High quality draft genomic sequence of *Flavihumibacter solisilvae* 3-3<sup>T</sup>

Gang Zhou¹, Chong Chen¹, Che Ok Jeon², Gejiao Wang¹* and Mingshun Li¹*

**Abstract**

*Flavihumibacter solisilvae* 3-3<sup>T</sup> (= KACC 17917<sup>T</sup> = JCM 19891<sup>T</sup>) represents a type strain of the genus *Flavihumibacter* within the family *Chitinophagaceae*. This strain can use various sole carbon sources, making it applicable in industry and bioremediation. In this study, the draft genomic information of *F. solisilvae* 3-3<sup>T</sup> is described. *F. solisilvae* 3-3<sup>T</sup> owns a genome size of 5.41 Mbp, 47% GC content and a total of 4,698 genes, including 4,215 protein coding genes, 439 pseudo genes and 44 RNA encoding genes. Analysis of its genome reveals high correlation between the genotypes and the phenotypes.

**Keywords:** *Flavihumibacter*, *Flavihumibacter solisilvae*, Genomic information, Genotype, Phenotype

**Introduction**

The genus *Flavihumibacter* was established in 2010 [1] and comprises three recognized species, *Flavihumibacter petaseus* T41<sup>T</sup> [1], *Flavihumibacter cheonanensis* WS16<sup>T</sup> [2] and *Flavihumibacter solisilvae* 3-3<sup>T</sup> [3], that were isolated from a subtropical rainforest soil, a shallow stream sediment and a forest soil, respectively. The *Flavihumibacter* members are Gram-positive, rod-shaped, strictly aerobic, non-motile, yellow-pigmented bacteria. The strains all contain phosphatidylethanolamine (as the major polar lipid, menaquinone-7 as the major respiratory quinine, iso-C<sub>15:0</sub> and iso-C<sub>15:1</sub> G as the principal fatty acids. In addition, the strains are oxidase- and catalase-positive and with a G + C content range of 45.9-49.5 mol% [1-3].

To the best of our knowledge, the genomic information of *Flavihumibacter* members still remains unknown. In this study, we present the draft genome information of *F. solisilvae* 3-3<sup>T</sup>. A polyphasic taxonomic study revealed that *F. solisilvae* 3-3<sup>T</sup> could utilize 33 kinds of sole carbon substrates, including 11 kinds of saccharides and 22 kinds of organic acids and amino acids [3]. Specialy, this strain could utilize aromatic compound 4-hydroxyphenylacetic acid as a sole carbon source making it applicable environmental bioremediation [4-6]. In addition, this strain could utilize quinic acid as a sole carbon. Quinic acid is the substrate used to synthesize aromatic amino acids (phenylalanine, tyrosine and tryptophan) via the shikimate pathway. These aromatic amino acids are very useful as food additives, sweetener and pharmaceutical intermediates [6, 7]. The genome analysis of *F. solisilvae* 3-3<sup>T</sup> will provide the genomic basis for better understanding these mechanisms and applying the strain to industries and bioremediation more efficiently.

**Organism information**

**Classification and features**

*F. solisilvae* 3-3<sup>T</sup> was isolated from forest soil of Bac Kan province in Vietnam [3]. The classification and features of *F. solisilvae* 3-3<sup>T</sup> are shown in Table 1. A maximum-likelihood tree was constructed based on the 16S rRNA gene sequences using MEGA 5.0 [8]. The bootstrap values were calculated based on 1,000 replications and distances were calculated in accordance with Kimura’s two-parameter method [9]. The phylogenetic tree showed that *F. solisilvae* 3-3<sup>T</sup> was clustered with the other *Flavihumibacter* members (Fig. 1).

Cells of *F. solisilvae* 3-3<sup>T</sup> (Fig. 2) are Gram-positive, aerobic, non-motile, and rod-shaped. Colony is yellow due to the production of flexirubin-type pigments [10]. *F. solisilvae* 3-3<sup>T</sup> grows well on NA and R2A agar...
It can hydrolyze aesculin, gelatin, casein and tyrosine [3].

F. solisilvae 3-3T can also utilize various carbohydrate substrates (Table 1) and produces several glycosyl hydrolases, such as β-N-acetylhexosaminidase, α-galactosidase, β-glucosidase, β-galactosidase, α-fucosidase, α-mannosidase and α-glucosidase [3].

F. solisilvae 3-3T contains iso-C15:0, iso-C16:1 G and summed feature 3 (C16:1ω6c/C16:1ω7c) as the principal fatty acids, MK-7 as the major respiratory quinine. The major polar lipids were PE, three unidentified aminolipids and three unidentified lipids [3].

**Genome sequencing information**

**Genome project history**

F. solisilvae 3-3T was selected for sequencing based on its taxonomic representativeness and the potential application in food industry and bioremediation. The genome of F. solisilvae 3-3T was sequenced at Wuhan Bio-Broad Co., Ltd, Hubei, China. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession JSVC00000000. The version described in this paper is version JSVC00000000.1. A summary of the genomic sequencing project information is given in Table 2.

**Table 1** Classification and general features of F. solisilvae 3-3T according to the MIGS recommendations [22]

| MIGS ID | Property       | Term                                                                 | Evidence code |
|---------|----------------|----------------------------------------------------------------------|---------------|
|         | Classification| Domain Bacteria                                                       | TAS [23]      |
|         |                | Phylum Bacteroidetes                                                 | TAS [24, 25]  |
|         |                | Class Sphingobacteria                                                | TAS [25, 26]  |
|         |                | Order Sphingobacterales                                              | TAS [25, 27]  |
|         |                | Family Chitinophagaceae                                              | TAS [28]      |
|         |                | Genus Flavihumibacter                                                | TAS [1]       |
|         |                | Species Flavihumibacter solisilvae                                   | TAS [3]       |
|         | Strain:        | 3-3T                                                                | TAS [3]       |
|         | Gram stain     | Positive                                                             | TAS [3]       |
|         | Cell shape     | Rod-shaped                                                           | TAS [3]       |
|         | Motility       | Non-motile                                                           | TAS [3]       |
|         | Sporulation    | Non-sporulating                                                      | NAS          |
|         | Temperature range | 20–37 °C                                                           | TAS [3]       |
|         | Optimum temperature | 28 °C                                                               | TAS [3]       |
|         | pH range; Optimum | 5.5–9.5; 7.5                                                        | TAS [3]       |
|         | Carbon source  | Sucrose, D-glucose, D-galactose, lactose, N-acetyl-glucosamine, L-arabinose, D-maltose, glycerol, dextrin, D-melibiose, glucuronamide, succinic acid mono-methyl ester, L-aspartic acid, D-galacturonic acid, D-glucosaminic acid, D-glucuronic acid, malonic acid, β-hydroxybutyric acid, 4-hydroxyphenylacetic acid, quinic acid, D-saccharic acid, bromosuccinic acid, succinic acid, L-pyroglutamic acid, urocanic acid, L-alanine, L-serine, D-alanine, L-histidine, L-alanyl-glycine, L-alaninamide and L-asparagine. | TAS [3]       |
|         | GS-6 Habitat   | Forest soil                                                          | TAS [3]       |
|         | MIGS-6.3 Salinity | 0–0.5 % NaCl (w/v)                                                   | TAS [3]       |
|         | MIGS-22 Oxygen requirement | Aerobic                                                              | TAS [3]       |
|         | MIGS-15 Biotic relationship | Free-living                                                          | NAS          |
|         | MIGS-14 Pathogenicity | Non-pathogenic                                                       | NAS          |
|         | MIGS-4 Geographic location | Bac Kan Province, Viet Nam                                           | TAS [3]       |
|         | MIGS-5 Sample collection | 2012                                                                 | TAS [3]       |
|         | MIGS-4.1 Latitude | Not reported                                                          | NAS          |
|         | MIGS-4.2 Longitude | Not reported                                                          | NAS          |
|         | MIGS-4.4 Altitude | Not reported                                                          | NAS          |

*Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [29]
Growth conditions and genomic DNA preparation

*F. solisilvae* 3-3 was grown aerobically in 50 ml R2A broth at 28 °C for 24 h with 160 r/min shaking. About 20 mg cells were harvested by centrifugation and suspended in normal saline, and then lysed using lysozyme. The DNA was obtained using the QiAamp kit according to the manufacturer's instruction (Qiagen, Germany).

Genome sequencing and assembly

The genome of *F. solisilvae* 3-3 was sequenced by Illumina Hiseq 2000 technology [11] with Paired-End library strategy (300 bp insert size). TruSeq DNA Sample Preparation Kits are used to prepare DNA libraries with insert sizes from 300–500 bp for single, paired-end, and multiplexed sequencing. The protocol supports shearing by either sonication or nebulization of 1 μg of DNA [12]. The genome of *F. solisilvae* 3-3 generated 7,041,525 reads totaling 1,422,388,050 bp data with an average coverage of 250 x. Sequence assembly and quality assessment were performed using velvet v.1.2.10 [13] software. Finally, all reads were assembled into 75 contigs (> 200 bp) with a genome size of 5.41 Mbp.

**Genome annotation**

Genome annotation was performed through the NCBI Prokaryotic Genome Annotation Pipeline which combined the Best-Placed reference protein set and the gene caller GeneMarkS+. WebMGA-server [14] with E-value cutoff 1-e3 was used to assess the COGs. The translated predicted CDSs were also used to search against the Pfam protein families database [15]. TMHMM Server v.2.0 [16], SignalP 4.1 Server [17] and CRISPRfinder program [18] were used to predict transmembrane helices,

![Fig. 1 A 16S rRNA gene based ML phylogenetic tree showing the phylogenetic position of *F. solisilvae* 3-3. Bootstrap values (>50 %) based on 1,000 replications are shown at branch nodes. Bar, 1 substitutions per 100 nucleotide positions. Sphingobacterium alimentarium DSM 2137 is used as the outgroup. The GenBank accession numbers are shown in parentheses.](image)

**Fig. 2** A transmission electron micrograph of *F. solisilvae* 3-3 cell. The bar indicates 0.5 μm
signal peptides and CRISPRs in the genome, respectively. The metabolic pathway analysis were constructed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) [19].

**Genome properties**

The draft genome size of *F. solisilvae* 3-3 is 5,410,659 bp with 47 % GC content and contains 75 contigs. From a total of 4,698 genes, 4,215 (89.72 %) genes are protein coding genes, 439 (9.34 %) are pseudo genes and 44 (0.94 %) are RNA encoding genes. The genome properties and statistics are shown in Table 3 and Fig. 3. Altogether, 3,137 (74.42 %) protein coding genes are distributed into COG functional categories (Table 4).

**Insights from the genome sequence**

*F. solisilvae* 3-3 could grow on 33 kinds of sole carbon substrates including saccharides, organic acids and amino acids [3] (Table 1). Analysis of the genome reveals that this strain possesses putative enzymes for central carbohydrate metabolism to assimilate these carbon sources through

---

**Table 3** Genome statistics of *F. solisilvae* 3-3

| Attribute                        | Value      | % of total |
|----------------------------------|------------|------------|
| Genome size (bp)                 | 5,410,659  | 100.00     |
| DNA coding (bp)                  | 4,540,989  | 83.93      |
| DNA G + C (bp)                   | 2,543,035  | 47.00      |
| DNA scaffolds                    | 75         | -          |
| Total genes                      | 4,698      | 100.00     |
| Protein coding genes             | 4,215      | 89.72      |
| RNA genes                        | 44         | 0.94       |
| Pseudo genes                     | 439        | 9.34       |
| Frameshifted Genes               | 9          | 0.19       |
| Genes with function prediction   | 1,893      | 44.91      |
| Genes assigned to COGs           | 3,137      | 74.42      |
| Genes with Pfam domains          | 3,511      | 83.30      |
| Genes with signal peptides       | 670        | 15.89      |
| Genes with transmembrane helices | 919        | 21.80      |
| CRISPR repeats                   | 1          | -          |

*The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.*

---

**Fig. 3** A Graphical circular map of *F. solisilvae* 3-3 genome. From outside to center, ring 1, 4 show protein-coding genes colored by COG categories on forward or reverse strand; ring 2, 3 denote genes on forward or reverse strand; ring 5 shows G + C content plot, and the innermost ring represents GC skew.
different metabolic pathways [20]. The putative enzymes that responsible to the utilization of 20 sole carbons were found in the genome (Table 5). All key enzymes in the Embden-Meyerhof-Parnas pathway (glucokinase, pyruvate kinase and 6-phosphofructokinase) and TCA cycle are present in *F. solisilvae* 3-3T. The key enzymes of Pentose Phosphate pathway (glucose-6-phosphate dehydrogenase, 6-phosphogluconolactonase and 6-phosphogluconate dehydrogenase) were also found.

The presence of 4-hydroxyphenylpyruvate dioxygenase (KIC95062), homogentisate 1,2-dioxygenase (KIC93392) and other related enzymes suggests that 4-hydroxyphenylacetic acid is degradable via homogentisic acid pathway [21]. In addition, the presence of 3-dehydroquinate dehydratase (KIC93382), shikimate dehydrogenase (KIC92987), shikimate kinase (KIC93265), 3-phosphoshikimate 1-carboxyvinyltransferase (KIC94147) and chorismate synthase (KIC94148) indicates that *F. solisilvae* 3-3T could probably utilize quinic acid to synthesize the three aromatic amino acids (tryptophan, tyrosine and phenylalanine) via shikimate pathway [7].

**Conclusion**

To the best of our knowledge, this report provides the first genomic information of the genus *Flavihumibacter*. Analysis of the genome shows high correlation between the genotypes and the phenotypes. The genome possesses many key proteins of central carbohydrate metabolism which provides the genomic basis to utilize the various carbon sources. In addition, analyzing its genome indicates that this strain has potential application for the production of aromatic amino acids and for environmental bioremediation.
Abbreviations
TCA: Tricarboxylic acid cycle; PE: Phosphatidylethanolamine; MK-7: Menaquinone-7; PGAP: Prokaryotic genome annotation pipeline.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
GZ performed laboratory experiments, analyzed the data and wrote the draft manuscript. CC and COJ cultured samples, analyzed the data and revised the manuscript. ML performed the genome comparison and revised the manuscript. GW organized the study and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgment
This work was supported by the National Natural Science Foundation of China (31470227).

Author details
1State Key Laboratory of Agricultural Microbiology, College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, People’s Republic of China. 2Department of Life Science, Chung-Ang University, Seoul 156-756, Republic of Korea.

Received: 10 February 2015 Accepted: 8 July 2015
Published online: 19 September 2015

References
1. Zhang NN, Qu JH, Yuan HL, Sun YM, Yang JS. Flavohumibacter petasus gen. nov., sp. nov., isolated from soil of a subtropical rainforest. Int J Syst Evol Microbiol. 2010;60:1609–12. doi:10.1099/ijs.0.011957-0.
2. Kim WH, Lee S, Ahn Y. Flavohumibacter cheniannensis sp. nov., isolated from sediment of a shallow stream. Int J Syst Evol Microbiol. 2014;64:3235–9. doi:10.1099/ijsem.0.063370-0.
3. Lee HJ, Jeong SE, Cho MS, Kim SH, Lee SS, Lee BH, et al. Flavohumibacter solisilvae sp. nov., isolated from forest soil. Int J Syst Evol Microbiol. 2014;64:2897–901. doi:10.1099/ijsem.0.063669-0.
4. Martin M, Cibello A, Fernandez J, Ferrer E, Garindo-Pertielra A. Catabolism of 3- and 4-hydroxyphenylacetic acid by Kelsbiella pneumoniae. J Gen Microbiol. 1991;137:621–8. doi:10.1099/00221287-137-3-621.
5. Méndez V, Agulló L, González M, Seeger M. The homogenitase and homoprotocatechuate central pathways are involved in 3- and 4-hydroxyphenylacetate degradation by Burkholderia xylorivae LB400. PLoS One. 2011;6:e17583. doi:10.1371/journal.pone.0017583.
6. Koma D, Yamanaoka H, Miroyoshii K, Omoto M, Sakai K. Production of aromatic compounds by metabolically engineered Eschenichia coli with an expanded shikimate central pathways. Appl Environ Microbiol. 2012;78:2023–16. doi:10.1128/AEM.01418-12.
7. Guo J, Carrington Y, Alber A, Ehting J. Molecular characterization of quinate and shikimate metabolism in Populus trichocarpa. J Biol Chem. 2014;289:28846–58. doi:10.1074/jbc.M114.558536.
8. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:753–9. doi:10.1093/molbev/msr121.
9. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16:111–20.
10. Bernardet JF, Nakagawa A, Höltje J, Besemer P. Proposed minimal standards for describing new taxa of the family Flavobacteriacae and emended description of the family. Int J Syst Evol Microbiol. 2002;52:1049–70. doi:10.1099/ijs.0.02136-0.
11. Bennett S. Solexa Ltd. Pharmacogenomics. 2004;5:433–8. doi:10.1517/14622416.5.4.433.
12. Illumina official website. www.illumina.com.cn
13. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9. doi:10.1101/gr.074921.107.
14. Wu S, Zhu ZW, Fu L, Niu BF, Li WZ. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 2011;12:444. doi:10.1186/1471-2164-12-444.
15. The Pfam protein families database. [http://pfam.xfam.org/search]