Indian Journal of Bioinformatics and Biotechnology, *Vol 3(1), 1–15, January 2014*

Role of Horizontal Gene Transfer Events in the Evolution of Phenol 2–Monooxygenase Gene: a Comparative Study across 75 Prokaryotic Genomes

Jaspreet Kaur[1*], Shailly Anand[2], Mansi Verma[2] and Rup Lal[2]

[1]S.G.T.B. Khalsa College, Delhi University, India; kaurpreet21@gmail.com
[2]Molecular Biology Laboratory, Department of Zoology, University of Delhi, Delhi–110007, India; shailly.anand@gmail.com, mansiverma20@gmail.com, ruplal@gmail.com

Abstract

Horizontal Gene Transfer (HGT) is a major evolutionary force with a significant role in microbial evolution. In the present study, we report an *in silico* evidence for HGT of Phenol 2–Monooxygenase Gene which is responsible for converting Phenol to Catechol and is the initial and rate–limiting step in phenol degradation. Five computational approaches were applied across 75 completely sequenced prokaryotic genomes for studying the prospects of horizontal gene transfer of Phenol 2–Monooxygenase Gene. The phylogenetic incongruencies between the species tree (based on 16S rRNA gene) and the gene tree (using Phenol 2–Monooxygenase Gene) supported by high bootstrap values, discrepancies in %GC content, %GC3 content, codon usage analysis based on Codon Adaptation Index (CAI) and effective Number of codons (Nc) led to the conclusion that Horizontal Gene Transfer events might have taken place during the evolution of this gene. Tree and Reticulogram Reconstruction (T–REX) was constructed that detected the HGT events with significant bootstrap values, further making HGT an evident event in the evolution of this gene.

Keywords: Horizontal Gene Transfer, Phenol 2 -Monooxygenase.

1. Introduction

Horizontal Gene Transfer (HGT) is the movement of genetic material from donor to a recipient organism other than by descent. This process of natural genetic transformation is not merely transfer of genes but is in fact a multi–step process[1]. Firstly, the gene or a set of genes are transferred from one organism to another by way of Transformation, Conjugation or Transduction. The transferred gene then begins to be maintained in the recipient through replication. Finally, after sustaining the strong selective forces, the acquired gene travels through the generations as the cell divides and thus begins to ameliorate to its new lineage.

HGT was discovered nearly half a century back[2] and is believed to play a significant role in genetic plasticity in many bacterial species. Acquisition of new genes by lateral transfer rather than due to alterations in gene functions have resulted in the adaptability of organisms (especially bacteria and archaea) to new environment in the course of evolution. An evolutionary force has been suggested to enhance bacterial adaptation to environment contaminated with heavy metals[3] and toxic compounds[4]. Along with this, HGT has also allowed adaptation of pathogenic bacteria by acquiring resistance to antibiotics[5]. Comparative studies of bacterial, archaeal and eukaryotic genomes indicates that a considerable proportion of genes in prokaryotic genomes have been subjected to HGT[6]. The common example of transfer of part of Tumor–inducing (Ti) plasmid of *Agrobacterium tumefaciens* to plants[7] and to yeast[8] demonstrates the role of HGT in transferring genetic material between different

*Author for correspondence

phylogenetic groups. There is growing evidence that gene transfer events have played a significant role in evolution of prokaryotic genomes, unlike eukaryotes that evolve through modification of existing genetic information[9]. Hence, to gain a complete insight into the evolutionary processes in prokaryotes and to uncover different societal implications, HGT deserves a detailed study. The current avalanche of genome sequence information has paved way for *in silico* analysis to realize a classified overview monitoring the lateral transfer of genes across completely sequenced genomes.

In the present study, we report the HGT of the Phenol 2–Monooxygenase Gene in a wide range of prokaryotic taxa. We found that HGT plays a vital role in the evolution of the Phenol 2–Monooxygenase Gene and transfers between distant families are presented with compelling evidences.

1.1 Phenol 2–Monooxygenase Gene

Phenolic compounds constitute the second largest group of natural compounds. In the past few decades, phenolic effluents of industrial origin have further increased the number of these compounds. Due to the toxic and hazardous impact of phenols and their derivatives on the environment, these have emerged as a major group of pollutants in the industrialized nations. Apart from being potent carcinogens, their entry into the food chain, the most important component of the ecosystem, has severe effects[10]. In light of the toxicity of these compounds to higher organisms and their occurrence in nature at an alarming rate, several scientists have focused their research on the bioremediation of these compounds. Conventional methods involve various chemical and physical means of detoxification but due to the problematic secondary effluents, the alternative cost effective remediation by way of microorganisms is preferred. There are many reports in literature documenting the degradation of phenols and their derivatives in diverse groups of organisms including bacteria, filamentous fungi, yeasts and algae[11]. The analysis of gene flow among different species and genera becomes essential for the evaluation of possible consequences of the deliberate environmental release of natural or recombinant bacteria for bioremediation purposes[12]. Such analyses are now possible, thanks to the completion of majority of prokaryotic genomes.

But, not all the genes are prone to the mechanism of HGT. According to the 'Complexity Hypothesis' proposed by Jain et al.[13], genes with fewer interactions with other genes (called as Operational Genes) are more prone to gene transfer. Phenol 2–Monooxygenase or Phenol Hydroxylase (E.C 14.3.7), also considered as an operational gene, catalyses the hydroxylation of phenol to catechol. This reaction is considered to be the first and rate–limiting step in the degradation of phenol. Two different types of phenol hydroxylases have been identified in bacteria: Single–chain Flavoproteins and Multicomponent Hydroxylases, with latter to be more widely distributed among different bacteria (Figure 1).

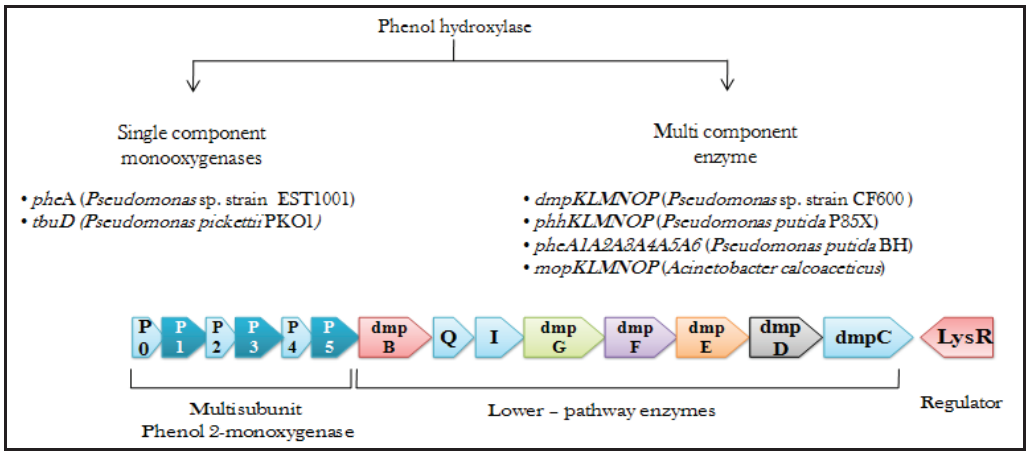

Figure 1. Schematic representation of two types of Phenol 2–Monooxygenase Gene. The multicomponent enzyme consists of a set of enzymes encoded by an operon. For example, *dmp* operon present in *Pseudomonas* sp. CF600. The *dmp* K, L M, N, O and P, and their corresponding products, P0, P1, P2, P3, P4 and P5, all code for components of Phenol 2–Monooxygenase multicomponent enzyme.

Degradation of aromatic hydrocarbons like phenol by bacteria is generally divided into an upper pathway, which produces di–hydroxylated aromatic intermediates by the action of monooxygenases (Phenol 2–Monooxygenase or Phenol Hydroxylase), followed by a lower pathway, which processes these intermediates to compounds that enter the citric acid cycle[14].

1.2 Motivation

Phenol catabolism too is a suitable model for analyzing horizontal evolution or adaptation in diverse microorganisms capable of degrading aromatic compounds. The transfer of Phenol 2–Monoxygenase Gene allows a microorganism to expand its ecological niche, allowing its proliferation in the presence of noxious compounds. But, detection and study of HGTs is not an easy task. Gene transfer experiments in natural environment are technically difficult. Though, studies using microcosm experiments have been done, they are only an approximation of natural environment. Such experiments enable manipulation of physicochemical variables (pH, temperature, carbon, nitrogen sources), which are impossible to manipulate in the natural environment. Hence, the results are viewed within the limitations of an experimental design. Moreover, frequency of transfer is shown to be lower in the presence of native microbial population[15,16]. Also, genetic transfer under laboratory conditions is done using plasmids and transposons. However, in the absence of suitable selective markers, the transformed cells may lose the newly incorporated gene by random mutation due to the absence of providing any adaptive advantage to the host cells. In other cases, the cells may die because the foreign gene may be toxic to the host cells[17].

These restrictions are overcome to a large extent by using *in–silico* analysis or predictions of gene transfer events over a wide range of a group of organisms. In addition to this, the availability of large number of complete genome sequences has given an implausible impetus to this growing field in computational biology and to study the mechanism of horizontal gene transfer in an efficient manner. Methods of HGT detection include Parametric and Phylogenetic methods[31,35]. Parametric or Compositional Methods detect atypical nucleotide composition for genes in question in comparison with the whole genome. Since base composition and codon usage vary in characteristic ways from one genome to another, the foreign or alien sequences which are not native to the host genome can be detected as deviants. Such statistical analysis can be performed by different softwares like the ones used in the present study. Additionally, phylogenetic methods search for topological conflicts between the phylogeny inferred for a gene under study (gene tree) and the corresponding organismal phylogeny (species tree). Since, conserved molecules like16S rRNA or 23S rRNA gene sequences are assumed to be refractory to HGT, comparing the phyletic groupings between the gene tree and species tree can reveal putative gene transfer events in a set of organisms under study[18]. These methods are based on the implications or consequences of HGT. They are:

Firstly, HGT will cause an unusually high degree of similarity between the donor and recipient organism for the character in question[9].

Secondly, the transferred genes can be identified by their atypical nucleotide compositions or patterns of codon usage bias. This is due to the fact that foreign sequences that are new to a recipient genome will retain their sequence characteristics and thus will be different from the recipient genome.

Thirdly, the genes introduced into a genome can also be identified by their unusual G+C and GC3 content. Otherwise, all the genes in a particular genome have fairly similar G+C content.

Fourthly, it is often seen that horizontally transferred genes are flanked by transposable elements, which further attest their foreign origin in the genome.

Lastly, phylogenetic incongruencies are observed between the 'gene tree' and the corresponding 'species tree'.

Here, we report the gene transfer events of Phenol 2–Monoxygenase Gene in prokaryotic taxa. To explore this issue, we performed an in–depth bioinformatic analysis using both parametric and phylogenetic methods. The data presented led to the conclusion that HGT has played a significant role in the evolution of this gene.

2. Materials and Methods

2.1 Sequence Retrieval and Analysis

Sequenced genomes of all the organisms were retrieved from Composition Vector Tree Version 2 (http://tlife.fudan.edu.cn/cvtree/)[19]. Gene sequences of single unit Phenol 2–Monooxygenase (EC:1.14.13.7) and catalytic subunit of multi subunit Phenol 2–Monooxygenase as well as 16S rRNA sequences from 75 prokaryotes were retrieved from KEGG (Kyoto Encyclopedia of Genes

and Genomes) (http://www.genome.ad.jp). Motif–finding algorithm, MEME 3.0 (http://meme.sdsc.edu/meme/website/)[20] was used to find conserved motifs in the Phenol 2–Monooxygenase Gene sequences. MEME discovers one or more motifs in a collection of unaligned DNA or protein sequences.

2.2 Phylogenetic Methods

75 selected sequences were aligned using the CLUSTAL X program version 1.81b[21] and phylogenetic trees for Phenol 2–Monooxygenase Gene and 16S rRNA gene sequences were constructed using the DNA Parsimony and Maximum Likelihood methods of PHYLIP package, version 3.5[22]. The resultant tree topologies were evaluated using 1000 replications by the program SEQBOOT[22] to estimate the statistical confidence for each node in phylogenetic trees. The trees were then viewed using TREEVIEW version 1.6.5[23]. Horizontal gene transfer events were detected by T–REX software[24]. Robinson and Foulds topological distance was used to construct the gene transfer tree[25]. Bootstrap validation was done for assessing the reliability of a particular gene transfer event.

2.3 GC Content

The GC Content of Phenol 2–Monooxygenase Gene for each organism was calculated using the online program Oligo Calc Software (http://www.basic.northwestern.edu/biotools/oligocalc.html)[26]. The GC Content for the whole genomes was obtained from NCBI (http://www.ncbi.nlm.nih.gov/genomes/).

2.4 Measures of Codon Bias

The software Codon W (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=codonw)[27] was used to calculate CAI, Nc and %GC3 values for Phenol 2–Monooxygenase Gene and all the ORFs of the genome sequences under study. Nc values compiled from all the ORF's of the genome (Expected values) as well as the Nc values from the single ORF–the Phenol 2–Monooxygenase (Observed value) were subjected to Chi Square Test.

3. Results and Discussion

In the present study, gene sequences of Phenol 2–Monoxygenase were retrieved from KEGG GENES database (Kyoto Encyclopedia of Genes and Genomes) from a large range of prokaryotic groups (75 organisms, 48 genera: 32 belonging to Proteobacteria, 1 to Acidobacteria, 13 to Actinobacteria, a single Firmicute and 2 to Archaea). Hence, it can be clearly seen that the gene showed an unusual distribution in certain bacterial classes and genera. This unusual species distribution of genes is a *prima facie* indicator of HGT events[29].

3.1 Phylogenetic Analysis

Discordant branching patterns were observed in 16S rRNA 'Species Tree' (Figure 2) and Phenol 2–Monooxygenase Gene tree (Figure 3), indicating the possibility of HGT. In other words, the placement of some organisms was not congruent with their taxonomic affliation, but was instead closer to an evolutionary distant organism. For example, a γ–proteobacterium (*Marinomonas* sp. MWYL1) clustered with members of α–proteobacteria rather than with other γ–proteobacterial species; the tight grouping of *Bacillus tusciae* DSM 2912 (Firmicutes) was observed with *Thermomonospora curvata* DSM 43183 (Actinobacteria) and members of β–proteobacteria & γ–proteobacteria. Further, *Azospirillum* sp. B510 (α–proteobacteria) and *Verminephrobacter eiseniae* EF01 2 (β–proteobacteria) formed a monophyletic clade. Interestingly, *Nocardia farcinica* IFM 10152 (Actinobacteria) grouped with members of both β–proteobacteria and γ–proteobacteria. These observations indicated the possibility of HGT between Gram Positive and Gram Negative organisms. The case for such anomalous placement was further strengthened when same topology was observed with more than one tree building algorithm (DNA Parsimony and Maximum Likelihood Method of PHYLIP package, version 3.63). To further study this incongruity between the species and gene trees, T–REX software[24] was used, which maps the gene tree into the species tree and then estimates the prospect of a HGT for each pair of branches of the species tree. Horizontal transfers of the considered gene are shown by arrows in the species Phylogeny (Supplementary Figure). A bootstrap validation procedure was also employed by this software to assess the reliability of a specific gene transfer. The gene transfer tree was constructed using Robinson and Foulds (RF) Topological Distance. This distance is equal to the minimum number of elementary operations, consisting of merging and splitting nodes, necessary to transform one tree to another[25–30].

3.2 Analysis based on GC Contents, Nc and Codon Usage

To further support these phylogenetic incongruencies and scrutinize HGT of Phenol 2–Monoxygenase Gene,

Figure 2. Evolutionary relationships of 16S rRNA gene sequences. The evolutionary histories were inferred using Maximum parsimony method. Phylogenetic analysis was conducted in PHYLIP package, version 3.5[22]. The numbers at node represent bootstrap values (based on 1000 resamplings). *Sulfolobus islandicus* Y.N.15.51 was taken as an outgroup.

parametric method comprising of detailed comparative analysis with parameters, namely %GC, %GC3 Content, Nc and CAI values of Phenol 2–Monooxygenase gene and the respective host genome was conducted. Unlike Phylogenetic Methods, Parametric Methods do not require alignment of homologous sequences and are therefore more suitable to examine less conserved and rapidly evolving genes[31]. In view of the fact that, base composition, codon usage and statistical features like %GC and %GC3 Content vary in significant ways from genome

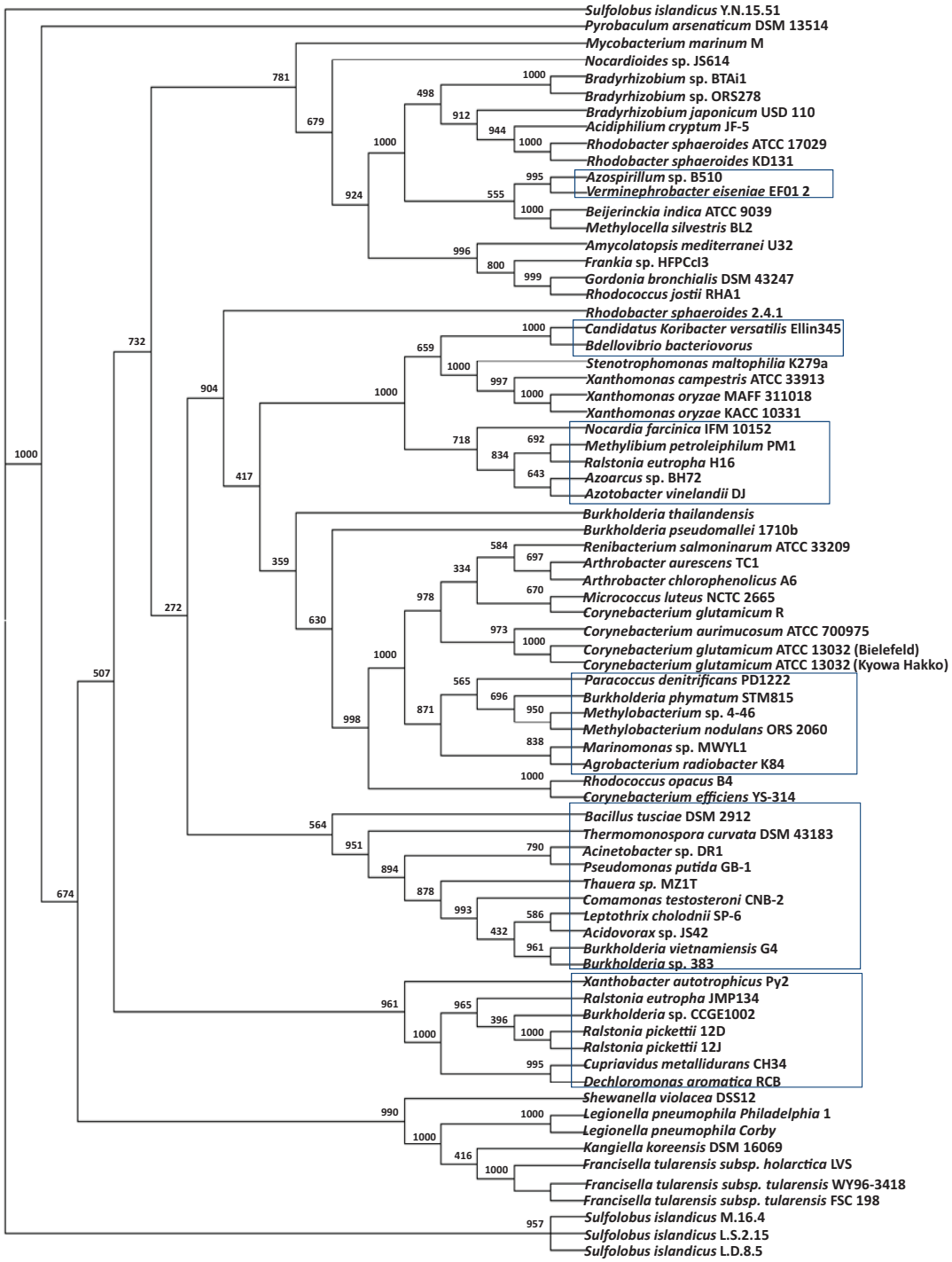

Figure 3. Evolutionary relationships of Phenol 2–Monooxygenase Gene sequences. The evolutionary histories were inferred using Maximum parsimony method. Phylogenetic analysis was conducted in PHYLIP package, version 3.5[22]. The numbers at node represent bootstrap values (based on 1000 resamplings). *Sulfolobus islandicus* Y.N.15.51 was taken as an outgroup. Names in the boxes represent the branches that are inconsistent with the 16S rRNA gene tree.

to genome, it is possible to identify foreign sequences as deviants from such genome specific characteristics. Since G+C content varies significantly as a function of the position within the codon, four discrete G+C content signatures can be identified. The first corresponds to the overall G+C content and is computed by considering all of the nucleotides in a genome. The remaining three signatures are denoted by G+C (n), with n = 1, 2, 3.

Each number corresponds to the value of the G+C content as the latter is determined by considering only those nucleotides occupying the *n*th position within each codon[32]. Nc (effective Number of codons) is an easy method to measure codon bias. It ranges from 20 (when just one codon is used per amino acid) to 61 (when each and every codon is used in equal probability)[33]. Nc is easily computed from codon usage data and its value provides an intuitively obvious measure of the extent of codon preference of a gene[34]. CAI (Codon Adaptation Index) estimates the extent of bias towards codons that are known to be favored in highly expressed genes and its value varies from 0 to 1.0[18]. It is evident from Table 1 that there is a good deal of deviation in %GC3 and Nc values between the Phenol 2–Monooxygenase Gene and the respective genomes of the studied strains, further indicating HGT of this gene. The difference in the GC content of Phenol 2–Monooxygenase Gene and whole genome is evident for α–proteobacteria (*Bradyrhizobium japonicum* USD 110, *Methylocella silvestris* BL2 & *Methylobacterium* sp. 4–46), β–proteobacteria (*Verminephrobacter eiseniae* EF01 2 & *Burkholderia thailandensis*) and γ–proteobacteria (*Acinetobacter* sp. DR1 & *Kangiella koreensis* DSM 16069). The GC content of newly acquisited gene is found to differ from the GC content of whole genome. This provides an important conclusion that the gene is transferred horizontally and has been acquired recently. In order to determine whether these observed deviations are true statistical deviations, we used the following equation: $(Nc_{Observed} - Nc_{Expected})^2/Nc_{Expected}$) for calculating Chi–Square statistical values. There were 8 cases of strong inconsistencies in the phylogenetic tree based on high bootstrap value (≥500) with organisms distantly related by 16S rRNA sequence and were further supported by significant chi square values and deviations in GC contents. These events might thus have arisen due to HGT (Table 2).

Demonstration of conserved sequence motifs would further provide a supporting evidence for an evolutionary relationship between different prokaryotic Phenol 2–Monooxygenases. Hence, in the present study, MEME 3.0 was used to analyze a divergent set of Phenol 2–Monooxygenases. For example, three strongly conserved motifs were identified between Phenol 2–Monoxygenase sequence of *Azospirillum* sp. B510 (α–proteobacteria) and *Verminephrobacter eiseniae* EF01 2 (β–proteobacteria) with significant sequence similarities between these organisms (data not shown). It was also observed that bacterial and archaeal Phenol 2–Monooxygenase Genes do not share homology as members of both domains form separate clades and are distinct from each other. Interestingly, *Candidatus Koribacter versatilis* Ellin 345 (Acidobacteria) and *Bdellovibrio bacteriovorus* (Deltaproteobacteria) form a monophyletic clade and supported by extremely high bootstrap value, but there is no significant overall sequence similarity between these organisms. This may be because the sequences have diverged so far that the evolutionary relationships are no longer readily evident. Further, these proteins may have arrived at a common function by convergent evolution from different progenitors.

4. Ecological Context of HGT

The close associations of bacteria belonging to different ecological niches make HGT possible. For instance, both *Burkholderia phymatum* STM815 and *Agrobacterium radiobacter* K84 are mesophillic and such a common ecological niche provides a propagation ground for HGT. Similarly, *Bacillus tusciae* DSM 2912 and *Thermomonospora curvata* DSM 43183 are thermophilic bacteria. *Azotobacter vinelandii* DJ is a soil bacterium that fixes nitrogen under aerobic conditions[36] and *Azoarcus* sp. strain BH72 is also of agrobiotechnological interest because it supplies biologically fixed nitrogen to its host[37]. *Xanthobacter autotrophicus* strain Py2 has multiple habitats, and hence HGT is not limited by geographical barriers.

5. Conclusion

Keeping in view the impact of horizontal gene transfers on the ecological and pathogenic character of genomes, algorithms were sought after, that can computationally determine which genes in a given genome are products of HGT events. Though, these algorithms or tools have been quite successful in their aim, but still many intricate and exciting challenges remain to be overcome before a clearer global picture of HGT patterns can come forward. Apart from the computationally challenging problems that arise from large datasets and quantifying disagreements among trees for detecting HGT, one of the major challenges that this approach faces include determining whether the disagreements are indeed due to HGT. For example, the compositional methods are usually applicable to recent transfers, because the older the transfer is, the more the gene adapts to the new genome. Moreover,

Vol 3(1) | January 2014 | www.jibb.informaticspublishing.com

Table 1. Characteristics (CAI, %GC, Nc and %GC3) of genes and genomes of different organisms involved in degradation of phenol by Phenol 2–Monooxygenase Gene

S. No.	Name	Length (bp)	Class	%GC whole Genome	%GC of gene	CAI of gene	Av. CAI of genome	Nc (Gene)	Nc (Genome)	%GC3 (Gene)	%GC3 (Genome)
1	*Bradyrhizobium* sp. ORS278	1092	α–proteobacteria	65.5	63	0.326	0.297	35.5	36.16	86.5	84.4
2	*Bradyrhizobium* sp. BTAi1	1095	α–proteobacteria	64.8	63	0.327	0.289	35.92	37.56	85.3	82.2
3	*Azospirillum* sp. B510	1140	α–proteobacteria	67.6	66	0.335	0.302	32.55	34.16	86.8	86.9
4	*Beijerinckia indica* ATCC 9039	1086	α–proteobacteria	57	57	0.296	0.245	37.89	46.55	71.3	64.9
5	*Bradyrhizobium japonicum* USD 110	1179	α–proteobacteria	64.1	60	0.286	0.280	43.25	39.44	77.3	80.5
6	*Acidiphilium cryptum* JF–5	1056	α–proteobacteria	67.1	64	0.318	0.274	32.32	35.94	86.1	84.9
7	*Rhodobacter sphaeroides* ATCC 17029	1059	α–proteobacteria	69	66	0.312	0.259	33.24	34.34	90.4	88.2
8	*Rhodobacter sphaeroides* 2.4.1	864	α–proteobacteria	68.8	69	0.348	0.258	31.31	34.58	91.3	87.8
9	*Methylocella silvestris* BL2	1074	α–proteobacteria	63.1	58	0.258	0.256	43.03	41.06	69.8	76.2
10	*Xanthobacter autotrophicus* Py2	1026	α–proteobacteria	67.3	69	0.313	0.289	29.94	35.34	93.3	85.7
11	*Agrobacterium radiobacter* K84	1947	α–proteobacteria	59.9	61	0.261	0.284	45.85	42.11	71.9	73.1
12	*Methylobacterium* sp. 4–46	1905	α–proteobacteria	71.5	66	0.279	0.254	37.12	32.76	83.4	92.2
13	*Methylobacterium nodulans* ORS 2060	1905	α–proteobacteria	68.4	69	0.309	0.257	33.43	35.72	88.9	87
14	*Paracoccus denitrificans* PD1222	1908	α–proteobacteria	66.8	68	0.288	0.288	33.25	36.16	83.8	83.3
15	*Rhodobacter sphaeroides* KD131	1038	α–proteobacteria	69.1	66	0.310	0.256	32.51	34.62	90.5	87.6
16	*Azoarcus* sp. BH72	1092	β–proteobacteria	67.9	67	0.374	0.340	30.57	33.54	91.8	80.5
17	*Verminephrobacter eiseniae* EF01 2	1086	β–proteobacteria	65.2	59	0.303	0.301	42.71	38.27	73.4	80.5

Indian Journal of Bioinformatics and Biotechnology | Print ISSN 2319-6580 | Online ISSN 2319-6599

18	*Burkholderia* sp. CCGE1002	987	β–proteobacteria	63.3	64	0.287	0.305	41.54	38.7	77.5	79
19	*Ralstonia pickettii* 12D	990	β–proteobacteria	63.3	67	0.314	0.324	34.22	38.19	84.2	79
20	*Ralstonia pickettii* 12J	990	β–proteobacteria	63.6	67	0.314	0.322	34.22	38.55	84.2	78.7
21	*Cupriavidus metallidurans* CH34	987	β–proteobacteria	63.5	56	0.386	0.346	48.76	33.3	67	86.5
22	*Dechloromonas aromatica* RCB	987	β–proteobacteria	59.2	63	0.390	0.311	31.42	42.84	83.4	70.2
23	*Ralstonia eutropha* JMP134	999	β–proteobacteria	64.4	68	0.351	0.325	33.97	37.29	87.8	80.7
24	*Ralstonia eutropha* H16	996	β–proteobacteria	66.3	68	0.424	0.341	29.32	34.58	91.9	84.7
25	*Burkholderia* sp. 383	996	β–proteobacteria	66.3	68	0.334	0.312	30.55	35	90.6	84.2
26	*Leptothrix cholodnii* SP–6	996	β–proteobacteria	68.9	69	0.406	0.344	27.83	31.51	91.9	89.6
27	*Thauera* sp. MZ1T	984	β–proteobacteria	68.3	67	0.363	0.320	29.18	32.81	93.9	88.8
28	*Burkholderia vietnamiensis* G4	996	β–proteobacteria	65.7	56	0.294	0.309	52.47	34.43	60.8	83.9
29	*Acidovorax* sp. JS42	996	β–proteobacteria	66.1	69	0.437	0.327	26.37	35.17	94.8	84
30	*Comamonas testosteroni* CNB–2	993	β–proteobacteria	61.5	65	0.347	0.317	33.97	39.57	80.4	75
31	*Burkholderia thailandensis*	1197	β–proteobacteria	67.6	73	0.287	0.289	32.45	34.48	91.2	85.5
32	*Burkholderia phymatum* STM815	1896	β–proteobacteria	62.3	63	0.284	0.297	42.68	39.38	71.5	77
33	*Burkholderia pseudomallei* 1710b	1008	β–proteobacteria	68	65	0.249	0.281	48.49	35.14	68.1	84.2
34	*Methylibium petroleiphilum* PM1	1098	β–proteobacteria	68.8	66	0.410	0.327	28.05	32.49	95	89.3
35	*Pseudomonas putida* GB–1	996	γ–proteobacteria	61.9	65	0.368	0.353	32.61	38.12	83.3	77.2
36	*Azotobacter vinelandii* DJ	999	γ–proteobacteria	65.7	66	0.372	0.316	33.52	35.68	87.7	82.5
37	*Acinetobacter* sp. DR1	1002	γ–proteobacteria	38	43	0.292	0.250	48.79	45.2	33.1	26.9

(*Continued*)

Vol 3(1) | January 2014 | www.ijbbinformaticspublishing.com

Table 1. *Continued*

S. No.	Name	Length (bp)	Class	%GC whole Genome	%GC of gene	CAI of gene	Av. CAI of genome	Nc (Gene)	Nc (Genome)	%GC3 (Gene)	%GC3 (Genome)
38	*Shewanella violacea* DSS12	1038	γ–proteobacteria	44	43.2	0.192	0.230	54.68	52.27	41.7	42.5
39	*Stenotrophomonas maltophilia* K279a	750	γ–proteobacteria	66.3	67	0.451	0.374	31.03	34.1	86.5	82.8
40	*Kangiella koreensis* DSM 16069	732	γ–proteobacteria	43.7	41	0.265	0.264	54.98	51.17	36.8	39.4
41	*Francisella tularensis subsp. tularensis* FSC 198	732	γ–proteobacteria	32.3	33	0.213	0.197	40.32	41.07	16.2	19.1
42	*Legionella pneumophila Corby*	693	γ–proteobacteria	38.5	37	0.190	0.203	43.43	50.26	24.4	30.2
43	*Legionella pneumophila Philadelphia* 1	747	γ–proteobacteria	38.3	37	0.184	0.202	46.04	50.6	26	30.5
44	*Xanthomonas campestris* ATCC 33913	768	γ–proteobacteria	65.1	68	0.362	0.346	33.4	36.48	84.1	80.6
45	*Xanthomonas oryzae* KACC 10331	762	γ–proteobacteria	63.7	67	0.369	0.323	34.79	39.04	86.4	77.1
46	*Xanthomonas oryzae* MAFF 311018	723	γ–proteobacteria	63.7	67	0.396	0.326	33.55	39.03	86.9	77.1
47	*Francisella tularensis* WY96–3418	732	γ–proteobacteria	32.3	33	0.213	0.198	40.38	41.04	16.2	19.1
48	*Francisella tularensis subsp. holarctica* LVS	732	γ–proteobacteria	32.2	33	0.215	0.196	40.19	41	16.2	18.9
49	*Marinomonas* sp. MWYL1	1914	γ–proteobacteria	42.6	45	0.273	0.248	53.99	49.36	36.9	35.8
50	*Bdellovibrio bacteriovorus*	723	δ–proteobacteria	50.6	48	0.374	0.336	47.52	47.33	57.1	56
51	*Bacillus tusciae* DSM 2912	1113	Firmicutes	59	56	0.255	0.205	52.96	47.45	63.8	68.5
52	*Amycolatopsis mediterranei* U32	1170	Actinobacteria	71.3	68	0.384	0.282	28.86	31.18	95.7	92.8
53	*Frankia* sp. HFPCcI3	1104	Actinobacteria	70.1	65	0.304	0.242	35.8	36.73	87.9	85.3

Indian Journal of Bioinformatics and Biotechnology | Print ISSN 2319-6580 | Online ISSN 2319-6599

54	*Gordonia bronchialis* DSM 43247	1107	Actinobacteria	67	64	0.317	0.270	34.87	37.34	85.1	83.5
55	*Rhodococcus jostii* RHA1	1107	Actinobacteria	67	65	0.348	0.258	32.78	37.32	92.8	84.2
56	*Nocardioides* sp. JS614	1050	Actinobacteria	71.4	62	0.275	0.265	42.01	31.59	82.5	93.4
57	*Mycobacterium marinum* M	1092	Actinobacteria	65.7	67	0.295	0.267	36.78	41.26	85.1	79.2
58	*Thermomonospora curvata* DSM 43183	1029	Actinobacteria	71.6	73	0.316	0.278	27.54	30.99	95.2	92.3
59	*Corynebacterium glutamicum* ATCC 13032 (Kyowa Hakko)	957	Actinobacteria	53.8	56	0.331	0.305	43.63	46.89	63.4	57
60	*Corynebacterium glutamicum* ATCC 13032 (Bielefeld)	957	Actinobacteria	53.8	56	0.164	0.303	57.09	46.96	59.6	56.9
61	*Corynebacterium glutamicum* R	1884	Actinobacteria	54.1	56	0.344	0.304	43.23	46.71	63.1	57.9
62	*Corynebacterium efficiens* YS–314	2064	Actinobacteria	63	63	0.317	0.299	37.61	38.78	78	77.2
63	*Corynebacterium aurimucosum* ATCC 700975	1851	Actinobacteria	60.6	60	0.374	0.315	35.46	41.91	75.5	72.3
64	*Nocardia farcinica* IFM 10152	1029	Actinobacteria	70.7	70	0.319	0.287	33.36	32.91	88.6	90.2
65	*Rhodococcus opacus* B4	1617	Actinobacteria	67.6	68	0.288	0.261	31.69	36.18	90.1	85.8
66	*Arthrobacter aurescens* TC1	1935	Actinobacteria	62.4	64	0.346	0.280	37.5	43.51	78.5	74.1
67	*Arthrobacter chlorophenolicus* A6	1899	Actinobacteria	66	68	0.329	0.295	32.14	36.98	89.7	82.3
68	*Renibacterium salmoninarum* ATCC 33209	987	Actinobacteria	56.3	58	0.307	0.253	50.14	52.26	65.2	59
69	*Micrococcus luteus* NCTC 2665	1914	Actinobacteria	73	70	0.398	0.289	25.9	29.97	98.7	94.7

(*Continued*)

Table 1. *Continued*

S. No.	Name	Length (bp)	Class	%GC whole Genome	%GC of gene	CAI of gene	Av. CAI of genome	Nc (Gene)	Nc (Genome)	%GC3 (Gene)	%GC3 (Genome)
70	*Candidatus Koribacter versatilis* Ellin345	738	Acidobacteria	58.4	57	0.283	0.281	41.31	44.64	73	70.4
71	*Pyrobaculum arsenaticum* DSM 13514	1137	Archaea	55.1	51	0.175	0.174	52.49	48.4	62.7	65.8
72	*Sulfolobus islandicus* L.S.2.15	1143	Archaea	35.1	37	0.157	0.152	43.01	44.6	31.9	28.3
73	*Sulfolobus islandicus* L.D.8.5	1131	Archaea	35.3	37	0.155	0.152	42.49	44.5	31.5	29
74	*Sulfolobus islandicus* M.16.4	1131	Archaea	35	37	0.155	0.152	42.38	44.53	32.1	28.4
75	*Sulfolobus islandicus* Y.N.15.51	993	Archaea	35.3	37	0.146	0.152	42.27	44.41	31.8	28.6

Table 2. Different associations of organisms indicating horizontal gene transfer of Phenol 2–Monooxygenase Gene

Name of Organism	Chi Square Value	Bootstrap Value	Associated Organisms
Verminephrobacter eiseniae EF01 2 (β–proteobacteria)	8834.70	995	*Azospirillum* sp. B510 (α–proteobacteria)
Bdellovibrio bacteriovorus (δ–proteobacteria)	2195.85	1000	*Candidatus Koribacter versatilis* Ellin345 (Acidobacteria)
Nocardia farcinica IFM 10152 (Actinobacteria)	2928.52	718	*Methylibium petroleiphilum* PM1, *Ralstonia eutropha* H16 (β–proteobacteria)
Azoarcus sp. BH72 (β–proteobacteria)	3354.08	643	*Azotobacter vinelandii* DJ (γ–proteobacteria)
Paracoccus denitrificans PD1222 (α–proteobacteria)	4362.43	565	*Burkholderia phymatum* STM815 (β–proteobacteria)
Marinomonas sp. MWYL1 (γ–proteobacteria)	4991.31	838	*Agrobacterium radiobacter* K84 (α–proteobacteria)
Bacillus tusciae DSM 2912 (Firmicutes)	4451.85	564	*Thermomonospora curvata* DSM 43183 (Actinobacteria)
Xanthobacter autotrophicus Py2 (α–proteobacteria)	6545.54	961	β–proteobacterial group

one drawback of the phylogenetic incongruence approach is that the horizontal transfers between taxa in the reference tree cannot be detected. Also, the results obtained are dependent on the choice of the reference tree. Thus, with voluminous data to interpret, the tree–based approaches are slow and tend to scale poorly to genome–wide applications. In addition, inference of correct phylogenetic trees is a difficult problem, and inferred gene trees can be incorrect, particularly when lineages evolve at different rates[38].

Despite the limitations of the current inference methods, the overall aim of this work was to provide additional insight into the areas of computational biology involving phylogenetics and HGT. More specifically, it benchmarks the phylogenetic reconstruction methods, using existing methods of HGT detection, and putting together a framework to help users with phylogenetic detection methods. Specifically, we were interested in how the sequence length, and total number of sequences affect the reconstruction methods. To answer this question, we selected a large data set comprising of 75 prokaryotic genomes. Given how imperfect HGT detection methods are, we obtained an approximate picture at best. Additionally, we have used genome data to estimate atypical statistical parameters of genomes, rather than providing a single phylogenetic tree only, which is an important step towards the improvement in the reconstruction of the evolutionary history of prokaryotes.

Thus, it is evident from this study that horizontal gene transfers are difficult to prove and various supporting lines of evidence are usually necessary for a convincing case. In order to study gene transfer phenomenon, sequences from a protein or a gene should be available from numerous and evolutionary distant organisms[28]. This is feasible now–a–days, thanks to the large number of genome sequences deposited in the Gene bank databases. From the present study, it was clearly observed that horizontal gene transfers should be considered carefully and there should be sufficient number of evidences, strongly in favour of a 'True HGT Event'. Because, as in every prediction, there will be false positives (genes that appear to be transferred but are not). Hence, ruling out these false positives and identifying evolution of a gene by HGT involves detailed study using a combined approach of various detection methods.

In conclusion, it is clear from the analysis of Phylogenetic Methods (incongruencies between Phenol 2–Monoxygenase Gene tree and 16S rRNA species tree) and Parametric Methods (statistically significant deviations in parameters such as %G+C and %GC3 Content, Nc, CAI and Chi–Square Tests) that HGT may be a major contributor for the evolution of Phenol 2–Monoxygenase Gene. Finally, our results are consistent with earlier claims of the important role of HGT in the evolution of microorganisms.

6. Acknowledgement

This work was supported by grants from National Bureau of Agricultural Important Microorganisms (NBAIM), Government of India. JK, SA and MV acknowledge NBAIM and CSIR (Council of Scientific and Industrial Research), Government of India for providing the research fellowships.

7. References

1. Eisen J. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Curr Opin Genet Dev. 2000; 10(6):606–611.
2. Syvanen M. Horizontal gene transfer: evidence and possible consequences. Annu Rev Genet. 1994; 28:237–261. doi:10.1146/annurev.ge.28.120194.001321.
3. Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome. Microbiol Mol Biol Rev. 1999 Sep; 63(3):507–522.
4. Top EM, Springael D. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Curr Opin Biotechnol. 2003 Jun; 14(3):262–269.
5. Walsh TR. Combinatorial genetic evolution of multiresistance. Curr Opin in Microbiol. 2006 Oct; 9(5):476–482.
6. Koonin EV, Makarova KS, AravindL. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol. 2001; 55:709–742. doi:10.1146/annurev.micro.55.1.709.
7. Zupan JR, Zambryski P. Transfer of T–DNA from *Agrobacterium* to the plant cell. Plant Physiol. 1995 Apr; 107(4):1041–1047.
8. Bundock P, den–Dulk–Ras A, Beijersbergen A, Hooykaas PJ. Transkingdom T–DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisae*. EMBO J. 1995 Jul; 14(13):3206–3214.
9. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000 May; 405(6784):299–304.
10. Enroth C, Neujahr H, Schneider G, Lindqvist Y. The crystal structure of Phenol 2–Monooxygenase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. Structure. 1998 May 15; 6(5):605–617.
11. Dos Santos VL, Monteiro Ade S, Braga DT, Santoro MM. Phenol degradation by *Aureobasidium pullulans* FE13 isolated from industrial effluents. J Hazard Mater. 2009 Jan 30; 161(2–3):1413–1420.
12. Davison J. Genetic exchange between bacteria in the environment. Plasmid. 1999 Sep; 42(2):73–91.
13. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci, USA. 1999 Mar 30; 96(7):3801–3806.
14. Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, Pucci P, Donato AD. Phenol hydroxylase and toluene/*o*–xylene monooxygenase from *Pseudomonas stutzeri* OX1: interplay between two enzymes. Appl Environ Microbiol. 2004 Apr; 70(4):2211–2219.
15. Stewart GJ, Sinigalliano CD. Exchange of chromosomal markers by natural transformation between the soil isolate *Pseudomonas stutzeri* JM 300 and the marine isolate *Pseudomonas stutzeri* strain ZoBell. Antonie van Leeuwenhoek. 1991 Jan; 59(1):19–25.
16. Top E, Mergeay M, Springael D, Verstraete W. Gene escape model: transfer of heavy metal resistance genes from *Escherichia coli* to *Alcaligenes eutrophus* on agar plates and in soil samples. Appl Environ Microbiol. 1990 Aug; 56(8):2471–2479.
17. Kurland CG, Canback B, Berg OG. Horizontal gene transfer: a critical view. Proc Natl Acad Sci, USA. 2003 Aug 19; 100(17):9658–9662.
18. Makarenkov V, Boc A, Delwiche CF, Philippe H. A new efficient method for detecting horizontal gene transfers: modeling partial and complete gene transfer scenarios. Université du Québec à Montréal. 2003; Départementd'informatique, Université du Québec à Montréal, C.P. 8888, Succ. Centre-Ville, Montréal (Québec), Canada, H3C 3P8.
19. Xu Z, Hao B. CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. Nucleic Acids Res. 2009 Jul 1; 37(Web Server issue):W174–W178.
20. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994; 2:28–36 .
21. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997 Dec 15; 25(24): 4876–4882.
22. Felsenstein J. PHYLIP (PHYLogeny Inference Package), version 3.5c. Department of Genome Sciences, University of Washington, Seattle, USA; 1993.
23. Page RDM. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci. 1996 Aug; 12(4):357–358.
24. Makarenkov V. T–Rex: reconstructing and visualizing phylogenetic trees and reticulation networks. Bioinformatics. 2001 Jul; 17(7):664–668.
25. Robinson DR, Foulds LR. Comparison of phylogenetic trees. Math Biosci. 1981; 53:131–147 .

26. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. 2007 Jul; 35(Web Server Issue):W43–W46. doi:10.1093/nar/gkm234.
27. Peden J. Analysis of codon usage [PhD Thesis]. United Kingdom: The University of Nottingham; 1999.
28. Smith MW, Feng D–F, Doolittle RF. Evolution by acquisition: the case for horizontal gene transfers. Trends Biochem Sci. 1992 Dec 1; 17(12):489–493.
29. Brown JR. Ancient horizontal gene transfer. Nat Rev Genet. 2003 Feb; 4(2):121–132.
30. Makarenkov V, Leclerc B. An optimal way to compare additive trees using circular orders. J Comput Biol. 2000; 7(5):731–744.
31. Zaneveld JR. Nemergut DR, Knight R. Are all horizontal gene transfers created equal? prospects for mechanism–based studies of HGT patterns. Microbiology. 2008 Jan; 154(Pt 1):1–15.
32. Tsirigos A, Rigoutsos I. A new computational method for the detection of horizontal gene transfer events. Nucleic Acids Res. 2005 Feb 16; 33(3):922–933.
33. Mondal UK, Sur S, Bothra AK, Sen A. Comparative analysis of codon usage patterns and identification of predicted highly expressed genes in five *Salmonella* genomes. Indian J of Med Microbiol. 2008 Oct–Dec; 26(4):313–321.
34. Wright F. The effective number of codons used in a gene. Gene. 1990 Mar 1; 87(1):23–29.
35. Azad RK, Lawrence JG. Towards more robust methods of alien gene detection. Nucleic Acids Res. 2011 May; 39(9) 1–11. doi:10.1093/nar/gkr059.
36. Setubal JC, dos Santos P, Goldman BS, Ertesvåg H, Espin G, Rubio LM, Valla S, Almeida NF, Balasubramanian D, Cromes L, Curatti L, Du Z, Godsy E, Goodner B, Hellner–Burris K, Hernandez JA, Houmiel K, Imperial J, Kennedy C, Larson TJ, Latreille P, Ligon LS, Lu J, Maerk M, Miller NM, Norton S, O'Carroll IP, Paulsen I, Raulfs EC, Roemer R, Rosser J, Segura D, Slater S, Stricklin SL, Studholme DJ, Sun J, Viana CJ, Wallin E, Wang B, Wheeler C, Zhu H, Dean DR, Dixon R, Wood D: Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. J Bacteriol. 2009 Jul; 191(14): 4534–4545.
37. Krause A, Ramakumar A, Bartels D, Battistoni F, Bekel T, Boch J, Böhm M, Friedrich F, Hurek T, Krause L, Linke B, McHardy AC, Sarkar A, Schneiker S, Syed A, Thauer R, Vorhölter FJ, Weidner S, Pühler A, Reinhold–Hurek B, Kaiser O, Goesmann A. Complete genome of the mutualistic, N2–fixing grass endophyte *Azoarcus* sp. strain BH72. Nat Biotechnol. 2006 Nov; 24(11): 1385–1391.
38. Anderson FE, Swofford DL. Should we be worried about long branch attraction in real data sets? investigations using metazoan18S rDNA. Mol Phylogenet Evol. 2004 Nov; 33(2):440–451.