

Phylogeny of Flavobacteria Group Isolated from Freshwater Using Multilocus Sequencing Analysis

Seyoung Mun¹, Jungnam Lee^{1†}, Siwon Lee², Kyudong Han¹, Tae-Young Ahn^{3*}

¹Department of Nanobiomedical Science & BK21 PLUS NBM Global Research Center for Regenerative Medicine, Dankook University, Cheonan 330-714, Korea, ²Plant Quarantine Technology Center, Animal and Plant Quarantine Agency, Suwon 443-440, Korea, ³Department of Microbiology, Dankook University, Cheonan 330-714, Korea

Sequence analysis of the 16S rRNA gene has been widely used for the classification of microorganisms. However, we have been unable to clearly identify five *Flavobacterium* species isolated from a freshwater by using the gene as a single marker, because the evolutionary history is incomplete and the pace of DNA substitutions is relatively rapid in the bacteria. In this study, we tried to classify *Flavobacterium* species through multilocus sequence analysis (MLSA), which is a practical and reliable technique for the identification or classification of bacteria. The five *Flavobacterium* species isolated from freshwater and 37 other strains were classified based on six housekeeping genes: *gyrB*, *dnaK*, *tuf*, *murG*, *atpA*, and *glyA*. The genes were amplified by PCR and subjected to DNA sequencing. Based on the combined DNA sequence (4,412 bp) of the six housekeeping genes, we analyzed the phylogenetic relationship among the *Flavobacterium* species. The results indicated that MLSA, based on the six housekeeping genes, is a trustworthy method for the identification of closely related *Flavobacterium* species.

Keywords: *Flavobacterium*, multilocus sequence typing, phylogeny

Introduction

Flavobacterium is a genus of gram-negative bacteria belonging to the family Flavobacteriaceae (order: Flavobacteriales, class: Flavobacteria, phylum: Bacteroidetes) [1]. The term *Flavobacterium* was first introduced by Bergey *et al.* in 1923 [1]. Since then, a number of *Flavobacterium* species have been identified, and the classification of the genus has been continuously amended. At present, it is widely accepted that the genus *Flavobacterium* contains bacteria that are yellow-pigmented, aerobic, and generally rod-shaped [2], and some *Flavobacterium* species mobilize through a gliding motility mechanism [3]. *Flavobacterium* strains could be isolated from various habitats, including freshwater, soil, plants, and food [3, 4].

The 16S rRNA gene has been widely used in many studies concerned with the identification or classification of microorganisms. However, using only one housekeeping gene as a single marker is not enough to distinguish some bacterial

species, especially when they are closely related. In addition, gene-specific changes caused by horizontal gene transfer or recombination could be an obstacle to the phylogenetic studies that are based on a single gene marker [5, 6]. To overcome the limitation, multilocus sequence analysis (MLSA) has been proposed as an alternative method in many studies. In reality, MLSA is a microbial genotyping method used to characterize a variety of microorganisms, including the genera of *Pantoea* [7], *Bifidobacterium* [8], *Borrelia* [9], *Bradyrhizobium* [10], and *Bacteroides* [11]. In MLSA, approximately seven housekeeping genes are required, but a standard for the given number and size of genes has not been established. This concise approach is based on concatenated sequences of multiple housekeeping genes, uses the concatenated sequences to identify the phylogenetic aspects of desired species, and investigates the patterns of genotypes between each species cluster [12].

This study was designed to classify the genetic diversity of collected *Flavobacterium* species. In addition, we also inve-

Received October 3, 2013; Revised November 15, 2013; Accepted November 16, 2013

*Corresponding author: Tel: +82-41-550-3451, Fax: +82-41-550-3450, E-mail: ahnty@dankook.ac.kr

†Present address: Department of Periodontology and Oral Biology, College of Dentistry, University of Florida, Gainesville, FL 32610, USA

Copyright © 2013 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>).

stigated their relationship using phylogenetic analysis and compared the results for phylogenetic markers.

Methods

Bacterial isolates

Seven *Flavobacterium* isolates were collected. The two *Flavobacterium* reference strains, *F. aquatile* ATCC 11947 and *F. frigidimaris* F3-JCM12218, were purchased from the Japan Collection of Microorganisms (JCM), and five *Flavobacterium* isolates – *F. koreense* ARSA-15, *F. koreense* ARSA-19, *F. koreense* ARSA-42, *F. chungnamense* ARSA-103, and *F. chungnamense* ARSA-108 – were previously isolated from the Cheon-ho reservoir in the Korea [13, 14].

Bacterial growth and genomic DNA isolation

This seven strains were cultured on R2A medium (Difco, Detroit, MI, USA) for 5 days at 25°C and produced cream-yellow, semi-transparent colonies. We extracted genomic DNA from the seven *Flavobacterium* strains using the Wizard genomic DNA purification kit (Promega, Madison, WI, USA), according to the manufacturer's instruction.

Polymerase chain reaction

Through a PCR assay, we amplified seven housekeeping genes from the bacteria: 16S rRNA (16S ribosomal RNA) gene, *atpA* (ATP synthase alpha subunit), *dnaK* (molecular chaperone protein), *glyA* (serine hydroxy-methyltransferase), *gyrB* (DNA gyrase, subunit B), *murG* (N-acetylglucosaminyl transferase), and *tuf* (elongation factor Tu). For the PCR amplification of the genes, we used primers

designed in a previous study [15]. However, we failed to amplify the *dnaK* and *atpA* genes through PCR with the previously designed primers. Thus, we manually designed new PCR primers to amplify these two genes. The primers are listed in Table 1. PCR amplification of each gene was performed in a 25- μ L reaction volume using 10 μ L of a DNA polymerase kit (TaKaRa EmeraldAmp GT PCR Master Mix; TaKaRa, Ohtsu, Japan), 200 nM of each oligonucleotide primer, and 50 ng of template DNA. The PCR conditions were an initial denaturation at 95°C for 5 min; followed by 35 cycles of 30 s of denaturation at 95°C, 30 s of annealing at 55°C, and 2 min of extension at 72°C; and followed by 5 min of a final extension at 72°C. Two microliters of the resulting PCR products was loaded on a 1% agarose gel, stained with ethidium bromide, and visualized using UV fluorescence. The remaining PCR products were purified with the Wizard SV gel and PCR clean-up kit (Promega, Madison, USA) according to the manufacturer's instruction and sequenced on an ABI 3130XL automated DNA sequencer (Applied Biosystems, Foster City, CA, USA).

DNA sequences analysis

For each of the seven genes, we aligned the DNA sequences obtained from the seven *Flavobacterium* isolates using the software BioEdit, version 7.0 (<http://www.mbio.ncsu.edu/Bioedit/bioedit.html/>). After that, the housekeeping gene sequences of the 14 added reference strains were collected using BLAST (<http://blast.ncbi.nlm.nih.gov/>) and aligned in the same way.

Table 1. Primers used for multilocus sequence analysis of the *Flavobacterim* species in this study

Gene	Gene product	Primer	Sequence (5' → 3')	Amplicon size (bp)	Analyzed size (bp)	Reference
<i>gyrB</i>	DNA gyrase, subunit B	<i>gyrB_F</i>	TTGTAATGACTAAAATTGGTGC	1,042	555	Nicolas et al. [15]
		<i>gyrB_R</i>	GCCGAATCTCCCTCGACAAG			
<i>dnaK</i>	Molecular chaperone protein	<i>dnaK_F</i>	AAGGTGGAGAAATTAAGTAGG	1,070	680	Nicolas et al. [15]
		<i>dnaK_AF</i>	GCTGAAGACTATTTAGGTC	828		This study
		<i>dnaK_R</i>	CCACCCATAGTTTCGATACC			Nicolas et al. [15]
<i>tuf</i>	Elongation factor Tu	<i>tuf_F</i>	GAAGAAAAAGAAAGAGGTATTAC	982	833	Nicolas et al. [15]
		<i>tuf_R</i>	CACCTTCACGGATAGCGAA			
<i>murG</i>	N-acetylglucosaminyl transferase	<i>murG_F</i>	TGGCGGTACAGGAGGACATAT	870	725	Nicolas et al. [15]
		<i>murG_R</i>	GCATTCTTGGTTTGATGGTCTTC			
<i>atpA</i>	ATP synthase alpha subunit	<i>atpA_F</i>	CTTGAAGAAGATAATGTGGG	1,020	796	Nicolas et al. [15]
		<i>atpA_R</i>	TGTTCCAGCTACTTTTTTTCAT			
		<i>atpA_AR</i>	GAGATACCGAAATACCTACG	970		Present study
<i>glyA</i>	Serine hydroxy-methyltransferase	<i>glyA_F</i>	AAAGATAGACAAAATTCACGG	1,042	823	Nicolas et al. [15]
		<i>glyA_R</i>	GGTGATTTATCATCAAAGG			
16S rRNA	16S ribosomal RNA	27_F	AGAGTTTGATCCTGGCTCAG	1,465	799	Universal
		1492_R	GGTTACCTTGTACGACTT			

Phylogenetic analysis

To construct the phylogenetic relationship of the 21 *Flavobacterium* strains, we used the molecular evolutionary genetics analysis (MEGA) software package, version 5.05 [16], with the Kimura two-parameter model and bootstrap analyses, based on 1,000 replications. In addition, we compared the genetic similarity of the six housekeeping genes and the 16S rRNA gene between the genus of *Flavobacterium* using the pair distance of ClustalW (weighted), which was supported by DNASTAR's MegAlign program in Lasergene, version 5.0 (<http://www.dnastar.com>). The MEGA program was also used to determine the pairwise distances, G+C contents, and substitution positions.

Nucleotide sequence accession numbers

The DNA sequences of the seven amplified *Flavobacterium* isolates have been deposited in GenBank under accession numbers JX256829-JX256870 (Supplementary Table 1).

Results and Discussion

In this study, to determine the phylogenetic relationship of the five closely related *Flavobacterium* species that were

previously isolated from a freshwater reservoir in Chonan, Korea [13, 14], we constructed a maximum-likelihood phylogenetic tree, but the tree showed a very low bootstrap value (<25%) between them. Previously, Nicolas *et al.* [15] recommended several housekeeping genes to analyze the phylogenetic relationship among fish-pathogenic *F. psychrophilum* strains. This approach is called MLSA, which is very useful for distinguishing closely related species and showing population structure. We have chosen seven housekeeping genes as phylogenetic markers—*atpA*, *dnaK*, *glyA*, *gyrB*, *murG*, *tuf* (elongation factor Tu), and *trpB* (tryptophan synthase, subunit B)—because these loci contain the largest amounts of polymorphisms among the 50 *F. psychrophilum* strains [15].

This study is the first to apply *Flavobacterium* species isolated from a freshwater reservoir. First, we tried to amplify seven housekeeping genes from five *Flavobacterium* isolates [13, 14], which are closely related to each other based on the 16S rRNA gene. However, we could not amplify the *trpB* gene from our samples from a freshwater reservoir. Therefore, we collected the sequences of six housekeeping genes and the 16S rRNA gene from five *Flavobacterium* isolates [13, 14], isolated from a freshwater reservoir in Cheonan, Korea, and two *Flavobacterium* reference strains, *F. aquatile* ATCC 11947 and *F. frigidimaris* F3-JCM122, as

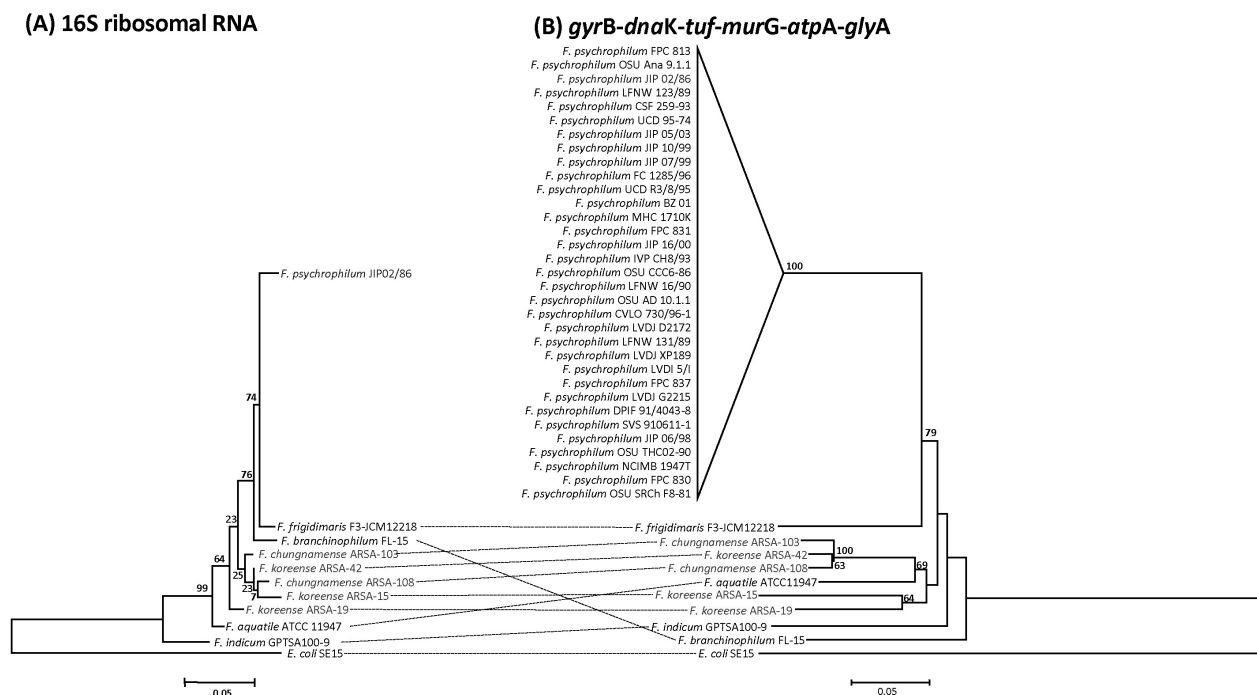


Fig. 1. Maximum-likelihood phylogenetic tree showing the phylogenetic relationship of 16S rRNA gene sequences (A) and the concatenated sequences of *gyrB-dnaK-tuf-murG-atpA-glyA* genes obtained from 42 *Flavobacterium* species strains and *Escherichia coli* (B). Numbers on the branches indicate bootstrap values (shown as percentages of 1,000 replications). Scale bar, 0.05 substitutions per nucleotide position. Purple box indicates discordant cluster between two phylogenetic trees.

interspecies out-groups. In addition, we searched for more *Flavobacterium* species that were available for the six housekeeping genes from the GenBank database. We found that 52 *Flavobacterium* species were available for all six housekeeping gene sequences in the database. Among them, 50 *Flavobacterium* species were *F. psychrophilum* that were identified and studied by Nicolas *et al.* [15]. The 50 *F. psychrophila* were assigned to 33 sequence types. Consequently, to evaluate the resolution and efficiency of MLSA, we aligned and trimmed each housekeeping gene sequence (555 bp of *gyrB*, 680 bp of *dnaK*, 833 bp of *tuf*, 725 bp of *murG*, 796 bp of *atpA*, and 823 bp of *glyA* genes) from the five *Flavobacterium* isolates, 33 distinct *F. psychrophila*, two other *Flavobacterium* species (*F. branchiophilum* FL-15 and *F. indicum* GPTSA100-9), and two reference strains. As shown in Fig. 1, we constructed two different phylogenetic trees of the *Flavobacterium* species, based on only 16S rRNA gene sequences (Fig. 1A) and based on the concatenated sequences of the six housekeeping genes (Fig. 1B). The bootstrap values of the tree branches clearly improved in Fig. 1B compared with those in Fig. 1A. Therefore, we could state that the phylogenetic tree of the genus *Flavobacterium*, produced by the MLSA approach, is more reliable than the tree based on a single locus (e.g., 16S rRNA gene).

To investigate the nucleotide diversity of each gene among the 10 *Flavobacterium* species, we analyzed the pairwise distances determined by MEGA5.05 (Table 2). Among the six housekeeping genes and the 16S rRNA gene, the highest average pairwise distance was 0.303 in *murG*, while the lowest average pairwise distance was 0.043 in the 16S rRNA gene. We found that the average pairwise distance within the *gyrB*, *dnaK*, *tuf*, *atpA*, and *glyA* genes was 0.218, 0.141, 0.115, 0.150, and 0.215, respectively, among *Flavobacterium* species. In addition, we estimated the pairwise distance of the six concatenated housekeeping genes, and thus, the pairwise distance was 0.164 on average. The pairwise distance of the six concatenated housekeeping genes was significantly greater than the pairwise distance of the 16S rRNA gene

Table 2. Average pairwise distances, calculated using Kimura two-parameter model, and average G + C content of the seven housekeeping genes in this study

Locus	Average pairwise distance	Average G + C content (%)
<i>gyrB</i>	0.218	36.2
<i>dnaK</i>	0.141	36.3
<i>tuf</i>	0.115	40.0
<i>murG</i>	0.303	35.4
<i>atpA</i>	0.150	40.3
<i>glyA</i>	0.215	38.1
16S rRNA	0.043	48.9

(0.043). Next, we analyzed G + C contents and variable nucleotide sites of six housekeeping genes in our MLSA data and 16S rRNA gene (Table 2). The G + C contents of the genes ranged from 35.4% (*murG*) to 48.9% (16S rRNA) (Table 2). The G + C content among housekeeping genes was found to be variable.

As a result, this study has demonstrated that MLSA, based on multiloci, produces more reliable phylogenetic trees among *Flavobacterium* species compared with the tree based on a single locus (e.g., 16S rRNA). When a high number of sequential similarities exist in genetic markers, it might be difficult to identify the phylogenetic relationship between microorganisms of the same genus. Therefore, MLSA will be a good approach to classify the genetic relationships among microorganisms.

Supplementary material

Supplementary data including one table can be found with this article online at <http://www.genominfo.org/src/sm/gni-11-272-s001.pdf>.

References

- Bergey DH, Harrison FC, Breed RS, Hammer BW, Huntoon FM. *Manual of Determinative Bacteriology*. Baltimore: Williams and Wilkins Co., 1923.
- Park M, Lu S, Ryu SH, Chung BS, Park W, Kim CJ, *et al.* *Flavobacterium croceum* sp. nov., isolated from activated sludge. *Int J Syst Evol Microbiol* 2006;56(Pt 10):2443-2447.
- Bernardet JF, Nakagawa Y, Holmes B; Subcommittee on the taxonomy of *Flavobacterium* and Cytophaga-like bacteria of the International Committee on Systematics of Prokaryotes. Proposed minimal standards for describing new taxa of the family Flavobacteriaceae and emended description of the family. *Int J Syst Evol Microbiol* 2002;52:1049-1070.
- Cousin S, Pauker O, Stackebrandt E. *Flavobacterium aquidurense* sp. nov. and *Flavobacterium hercynium* sp. nov., from a hard-water creek. *Int J Syst Evol Microbiol* 2007;57(Pt 2):243-249.
- Boucher Y, Douady CJ, Sharma AK, Kamekura M, Doolittle WF. Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J Bacteriol* 2004;186:3980-3990.
- Gogarten JP, Doolittle WF, Lawrence JG. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 2002;19:2226-2238.
- Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J, Coutinho T. Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Syst Appl Microbiol* 2008;31:447-460.
- Deletoile A, Passet V, Aires J, Chambaud I, Butel MJ, Smokvina T, *et al.* Species delineation and clonal diversity in four *Bifidobacterium* species as revealed by multilocus sequen-

- cing. *Res Microbiol* 2010;161:82-90.
9. Margos G, Vollmer SA, Cornet M, Garnier M, Fingerle V, Wilske B, et al. A new *Borrelia* species defined by multilocus sequence analysis of housekeeping genes. *Appl Environ Microbiol* 2009;75:5410-5416.
 10. Rivas R, Martens M, de Lajudie P, Willems A. Multilocus sequence analysis of the genus *Bradyrhizobium*. *Syst Appl Microbiol* 2009;32:101-110.
 11. Sakamoto M, Ohkuma M. Identification and classification of the genus *Bacteroides* by multilocus sequence analysis. *Microbiology* 2011;157(Pt 12):3388-3397.
 12. Hanage WP, Fraser C, Spratt BG. Sequences, sequence clusters and bacterial species. *Philos Trans R Soc Lond B Biol Sci* 2006; 361:1917-1927.
 13. Lee S, Weon HY, Han K, Ahn TY. *Flavobacterium dankookense* sp. nov., isolated from a freshwater reservoir, and emended descriptions of *Flavobacterium cheonanense*, *F. chungnamense*, *F. koreense* and *F. aquatile*. *Int J Syst Evol Microbiol* 2012;62(Pt 10):2378-2382.
 14. Lee S, Weon HY, Kim SJ, Ahn TY. *Flavobacterium koreense* sp. nov., *Flavobacterium chungnamense* sp. nov., and *Flavobacterium cheonanense* sp. nov., isolated from a freshwater reservoir. *J Microbiol* 2011;49:387-392.
 15. Nicolas P, Mondot S, Achaz G, Bouchenot C, Bernardet JF, Duchaud E. Population structure of the fish-pathogenic bacterium *Flavobacterium psychrophilum*. *Appl Environ Microbiol* 2008; 74:3702-3709.
 16. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731-2739.