Genome-Guided Insights Reveal Organophosphate-Degrading *Brevundimonas diminuta* as *Sphingopyxis wildii* and Define Its Versatile Metabolic Capabilities and Environmental Adaptations

Sunil Parthasarathy[1,†], Sarwar Azam[2,†], Annapoorni Lakshman Sagar[1], Veera Narasimha Rao[2], Ramurthy Gudla[1], Hari Parapatla[1], Harshita Yakkala[1], Sujana Ghanta Vemuri[1], and Dayananda Siddavattam[1,*]

[1]Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad, India

[2]National Institute of Animal Biotechnology, Hyderabad, India

†These authors contributed equally to this work.

*Corresponding authors: E-mails: sdsl@uohyd.ernet.in; siddavattam@gmail.com.

Abstract

The complete genome sequence of *Brevundimonas diminuta* represented a chromosome (~4.15 Mb) and two plasmids (pCMS1 and pCMS2) with sizes of 65,908 and 30,654 bp, respectively. The sequence of the genome showed no significant similarity with the known bacterial genome sequences, instead showed weak similarity with the members of different genera of family, Sphingomonadaceae. Contradicting existing taxonomic position, the core genome-guided phylogenetic tree placed *B. diminuta* in the genus Sphingopyxis and showed sufficient genome-to-genome distance warranting a new species name. Reflecting the strains ability to grow in harsh environments, the genome-contained genetic repertoire required for mineralization of several recalcitrant man-made aromatic compounds.

Key words: biodegradation, aromatic compound degradation, biotransformation, Sphingomonadales.

Introduction

The safe disposal of neurotoxic organophosphate (OP) residues has attracted the attention of several microbiologists. Bacterial strains possessing organophosphate hydrolase (OPH) activity have been isolated from sewage and soil samples (Munnecke and Hsieh 1974). Using conventional taxonomic tools, the isolated OP-degrading bacterial strains have been placed in the genus, *Pseudomonas* (Munnecke and Hsieh 1974). However, when the genus *Pseudomonas* was reclassified, bacterial strains that were previously named as *Pseudomonas diminuta* and *Pseudomonas vesicularis* were moved to a separate genus known as *Brevundimonas*. Accordingly, the *P. diminuta* MG was renamed *Brevundimonas diminuta* MG (Segers et al. 1994). In this study, we generated the complete genome sequence of *B. diminuta* MG and report several interesting features pertaining to its taxonomy, evolution, and degradation potential.

Results and Discussion

A total of 6.8 Gbp data was generated to assemble the complete genome of *B. diminuta*. The scaffold level assembly contained 28 scaffolds with N50 of 1,786,567 bp. These scaffolds were further merged to obtain three super scaffolds using data generated from a mate-pair library. The super scaffolds were circularized by manually generating sequence for the DNA, amplified using primers specific to the right and left flanks of the assembled scaffold sequences. The largest super scaffold with a length of 4,147,822 bp is regarded as the chromosome sequence of *B. diminuta* (fig. 1A). The other two scaffolds gave circular DNA molecules with a size of 65,908 bp (fig. 1B) and 30,654 bp (fig. 1C), respectively. The 65,908 bp circular sequence matched perfectly with the physical map and partial sequence of pCMS1, a previously reported indigenous plasmid of *B. diminuta* (Pandeeti et al. 2011). However, the second circular DNA showed no complete similarity to any other
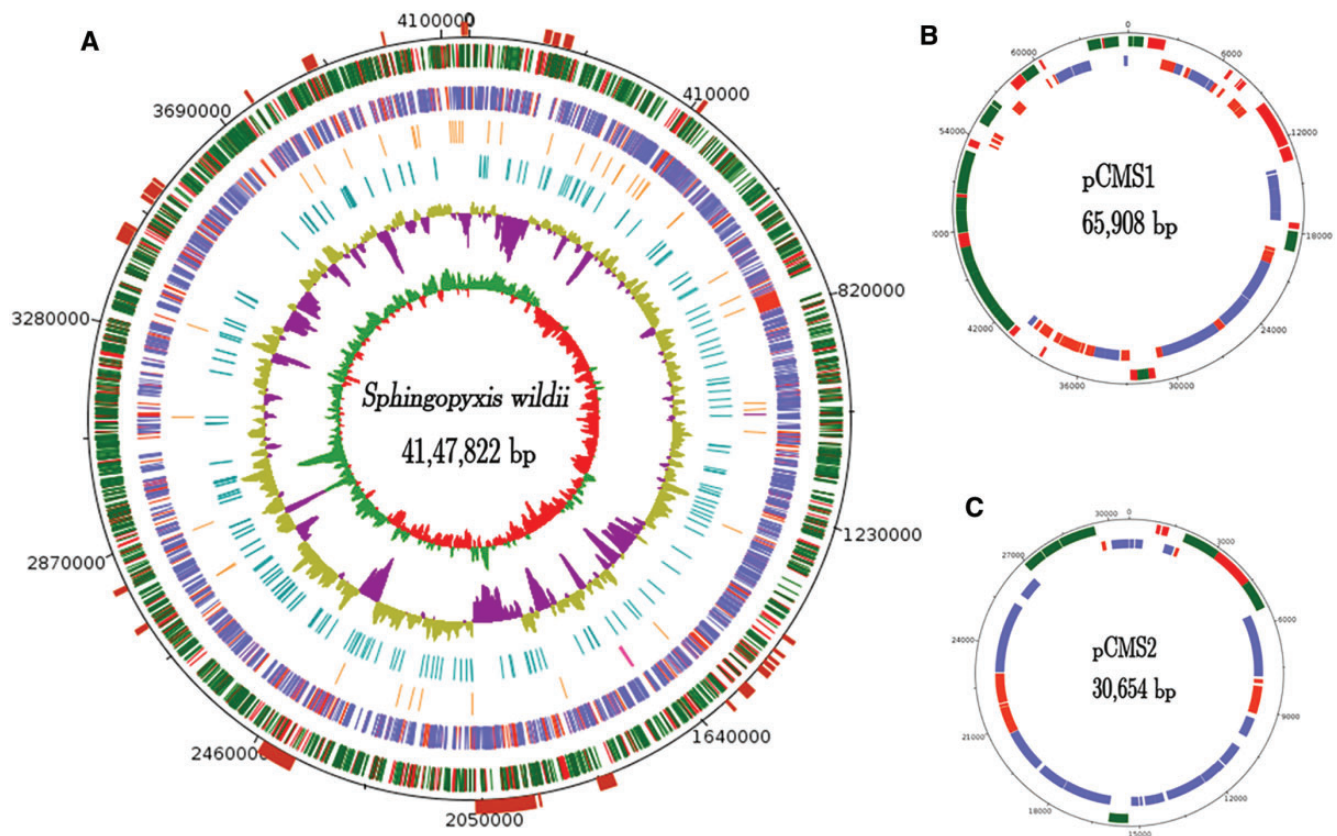
Fig. 1.—The circular maps of the *S. wildii* genome and plasmids: (*A*) Circles 1 and 2 (from exterior to interior) represent CDS on forward (green-annotated, red-hypothetical) and reverse strands (blue annotated, red hypothetical); circles 3 and 4 show RNA genes (orange for tRNA, pink for rRNA, and purple for tmRNA); and VNTRS (turquoise). The GC content (olive for positive and purple for negative) and GC skew (green for positive and red for negative) are shown in 5 and 6. The maroon blocks shown above circle 1 represent genomic islands. CDS on the forward (green for annotated, red for hypothetical) and reverse strands (blue for annotated, red for hypothetical) of plasmids pCMS1 and pCMS2 in (*B, C*).

known sequences, therefore the 30,654 bp second circular DNA is designated as plasmid pCMS2. Both chromosomal and plasmid sequences were comprehensively annotated for protein-coding genes (Supplementary Material online). The sequenced bacterial genome contained 3,966 protein-coding genes, among which 2,758 genes had an assigned function, and the remaining 1,208 were hypothetical genes (fig. 1*A*). Plasmids pCMS1 and pCMS2 contained 87 and 39 genes, respectively, and displayed no rRNA or tRNA coding sequences (fig. 1*B* and *C*).

Unique Genome Sequence

The sequenced genome was used to search against the NCBI representative database to identify species with maximum genome similarity. Considering the present classification, we expected the sequence to show maximum similarity with the genomes of the members of *Brevundimonas* genus. Instead, the sequence exhibited weak similarity with the genome sequences of strains belonging to four different genera, that is,

Sphingomonas, *Sphingopyxis*, *Sphingobium*, and *Novosphingobium*. The sequence of *B. diminuta* showed the highest similarity (as per Max score: 41,422) with *Sphingomonas wittichii*, whereas maximum coverage (57%) was observed with the genome sequence of *Sphingopyxis alaskensis* and maximum sequence identity with *Novosphingobium* sp. PP1Y (total score: 893,300). None of the genomes covered the entire sequences of the query genome suggesting that the *B. diminuta* sequence has no closely related genome sequences in the NCBI database. This appears to be the first report of the genome sequence of a novel species.

Comparative Genome Analysis

Interestingly, all the sequences that displayed a maximum similarity to the sequenced genome belonged to the Sphingomonadaceae family in the order Sphingomonadales. Since the genome sequence showed no continuous similarity, orthology analysis was done including available (120) genome sequences of Sphingomonadales and *E. coli* being included as
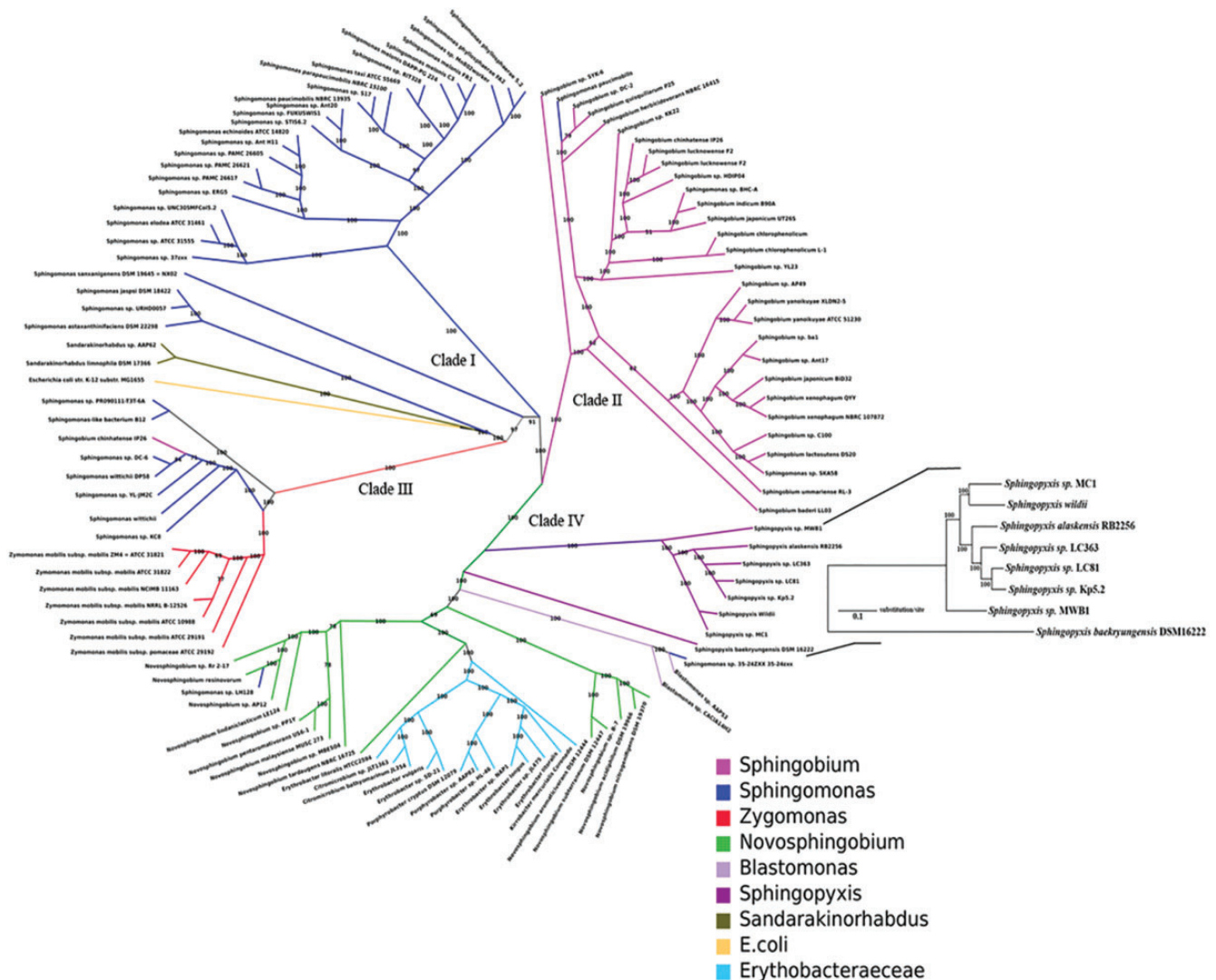
FIG. 2.—Phylogenetic tree: In the radial cladogram, the branches of each genus are colored uniquely. The branch length is not proportional to time. The scale bar indicates the number of substitutions per site. Each node of the tree is supported by a bootstrap value of 100. The tree in the inset depicts the relationship of *S. wildii* with other members of *Sphingopyxis*.

an out-group. Among the 122 bacterial genomes only 114 have shared more than 1,500 orthologous groups. Remaining eight genomes have displayed less number of orthologous groups which might be due to incomplete genome information. Since inclusion of such sequences causes bias in establishing the evolutionary distance and phylogenetic relationship, these bacterial genomes were discarded. The curated set of 114 genomes was further analyzed for single-copy orthologous. A total of 396 genes showing orthology across the Sphingomonadales were selected to infer phylogenetic relationships. A resultant phylogenetic tree showing the taxonomic placement and evolutionary distances of all 114 bacteria was constructed (fig. 2). Interestingly, *B. diminuta* clustered with the bacteria belonging

to the *Sphingopyxis* genus. As shown in figure 2, the complete phylogenetic tree was divided into four major clades, of which two major clades (I & II) contained only strains belonging to the *Sphingobium* and *Sphingomonas*. The third clade branched into two groups, that is, *Zygomonas* and *Sphingomonas* members. The fourth clade again diverged, separating the members of *Novosphingobium* and *Sphingopyxis*. Surprisingly, *Erythrobacter*, which belongs to a separate family of the Sphingomonadales, actually clustered in this clade, displaying evolutionary similarity with *Novosphingobium*. However, a few bacterial strains, such as *Sphingomonas* sp. SKA58 and *Sphingomonas paucimobilis*, showed unusual clustering with *Sphingobium* in clade I, while *Sphingomonas* sp. 35-24ZXX clustered with *Blastomonas*. This finding indicates that these

A

B

Fig. 3.—Degradation pathways: The pathways (A) and the genomic location of degradative traits are shown (B).

bacteria were either incorrectly classified or requires further refinement to determine their exact taxonomic position.

Pan/Core Genome Analysis

While gaining further clarity on the intra-genus relationship, we further performed pan/core genome analysis for the available seven genomes of *Sphingopyxis* together with the *B. diminuta*. A total of 4,575 orthologous groups were obtained for the genus *Sphingopyxis*, each containing anywhere

between 2 and 20 genes. Among the generated 4,575 orthologue groups only 1,515 clusters had at least one gene from each species. Therefore, these 1,515 single-copy orthologous present in each species were regarded as the core genome of the *Sphingopyxis* genus. We also identified unique genes (9–25%) in each species of the *Sphingopyxis* genus. The maximum number of unique genes was found in *Sphingopyxis* Kp5.2 (1,089), and relatively fewer unique genes were found in *S. baekryungensis* (786) showing considerable evolutionary distance with the remaining members of the

Sphingopyxis genus (fig. 2). A high-resolution phylogram inferred from the core genome of *Sphingopyxis* species confirmed that *B. diminuta* was closest to *Sphingopyxis* sp. MC1, while *S. baekryungensis* DSM 16222 was evolutionarily distant from other *Sphingopyxis* members. Further, these two phylogenetic trees clearly suggested existence of genome-to-genome distance between the sequenced genome and the genomes of other *Sphingopyxis* members warranting a new species name. Therefore, the organophosphate degrading *B. diminuta* is renamed as *Sphingopyxis wildii*.

Analysis of Natural Selection Pressure

Since *Sphingopyxis* members survive in harsh environments, the identified core genome (1,515) was analyzed for positive selection. Interestingly, ~18% of the core genome is positively selected. These positively selected genes are found in various Cluster of Orthologous Genes (COG) showing genome wide selection pressure (Tatusov et al. 2000). Interestingly, the genes related to cell wall/membrane/envelope biogenesis (M) and energy production and conversion (C) are selectively enriched than the other positively selected genes (supplementary material fig. S1, Supplementary Material online). The *bhpA*, which codes for cytochrome P450 monooxygenase, is found among positively selected genes. These monooxygenases perform regioselective hydroxylation of several man-made and natural aromatic compounds and serve as progenitors for evolution of peripheral oxygenases involved in degradation of aromatic compounds (Urlacher and Girhard 2012). Several peripheral oxygenase and dehalogenase coding genes are found in chromosome and plasmid sequences of *S. wildii* (fig. 3B). These enzymes generate catabolic intermediates from a variety of aromatic compounds that are funneled into β-ketoadipate pathway (fig. 3A) encoded by the indigenous plasmids pCMS1 and pCMS2 (fig. 3B). In support of the identified genetic repertoire

S. wildii has grown on these aromatic compounds using them as sole source of carbon.

Supplementary Material

Supplementary material and fig. S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Literature Cited

Munnecke DM, Hsieh DP. 1974. Microbial decontamination of parathion and p-nitrophenol in aqueous media. Appl. Microbiol. 28:212–217.

Pandeeti EVP, Chakka D, Pandey JP, Siddavattam D. 2011. Indigenous organophosphate-degrading (opd) plasmid pCMS1 of *Brevundimonas diminuta* is self-transmissible and plays a key role in horizontal mobility of the *opd* gene. Plasmid 65:226–231.

Segers P, et al. 1994. Classification of *Pseudomonas diminuta* Leifson and Hugh 1954 and *Pseudomonas vesicularis* Büsing, Döll, and Freytag 1953 in Brevundimonas gen. nov. as *Brevundimonas diminuta* comb. nov. and *Brevundimonas vesicularis* comb. nov., respectively. Int J Syst Bacteriol. 44:499–510.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28:33–36.

Urlacher VB, Girhard M. 2012. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. Trends Biotechnol. 30:26–36.

Associate editor: Bill Martin