine synthesis [12]. Lys20 of *S. cerevisiae* is linked to the DNA damage repair process via the histone acetyltransferase Esa1 and the H2A.Z histone variant [12]. Thus, it is possible that this bifunctionality is associated with HCS gene duplication.

2. Materials and Methods

2.1. Phylogenetic Analyses. I selected 71 HCSs (31 from Saccharomycotina species, 30 from Pezizomycotina species, 2 from Taphrinomycotina species, and 8 from Basidiomycota species) based on BLASTP results in the fungal genome database at NCBI (http://www.ncbi.nlm.nih.gov/projects/genome/guide/fungi/). Multiple alignments were generated with CLUSTAL W. A maximum likelihood tree was reconstructed using MEGA version 5 [10]. The WAG model was

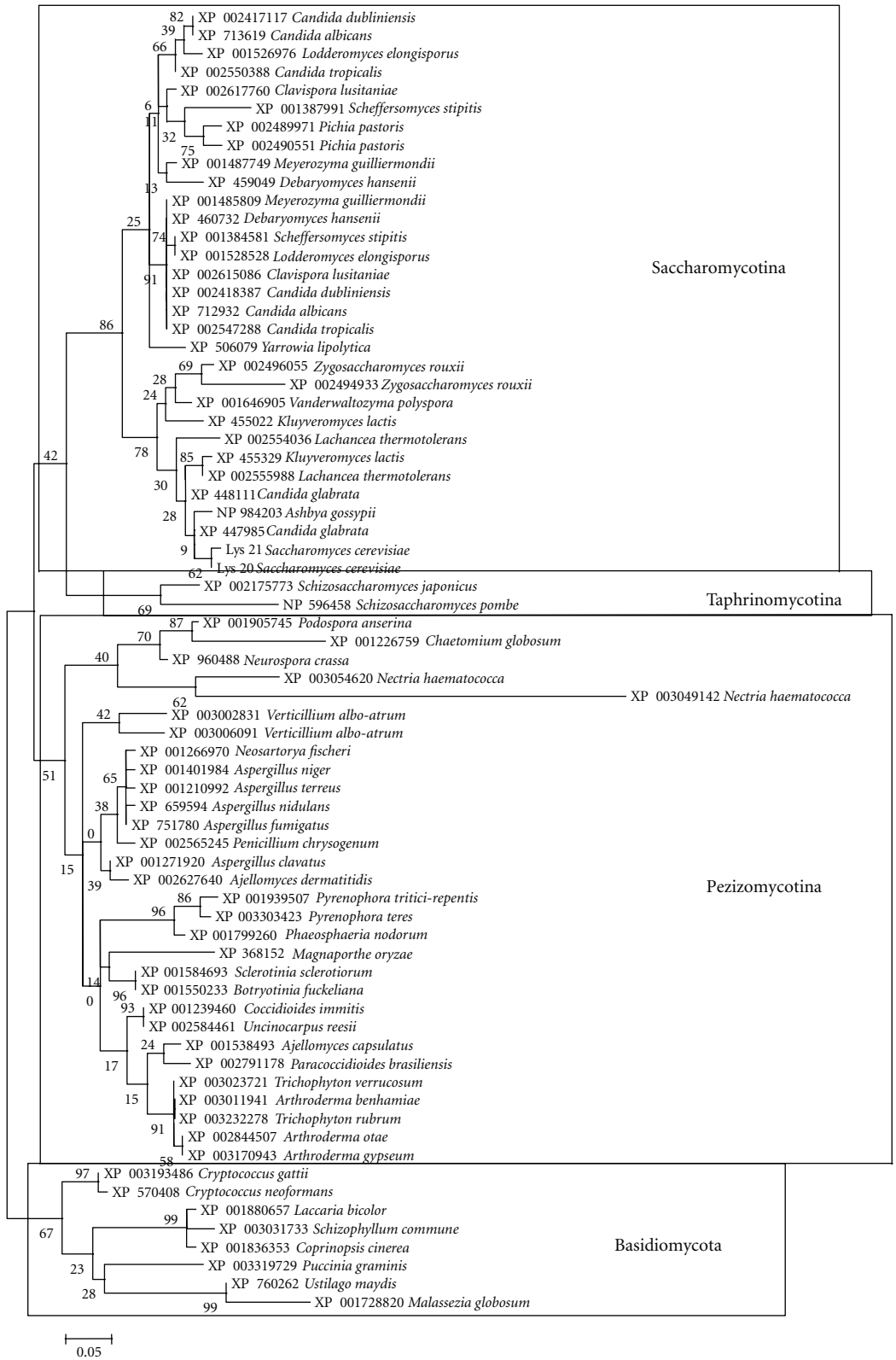


Figure 1: Phylogenetic relationships among 71 fungal homocitrate synthases. The phylogenetic tree was constructed based on multiple alignment with complete deletion of gap sites using the maximum likelihood method of MEGA software [10] with 100 bootstrap analyses. The WAG model was used as the amino acid substitution model. A total of 103 amino acid sites were considered. The γ-distributed rate was considered, and the number of discrete gamma categories was 3. The gamma was 0.81; the discrete rates were 0.14, 0.65, and 2.2.

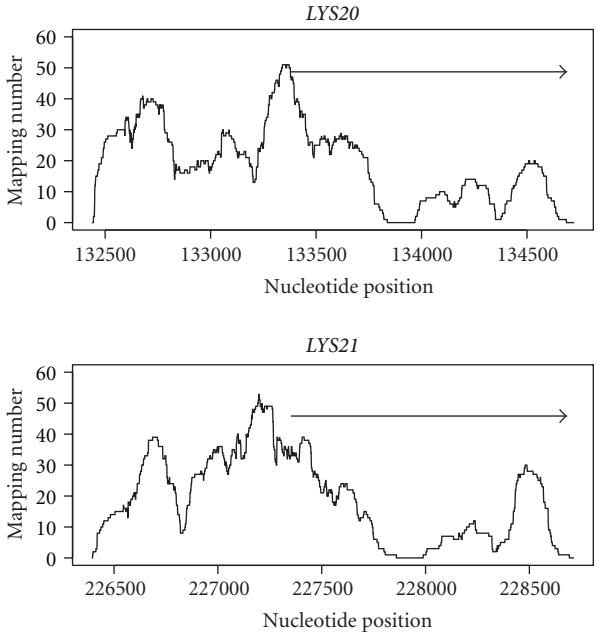


Figure 2: Mapping of nucleosomes around *Saccharomyces cerevisiae* *LYS20* and *LYS21*. In this study, I used nucleosome position data from *S. cerevisiae* BY4741 [13]. Based on each nucleosomal DNA fragment sequence, nucleosomal mapping numbers were estimated for each nucleotide position [14]. Arrows indicate the coding region.

used as the amino acid substitution model. The nearest neighbor interchange was used as the maximum likelihood heuristic method. The γ-distributed rate was considered, and the number of discrete gamma categories was 3.

2.2. Nucleosome Position Comparison. Nucleosome positioning was used to compare gene promoter regions. I used nucleosome position data from *S. cerevisiae* BY4741 [13]. The nucleosome position profiles were compared between the promoter (1000 bases upstream of the translational start site) and coding regions (between the translational start and end site) of the HCS genes, according to a previously described method [14]. Similarity between the two nucleosome position profiles was estimated using the Spearman's rank correlation coefficient.

3. Results and Discussion

The HCS phylogenetic tree (Figure 1) indicates that the HCS gene has been duplicated multiple times in the course of ascomycete evolution. The 31 HCSs of the Saccharomycotina species (ascomycetous yeasts) are encoded in 17 organisms. In contrast, the 30 HCSs of the Pezizomycotina species (filamentous ascomycetes) are encoded in 28 organisms. Thus, 14 of the 17 Saccharomycotina species and 2 of the 28 Pezizomycotina species have 2 HCSs (Figure 1).

Gene duplication is not found in *LYS1*, *LYS2*, *LYS5*, *LYS9*, and their homologues [15]. In addition, no duplication was found in *LYS4*, *LYS12*, and their homologues (data not shown). Therefore, among the fungal lysine biosynthesis-related genes, only the HCS gene has been duplicated. Phylogenetic analysis of HCSs in ascomycetous yeasts showed that the *S. cerevisiae* HCSs (Lys20 and Lys21) are most closely related to each other (Figure 1), suggesting that HCS gene duplication occurred during evolution of the genus *Saccharomyces*. On the other hand, all Saccharomycotina species except *Ashbya gossypii*, *Vanderwaltozyma polyspora*, and *Yarrowia lipolytica* have duplicated HCS genes (Figure 1). Thus, HCS gene duplication may be related to genome duplication events in Saccharomycotina [16–18].

In addition to the phylogenetic analysis based on HCS amino acid sequences, I compared the nucleosome positioning of *LYS20* and *LYS21*. Interestingly, nucleosomes were mapped to the HCS gene promoters more often than to the coding regions (Figure 2). Nucleosome position profiles in the coding regions were highly correlated (Spearman's rank correlation coefficient = 0.833) between *LYS20* and *LYS21*. On the other hand, those in the gene promoter regions were poorly correlated (Spearman's rank correlation coefficient = 0.396). This result suggests that these 2 HCS genes have different regulatory systems.

On the other hand, *LYS20* expression is most similar to *LYS21* expression, and *LYS21* is most similar to *LYS20* expression, based on the SPELL version 2.0.2 [19]. In addition, recent comparative analyses of orthologous genes in evolutionarily close yeasts indicated that divergence of nucleosome positioning is not correlated with divergence of gene expression [20, 21].

Although HCS (Lys20 and Lys21) is located in the nucleus of *S. cerevisiae* [22], HCS is located in the cytoplasm of *Penicillium chrysogenum* [23, 24]. *P. chrysogenum* has a single HCS gene (Figure 1). The phylogenetic tree (Figure 1) showed that gene duplication is not found in Basidiomycota and Taphrinomycotina. In addition, gene duplication has occurred rarely in Pezizomycotina, suggesting that a common ancestor of the Dikarya lacked the nuclear function of chromatin regulation. Considering that duplication of the HCS gene occurred in a limited number of ascomycetes, it may not be an essential event in the evolution of Dikarya. I hypothesize that after divergence of the phyla Ascomycota and Basidiomycota, *S. cerevisiae* obtained HCS bifunctionality.

Acknowledgment

This work was supported in part by a research grant from the Institute for Fermentation, Osaka (IFO), Japan.

References

[1] H. Xu, B. Andi, J. Qian, A. H. West, and P. F. Cook, "The α-aminoadipate pathway for lysine biosynthesis in fungi," *Cell Biochemistry and Biophysics*, vol. 46, no. 1, pp. 43–64, 2006.

[2] H. P. Broquist, "Lysine biosynthesis (yeast)," *Methods in Enzymology*, vol. 17, no. 2, pp. 112–113, 1971.

[3] H. J. Vogel, "Distribution of lysine pathways among fungi: evolutionary implications," *American Naturalist*, vol. 98, pp. 435–446, 1964.

[4] T. M. Zabriskie and M. D. Jackson, "Lysine biosynthesis and metabolism in fungi," *Natural Product Reports*, vol. 17, no. 1, pp. 85–97, 2000.

[5] N. Kobashi, M. Nishiyama, and M. Tanokura, "Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus* lysine is synthesized via α-aminoadipic acid not via diaminopimelic acid," *Journal of Bacteriology*, vol. 181, no. 6, pp. 1713–1718, 1999.

[6] T. Kosuge and T. Hoshino, "Lysine is synthesized through the α-aminoadipate pathway in *Thermus thermophilus*," *FEMS Microbiology Letters*, vol. 169, no. 2, pp. 361–367, 1998.

[7] H. Nishida, M. Nishiyama, N. Kobashi, T. Kosuge, T. Hoshino, and H. Yamane, "A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis," *Genome Research*, vol. 9, no. 12, pp. 1175–1183, 1999.

[8] H. Nishida, "Distribution of genes for lysine biosynthesis through the aminoadipate pathway among prokaryotic genomes," *Bioinformatics*, vol. 17, no. 2, pp. 189–191, 2001.

[9] I. Iraqui, S. Vissers, M. Cartiaux, and A. Urrestarazu, "Characterisation of *Saccharomyces cerevisiae ARO8* and *ARO9* genes encoding aromatic aminotransferases I and II reveals a new aminotransferase subfamily," *Molecular and General Genetics*, vol. 257, no. 2, pp. 238–248, 1998.

[10] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.

[11] H. Quezada, C. Aranda, A. DeLuna et al., "Specialization of the paralogue *LYS21* determines lysine biosynthesis under respiratory metabolism in *Saccharomyces cerevisiae*," *Microbiology*, vol. 154, no. 6, pp. 1656–1667, 2008.

[12] E. M. Scott and L. Pillus, "Homocitrate synthase connects amino acid metabolism to chromatin functions through Esa1 and DNA damage," *Genes and Development*, vol. 24, no. 17, pp. 1903–1913, 2010.

[13] T. Matsumoto, C.-S. Yun, H. Yoshikawa, and H. Nishida, "Comparative studies of genome-wide maps of nucleosomes between deletion mutants of *elp3* and *hos2* genes of *Saccharomyces cerevisiae*," *Plos ONE*, vol. 6, no. 1, Article ID e16372, 2011.

[14] H. Nishida, "Calculation of the ratio of the mononucleosome mapping number to the dinucleosome mapping number for each nucleotide position in the *Aspergillus fumigatus* genome," *Open Access Bioinformatics*, vol. 1, pp. 1–6, 2009.

[15] H. Nishida and M. Nishiyama, "What is characteristic of fungal lysine synthesis through the α-aminoadipate pathway?" *Journal of Molecular Evolution*, vol. 51, no. 3, pp. 299–302, 2000.

[16] M. Kellis, B. W. Birren, and E. S. Lander, "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*," *Nature*, vol. 428, no. 6983, pp. 617–624, 2004.

[17] R. T. Morris and G. Drouin, "Ectopic gene conversions in the genome of ten hemiasomycete yeast species," *International Journal of Evolutionary Biology*, vol. 2011, Article ID 970768, 11 pages, 2011.

[18] K. H. Wolfe and D. C. Shields, "Molecular evidence for an ancient duplication of the entire yeast genome," *Nature*, vol. 387, no. 6634, pp. 708–713, 1997.

[19] M. A. Hibbs, D. C. Hess, C. L. Myers, C. Huttenhower, K. Li, and O. G. Troyanskaya, "Exploring the functional landscape of gene expression: directed search of large microarray compendia," *Bioinformatics*, vol. 23, no. 20, pp. 2692–2699, 2007.

[20] I. Tirosh, N. Sigal, and N. Barkai, "Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences," *Molecular Systems Biology*, vol. 6, article 365, 2010.

[21] K. Tsui, S. Dubuis, M. Gebbia et al., "Evolution of nucleosome occupancy: conservation of global properties and divergence of gene-specific patterns," *Molecular and Cellular Biology*, vol. 31, no. 21, pp. 4348–4355, 2011.

[22] S. Chen, J. S. Brockenbrough, J. E. Dove, and J. P. Aris, "Homocitrate synthase is located in the nucleus in the yeast *Saccharomyces cerevisiae*," *Journal of Biological Chemistry*, vol. 272, no. 16, pp. 10839–10846, 1997.

[23] O. Bañuelos, J. Casqueiro, S. Steidl, S. Gutiérrez, A. Brakhage, and J. F. Martín, "Subcellular localization of the homocitrate synthase in *Penicillium chrysogenum*," *Molecular Genetics and Genomics*, vol. 266, no. 5, pp. 711–719, 2001.

[24] W. M. Jaklitsch and C. P. Kubicek, "Homocitrate synthase from *Penicillium chrysogenum*. Localization, purification of the cytosolic isoenzyme, and sensitivity to lysine," *Biochemical Journal*, vol. 269, no. 1, pp. 247–253, 1990.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 205274, 6 pages
doi:10.1155/2012/205274

Research Article

Genome Signature Difference between *Deinococcus radiodurans* and *Thermus thermophilus*

Hiromi Nishida,[1] Reina Abe,[2] Taishi Nagayama,[2] and Kentaro Yano[2]

[1] *Agricultural Bioinformatics Research Unit, Graduate School of Agricultural and Life Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan*
[2] *Bioinformatics Laboratory, Department of Life Sciences, School of Agriculture, Meiji University, 1-1-1 Higashi-Mita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan*

Correspondence should be addressed to Hiromi Nishida, hnishida@iu.a.u-tokyo.ac.jp

Received 7 November 2011; Accepted 8 December 2011

Academic Editor: Kenro Oshima

The extremely radioresistant bacteria of the genus *Deinococcus* and the extremely thermophilic bacteria of the genus *Thermus* belong to a common taxonomic group. Considering the distinct living environments of *Deinococcus* and *Thermus*, different genes would have been acquired through horizontal gene transfer after their divergence from a common ancestor. Their guanine-cytosine (GC) contents are similar; however, we hypothesized that their genomic signatures would be different. Our findings indicated that the genomes of *Deinococcus radiodurans* and *Thermus thermophilus* have different tetranucleotide frequencies. This analysis showed that the genome signature of *D. radiodurans* is most similar to that of *Pseudomonas aeruginosa*, whereas the genome signature of *T. thermophilus* is most similar to that of *Thermanaerovibrio acidaminovorans*. This difference in genome signatures may be related to the different evolutionary backgrounds of the 2 genera after their divergence from a common ancestor.

1. Introduction

In the present bacterial taxonomic system, the extremely radioresistant bacteria of the genus *Deinococcus* and the extremely thermophilic bacteria of the genus *Thermus* belong to a common lineage with remarkably different characteristics [1, 2]. Comparative genomic analyses have shown that after their divergence from a common ancestor, *Deinococcus* species seem to have acquired numerous genes from various other bacteria to survive different kinds of environmental stresses, whereas *Thermus* species have acquired genes from thermophilic archaea and bacteria to adapt to high-temperature environments [3]. For example, the aspartate kinase gene of *Deinococcus radiodurans* has a different evolutionary history from that of *Thermus thermophilus* [4]. In addition, *D. radiodurans* has several unique protein families [5] and genomic characters [6], and there is no genome-wide synteny between *D. radiodurans* and *T. thermophilus* [7]. However, phylogenetic analyses based on both orthologous protein sequence comparison and gene content comparison have shown that the genomes of *Deinococcus* and *Thermus* are most closely related with each other [3, 8]. The trinucleotide usage correlations have been used to predict the functional similarity between two RecA orthologs of bacteria including *D. radiodurans* and *T. thermophilus* [9].

If the genes acquired through horizontal gene transfers are different between *Deinococcus* and *Thermus*, then the genomic base composition (GC content) and/or genome signature can be hypothesized to also be different between these 2 genera. However, the GC content of *D. radiodurans* (67%) is similar to that of *T. thermophilus* (69.4%). The genome signature, on the other hand, is a powerful basis for comparing different bacterial genomes [11–19].

Phylogenetic analyses based on genome signature comparison have been developed, and these analyses are useful for metagenomics studies [20]. It was reported that comparative study using the frequency of tetranucleotides is a powerful tool for the bacterial genome comparison [21]. In this study, we compared the relative frequencies

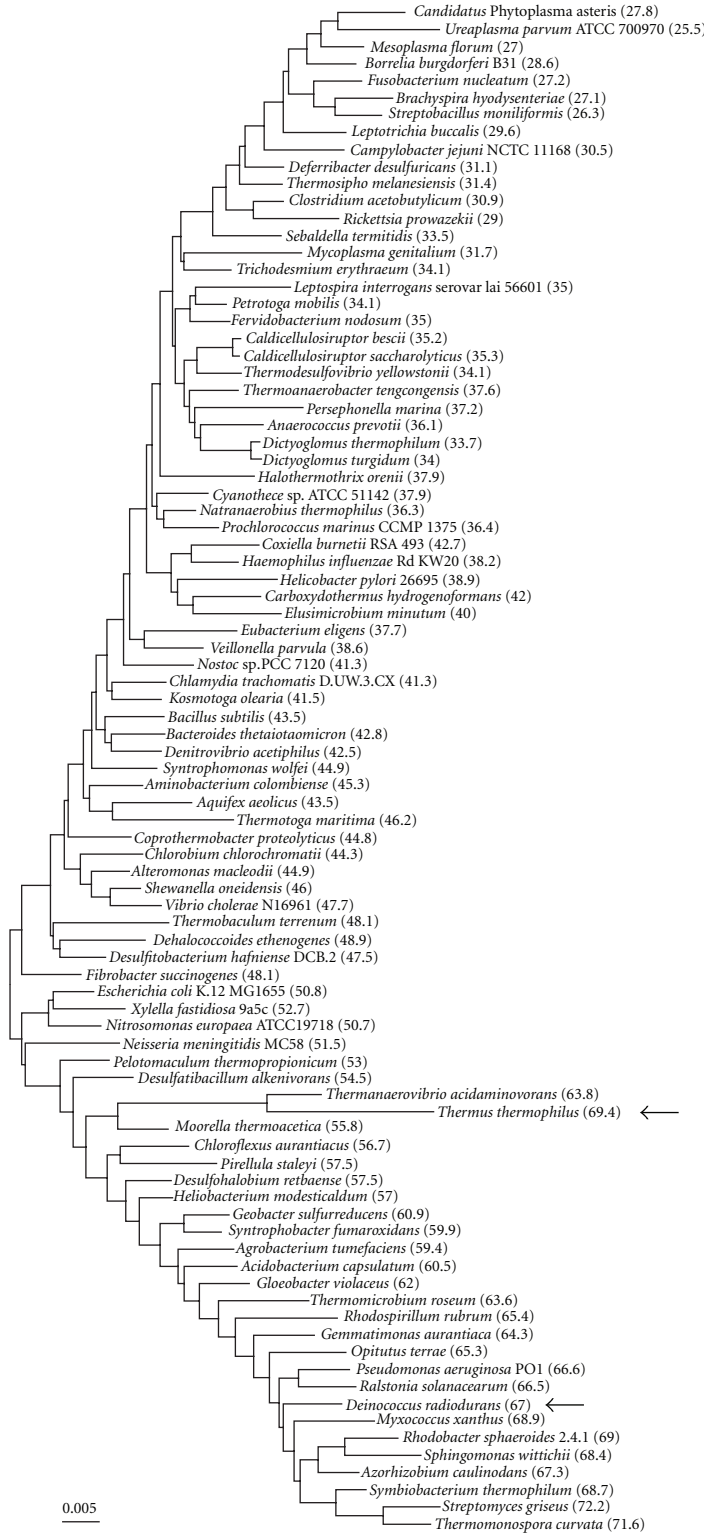


Figure 1: Neighbor-joining tree based on tetranucleotide sequence frequencies in 89 genomes. The frequencies for 89 bacteria were obtained from OligoWeb (oligonucleotide frequency search, http://insilico.ehu.es/oligoweb/). Each frequency vector consisted of 256 elements. The Euclidean distance between 2 vectors was calculated using the software package R (language and environment for statistical computing, http://www.R-project.org). On the basis of the distance matrix, a neighbor-joining tree was constructed using the MEGA software [10]. Numbers in parentheses indicate the GC content (percentage) of each genome sequence. Arrows indicate the positions of *Thermus thermophilus* and *Deinococcus radiodurans*.

Table 1: Distance between *Deinococcus radiodurans* and each bacterium using correspondence analysis.

Bacterial species	Distance
Pseudomonas aeruginosa PO1	0.297932379
Myxococcus xanthus	0.305390764
Azorhizobium caulinodans	0.308895493
Ralstonia solanacearum	0.309212661
Gloeobacter violaceus	0.317496648
Symbiobacterium thermophilum	0.324422553
Thermomonospora curvata	0.347077134
Opitutus terrae	0.376683191
Acidobacterium capsulatum	0.378916616
Gemmatimonas aurantiaca	0.383939504
Rhodobacter sphaeroides 2.4.1	0.386383492
Rhodospirillum rubrum	0.392789705
Streptomyces griseus	0.415746597
Geobacter sulfurreducens	0.425877427
Agrobacterium tumefaciens	0.457788385
Thermomicrobium roseum	0.460897799
Syntrophobacter fumaroxidans	0.470005872
Sphingomonas wittichii	0.478630032
Desulfohalobium retbaense	0.50752939
Heliobacterium modesticaldum	0.512911658
Chloroflexus aurantiacus	0.53688488
Pirellula staleyi	0.540489386
Desulfatibacillum alkenivorans	0.618176651
Pelotomaculum thermopropionicum	0.636637282
Moorella thermoacetica	0.637983756
Xylella fastidiosa 9a5c	0.655118109
Escherichia coli K-12 MG1655	0.671407958
Neisseria meningitidis MC58	0.679417806
Thermanaerovibrio acidaminovorans	0.707497366
Nitrosomonas europaea ATCC 19718	0.718956013
Fibrobacter succinogenes	0.773393097
Dehalococcoides ethenogenes	0.793600646
Vibrio cholerae N16961	0.794460696
Desulfitobacterium hafniense DCB-2	0.823845007
Thermus thermophilus	0.831109438
Shewanella oneidensis	0.84848937
Alteromonas macleodii	0.881823764
Aminobacterium colombiense	0.889238858
Thermobaculum terrenum	0.899243716
Syntrophomonas wolfei	0.90576094
Bacillus subtilis	0.913613719
Coprothermobacter proteolyticus	0.923779953
Chlorobium chlorochromatii	0.926043337
Coxiella burnetii RSA 493	0.929681834
Thermotoga maritima	0.952651677
Bacteroides thetaiotaomicron	0.958944885
Denitrovibrio acetiphilus	0.966489936
Kosmotoga olearia	0.998958025
Carboxydothermus hydrogenoformans	1.012583789
Nostoc sp. PCC 7120	1.014447775
Aquifex aeolicus	1.03027576
Chlamydia trachomatis D/UW-3/CX	1.041383827
Elusimicrobium minutum	1.06077929
Haemophilus influenzae Rd KW20	1.084974973
Veillonella parvula	1.10092918
Helicobacter pylori 26695	1.124775019
Cyanothece sp. ATCC 51142	1.126779861
Thermoanaerobacter tengcongensis	1.139238445
Halothermothrix orenii	1.149150516
Eubacterium eligens	1.164829099
Natranaerobius thermophilus	1.167863816
Prochlorococcus marinus CCMP1375	1.174664974
Fervidobacterium nodosum	1.195233916
Caldicellulosiruptor saccharolyticus	1.19880562
Caldicellulosiruptor bescii	1.199097055
Persephonella marina	1.209862481
Leptospira interrogans serovar lai 56601	1.221066506
Anaerococcus prevotii	1.224535688
Petrotoga mobilis	1.231307366
Thermodesulfovibrio yellowstonii	1.242134666
Trichodesmium erythraeum	1.246593564
Sebaldella termitidis	1.270114395
Dictyoglomus turgidum	1.29240584
Dictyoglomus thermophilum	1.297069077
Thermosipho melanesiensis	1.324630145
Deferribacter desulfuricans	1.331638037
Clostridium acetobutylicum	1.357082068
Mycoplasma genitalium	1.360597739
Campylobacter jejuni NCTC 11168	1.374681774
Leptotrichia buccalis	1.383345312
Rickettsia prowazekii	1.426681449
Borrelia burgdorferi B31	1.431569209
Candidatus Phytoplasma asteris	1.471567529
Mesoplasma florum	1.477622916
Fusobacterium nucleatum	1.487576702
Brachyspira hyodysenteriae	1.517447262
Streptobacillus moniliformis	1.535004291
Ureaplasma parvum ATCC 700970	1.559892696

of tetranucleotides in 89 bacterial genome sequences and determined the phylogenetic positions of *D. radiodurans* and *T. thermophilus*.

2. Methods

2.1. Construction of Phylogenetic Relationships Based on the Relative Frequencies of Tetranucleotides in 89 Genome Sequences. We compared the relative frequencies of tetranucleotides in the genome sequences. The frequencies of the 89 bacteria were obtained from OligoWeb (oligonucleotide frequency search, http://insilico.ehu.es/oligoweb/). The 89

Table 2: Distance between *Thermus thermophilus* and each bacterium using correspondence analysis.

Bacterial species	Distance
Thermanaerovibrio acidaminovorans	0.468763255
Symbiobacterium thermophilum	0.686400076
Geobacter sulfurreducens	0.756754453
Myxococcus xanthus	0.772836176
Streptomyces griseus	0.786527308
Thermomonospora curvata	0.791039191
Moorella thermoacetica	0.806329416
Syntrophobacter fumaroxidans	0.825184063
Deinococcus radiodurans	0.831109438
Desulfohalobium retbaense	0.835469081
Rhodospirillum rubrum	0.836862939
Azorhizobium caulinodans	0.837497899
Gloeobacter violaceus	0.847382695
Rhodobacter sphaeroides 2.4.1	0.857474011
Desulfatibacillum alkenivorans	0.876877944
Heliobacterium modesticaldum	0.886943785
Pseudomonas aeruginosa PO1	0.902403886
Pelotomaculum thermopropionicum	0.910464775
Acidobacterium capsulatum	0.940977424
Thermomicrobium roseum	0.958396462
Agrobacterium tumefaciens	0.993864461
Gemmatimonas aurantiaca	0.993867563
Ralstonia solanacearum	0.99540692
Opitutus terrae	1.014357577
Sphingomonas wittichii	1.018425039
Chloroflexus aurantiacus	1.027585883
Pirellula staleyi	1.047176443
Desulfitobacterium hafniense DCB-2	1.051272244
Dehalococcoides ethenogenes	1.071801398
Xylella fastidiosa 9a5c	1.080146527
Thermobaculum terrenum	1.103102039
Aminobacterium colombiense	1.103447745
Syntrophomonas wolfei	1.119525557
Nitrosomonas europaea ATCC 19718	1.125942985
Escherichia coli K-12 MG1655	1.136087269
Neisseria meningitidis MC58	1.137392967
Fibrobacter succinogenes	1.147727362
Aquifex aeolicus	1.154770307
Thermotoga maritima	1.163190235
Coprothermobacter proteolyticus	1.187035315
Vibrio cholerae N16961	1.194131544
Carboxydothermus hydrogenoformans	1.202997317
Shewanella oneidensis	1.207081448
Bacillus subtilis	1.236980427
Coxiella burnetii RSA 493	1.237627206
Kosmotoga olearia	1.240198963
Alteromonas macleodii	1.241401986
Bacteroides thetaiotaomicron	1.250498401
Chlamydia trachomatis D/UW-3/CX	1.259097769
Chlorobium chlorochromatii	1.264256111
Denitrovibrio acetiphilus	1.264320363
Nostoc sp. PCC 7120	1.283892849
Halothermothrix orenii	1.307140057
Thermoanaerobacter tengcongensis	1.321852789
Elusimicrobium minutum	1.327006319
Cyanothece sp. ATCC 51142	1.338924672
Helicobacter pylori 26695	1.353623157
Veillonella parvula	1.366604516
Natranaerobius thermophilus	1.374016605
Persephonella marina	1.384851067
Prochlorococcus marinus CCMP1375	1.392425502
Haemophilus influenzae Rd KW20	1.392980033
Anaerococcus prevotii	1.394012634
Eubacterium eligens	1.420199298
Dictyoglomus turgidum	1.42068199
Caldicellulosiruptor saccharolyticus	1.428805275
Caldicellulosiruptor bescii	1.430940559
Dictyoglomus thermophilum	1.432160811
Petrotoga mobilis	1.43247619
Fervidobacterium nodosum	1.436232766
Leptospira interrogans serovar lai 56601	1.445508054
Thermodesulfovibrio yellowstonii	1.445636432
Trichodesmium erythraeum	1.459523665
Sebaldella termitidis	1.491593819
Thermosipho melanesiensis	1.522817305
Deferribacter desulfuricans	1.541728701
Clostridium acetobutylicum	1.553667164
Mycoplasma genitalium	1.586376378
Campylobacter jejuni NCTC 11168	1.590027263
Leptotrichia buccalis	1.598390503
Borrelia burgdorferi B31	1.626448618
Rickettsia prowazekii	1.653875547
Candidatus Phytoplasma asteris	1.673704846
Fusobacterium nucleatum	1.674099107
Mesoplasma florum	1.701326765
Streptobacillus moniliformis	1.715886446
Brachyspira hyodysenteriae	1.717967185
Ureaplasma parvum ATCC 700970	1.784252531

bacterial species are part of a list that which covers a wide range of bacterial species published in a previous report [8]. Each frequency vector consisted of 256 ($= 4^4$) elements. The Euclidean distance between 2 vectors was calculated using the software package R (language and environment for statistical computing, http://www.R-project.org). On the basis of the distance matrix, a neighbor-joining tree was constructed using the MEGA software [10].

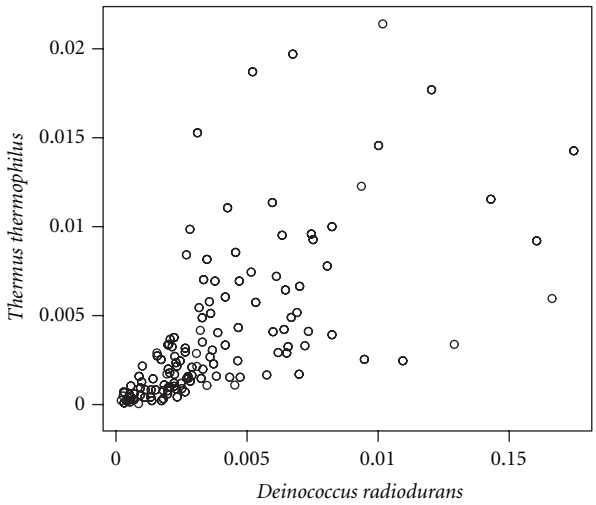


Figure 2: Scatter plot between the tetranucleotide frequencies of the genomes of *Deinococcus radiodurans* and *Thermus thermophilus*.

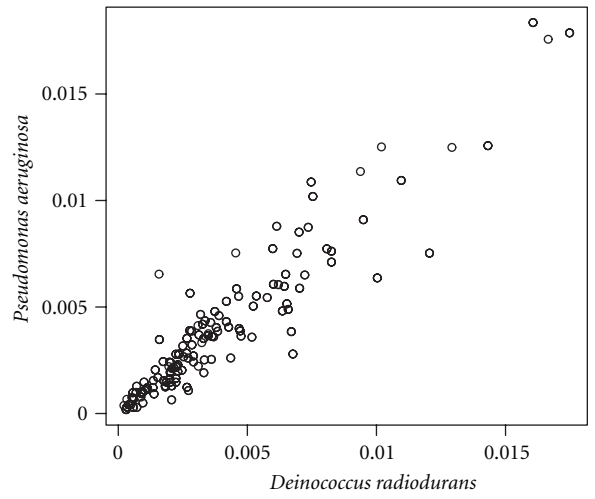


Figure 3: Scatter plot between the tetranucleotide frequencies of the genomes of *Deinococcus radiodurans* and *Pseudomonas aeruginosa*.

2.2. Ranking Based on Similarities between the Relative Frequencies of Tetranucleotides according to Correspondence Analysis. Correspondence analysis [22], which is a multivariate analysis method for profile data, was performed against the relative frequencies of tetranucleotides in 89 genomes. Correspondence analysis summarizes an originally high-dimensional data matrix (rows (tetranucleotides) and columns (genomes)) into a low-dimensional projection (space) [23, 24]. Scores (coordinates) in the low-dimensional space are given to each genome. The distance between plots (genomes) in a low-dimensional space theoretically depends on the degree of similarity in the relative frequencies of tetranucleotides: a short distance means similar relative frequencies of tetranucleotides between genomes, whereas a long distance means different relative frequencies. Thus, distance can be used as an index for similarity among genomes in the relative frequencies of tetranucleotides.

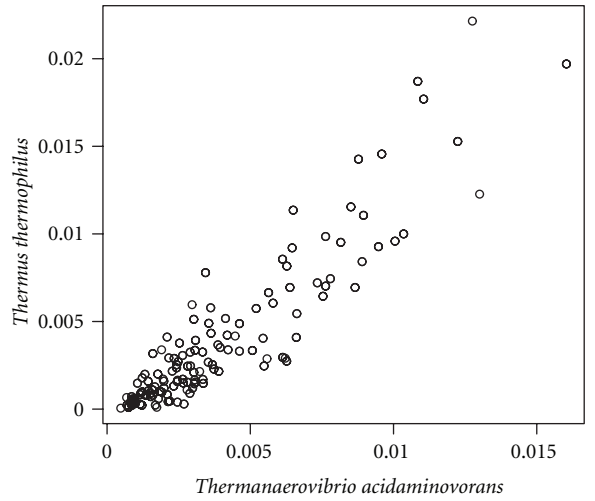


Figure 4: Scatter plot between the tetranucleotide frequencies of the genomes of *Thermanaerovibrio acidaminovorans* and *Thermus thermophilus*.

Distances between all genome pairs were calculated, and then a ranking for distances was obtained.

3. Results and Discussion

In the neighbor-joining tree (Figure 1), *D. radiodurans* is located in the high-GC-content cluster, whereas *T. thermophilus* is grouped with *Thermanaerovibrio acidaminovorans* and their group is located away from the high-GC-content cluster. The neighbor-joining tree (Figure 1) was greatly influenced by the genomic GC content bias; most of the well-defined major taxonomic groups did not form a monophyletic lineage. This result indicates that each constituent of the well-defined major group has diversified by changing its genome signature during evolution. It is consistent with a previous paper indicating that microorganisms with a similar GC content have similar genome signature patterns [25].

Phylogenetic analysis according to genome signature comparison is not based on multiple alignment data. Thus, bootstrap analysis cannot be performed. In this paper, we estimated the similarity between 2 different tetranucleotide frequencies by using correspondence analysis. The correspondence analysis showed that the genome signature of *D. radiodurans* is most similar to that of *Pseudomonas aeruginosa* (Table 1), whereas the genome signature of *T. thermophilus* is most similar to that of *Th. acidaminovorans* (Table 2). Although the *D. radiodurans* genome signature has similarity to 18 bacterial species within the distance 0.5, the *T. thermophilus* genome signature has similarity only to *Th. acidaminovorans* within the same distance (Table 2). These results indicate that *T. thermophilus* has a different genome signature from those of bacteria included in the high-GC-content cluster (Figure 1).

Although Pearson's correlation coefficient between the tetranucleotide frequencies of genomes of *D. radiodurans*

and *T. thermophilus* is 0.630 (Figure 2), that between the tetranucleotide frequencies of genomes of *D. radiodurans* and *Pseudomonas aeruginosa* is 0.935 (Figure 3) and that between the tetranucleotide frequencies of genomes of *Th. acidaminovorans* and *T. thermophilus* is 0.914 (Figure 4). These results support the results of the neighbor-joining and correspondence analyses.

The frequency of horizontal gene transfer between different bacteria may be associated with genome signature similarity. However, the tree topology based on genome signature (Figure 1) is different from that based on gene content [8]. This is caused by, among others, an amelioration of the horizontally transferred genes [26]. Our findings strongly support the previous report that *Deinococcus* has acquired genes from various other bacteria to survive different kinds of environmental stresses, whereas *Thermus* has acquired genes from thermophilic bacteria to adapt to high-temperature environments [3].

Acknowledgment

The authors thank Professor Teruhiko Beppu for his valuable comments and encouragement.

References

[1] W. G. Weisburg, S. J. Giovannoni, and C. R. Woese, "The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction," *Systematic and Applied Microbiology*, vol. 11, pp. 128–134, 1989.

[2] E. Griffiths and R. S. Gupta, "Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the *Deinococcus-Thermus* phylum," *Journal of Bacteriology*, vol. 186, no. 10, pp. 3097–3107, 2004.

[3] M. V. Omelchenko, Y. I. Wolf, E. K. Gaidamakova et al., "Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance," *BMC Evolutionary Biology*, vol. 5, article no. 57, 2005.

[4] H. Nishida and I. Narumi, "Phylogenetic and disruption analyses of aspartate kinase of *Deinococcus radiodurans*," *Bioscience, Biotechnology, and Biochemistry*, vol. 71, no. 4, pp. 1015–1020, 2007.

[5] K. S. Makarova, L. Aravind, Y. I. Wolf et al., "Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics," *Microbiology and Molecular Biology Reviews*, vol. 65, no. 1, pp. 44–79, 2001.

[6] H. Sghaier, H. Mitomo, and I. Narumi, "Genomic confession of *Deinococcus radiodurans*: it started out as a radiation resistant organism," *Viva Origino*, vol. 33, no. 4, pp. 243–257, 2005.

[7] A. Henne, H. Brüggemann, C. Raasch et al., "The genome sequence of the extreme thermophile *Thermus thermophilus*," *Nature Biotechnology*, vol. 22, no. 5, pp. 547–553, 2004.

[8] H. Nishida, T. Beppu, and K. Ueda, "Whole-genome comparison clarifies close phylogenetic relationships between phyla Dictyoglomi and Thermotogae," *Genomics*, vol. 98, no. 5, pp. 370–375, 2011.

[9] H. Sghaier, K. Satoh, and I. Narumi, "In silico method to predict functional similarity between two RecA orthologs," *Journal of Biomolecular Screening*, vol. 16, no. 4, pp. 457–459, 2011.

[10] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.

[11] A. Campbell, J. Mrázek, and S. Karlin, "Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 16, pp. 9184–9189, 1999.

[12] T. Abe, S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, and T. Ikemura, "Informatics for unveiling hidden genome signatures," *Genome Research*, vol. 13, no. 4, pp. 693–702, 2003.

[13] M. W. J. van Passel, A. Bart, A. C. M. Luyf, A. H. C. van Kampen, and A. van der Ende, "Compositional discordance between prokaryotic plasmids and host chromosomes," *BMC Genomics*, vol. 7, article 26, 2006.

[14] M. W. J. Van Passel, E. E. Kuramae, A. C. M. Luyf, A. Bart, and T. Boekhout, "The reach of the genome signature in prokaryotes," *BMC Evolutionary Biology*, vol. 6, article 84, 2006.

[15] J. Mrázek, "Phylogenetic signals in DNA composition: limitations and prospects," *Molecular Biology and Evolution*, vol. 26, no. 5, pp. 1163–1169, 2009.

[16] H. Suzuki, H. Yano, C. J. Brown, and E. M. Top, "Predicting plasmid promiscuity based on genomic signature," *Journal of Bacteriology*, vol. 192, no. 22, pp. 6045–6055, 2010.

[17] H. Suzuki, M. Sota, C. J. Brown, and E. M. Top, "Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes," *Nucleic Acids Research*, vol. 36, no. 22, article e147, 2008.

[18] S. C. Perry and R. G. Beiko, "Distinguishing microbial genome fragments based on their composition: evolutionary and comparative genomic perspectives," *Genome Biology and Evolution*, vol. 2, no. 1, pp. 117–131, 2010.

[19] F. Reyes-Prieto, A. J. García-Chéquer, H. Jaimes-Díaz et al., "LifePrint: a novel *k*-tuple distance method," *Advances and Applications in Bioinformatics and Chemistry*, vol. 2011, pp. 13–27, 2011.

[20] G. J. Dick, A. F. Andersson, B. J. Baker et al., "Community-wide analysis of microbial genome sequence signatures," *Genome Biology*, vol. 10, article R85, no. 8, 2009.

[21] D. T. Pride, R. J. Meinersmann, T. M. Wassenaar, and M. J. Blaser, "Evolutionary implications of microbial genome tetranucleotide frequency biases," *Genome Research*, vol. 13, no. 2, pp. 145–156, 2003.

[22] M. J. Greenacre, *Correspondence Analysis in Practice*, Chapman and Hall, London, UK, 2nd edition, 2007.

[23] K. Yano, K. Imai, A. Shimizu, and T. Hanashita, "A new method for gene discovery in large-scale microarray data," *Nucleic Acids Research*, vol. 34, no. 5, pp. 1532–1539, 2006.

[24] K. Hamada, K. Hongo, K. Suwabe et al., "OryzaExpress: an integrated database of gene expression networks and omics annotations in rice," *Plant and Cell Physiology*, vol. 52, no. 2, pp. 220–229, 2011.

[25] S. H. Zhang and L. Wang, "A novel common triplet profile for GC-rich prokaryotic genomes," *Genomics*, vol. 97, no. 5, pp. 330–331, 2011.

[26] J. G. Lawrence and H. Ochman, "Amelioration of bacterial genomes: rates of change and exchange," *Journal of Molecular Evolution*, vol. 44, no. 4, pp. 383–397, 1997.