Variation and predicted structure of the flagellin gene in Actinoplanes species

Hideki Yamamura1, Keitaro Hanawa1, Masami Kusunoki1, Kazuo Nakamura1, Youji Nakagawa1, Yuumi Ishida2, Moriyuki Hamada2, Misa Otoguro2, Tomohiko Tamura2, Nobuyuki Fujita2 & Masayuki Hayakawa1
1Division of Applied Biological Sciences, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Yamanashi, Japan; and 2NITE Biological Resource Center, National Institute of Technology and Evaluation, Chiba, Japan

Correspondence: Hideki Yamamura, Division of Applied Biological Sciences, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Takeda-4, Kofu 400-8511, Japan. Tel./fax: +81 55 220 8546; e-mail: hyamamura@yamanashi.ac.jp

Received 20 June 2011; revised 4 August 2011; accepted 5 August 2011. Final version published online 8 September 2011.

DOI: 10.1111/j.1574-6968.2011.02380.x

Editor: André Klier

Keywords
motile actinomycetes; flagellin; structural prediction; Actinoplanes.

Abstract

Members of the genus Actinoplanes are considered to be representative of motile actinomycetes. To infer the flagellar diversity of Actinoplanes species, novel degenerate primers were designed for the flagellin (fliC) gene. The fliC gene of 21 Actinoplanes strains was successfully amplified and classified into two groups based on whether they were large (type I) or small (type II). Comparison of the translated amino acid sequences revealed that this size difference could be attributed to large number of gaps located in the central variable region. However, the C- and N-terminal regions were conserved. Except for a region on the flagellum surface, structural predictions of type I and II flagellins revealed that the two flagellin types were strongly correlated with each other. Phylogenetic analysis of the 115-amino acid N-terminal sequences revealed that the Actinoplanes species formed three clusters, and type II flagellin gene containing three type strains were phylogenetically closely related each other.

Introduction

The genus Actinoplanes (Couch, 1950; Stackebrandt & Kroppenstedt, 1987) is a member of the family Micromonosporaceae (Krasil’nikov, 1938; Zhi et al., 2009), and is characterized by the presence of spherical, subspherical, cylindrical or very irregular sporangia (Lechevalier et al., 1966). The motile sporangiospores move by means of polar or peritrichous flagella (Couch, 1950). The flagellated spores exhibit chemotactic properties and are attracted to a variety of substrates, including those that contain bromide or chloride ions (Palleroni, 1976), fungal conidia, chlamydospores, sclerotia, or exudates of these (Arora, 1986), γ-collidine, d-Xylose, and pollen (Hayakawa et al., 1991a, b).

Phylogenetic analyses based on the 16S rRNA gene sequences of members of the family Micromonosporaceae revealed that motile genera, such as the Actinoplanes, do not form coherent clusters or linageas (Inahashi et al., 2010). Similarly, other motile actinomycetes were phylogenetically distributed among at least 20 families in the order Actinomycetales. Indeed, these findings indicate that the relationship between phylogeny and the propagation of the gene(s) encoding the flagellar system in prokaryotic organisms, including actinomycetes, is unclear.

Bacterial flagella are considered to be composed of three parts: a basal body, a hook, and a filament (Macnab, 1992). The filament is composed of the flagellin protein, which is synthesized internally and transported through the cell membrane to an external site for flagellum assembly (Snyder et al., 2009). The flagellin-encoding gene, fliC, has been used previously as a biomarker in studies of the taxonomy, epidemiology, and virulence of Burkholderia cepacia, Borrelia spp., and Clostridium difficile (Fukunaga & Koreki, 1996; Hales et al., 1998; Tasteyre et al., 2000). However, few studies have been conducted to date on the flagellar protein (Vesselinova & Ensign, 1996; Uchida et al., 2011) of motile actinomycetes. Vesselinova & Ensign (1996) reported that flagellins show two different sizes (32–43 and 42–43 kDa) in Actinoplanes spp.
Recent advances in whole genome sequence analysis have facilitated examinations of bacterial flagellar diversity. Snyder et al. (2009) reported the distribution of flagellar genes and the predicted nucleotide sequences of the genes responsible for synthesis of flagellar systems using BLASTP in a mutual-best-hit approach (e-value < 0.05) and a text-based search of genome annotations for 516 bacterial genomes. Their article also described the distribution of flagellar systems in 43 actinobacterial genomes, as well as in four actinomycetes that possessed a flagellin gene (e.g., Nocardioidea sp. JS614, Leifsonia xyli, Acidothermus cellulolyticus, and Kineococcus radiotolerans). Analysis of these four actinomycete genomes revealed that there were no genes encoding FlgF (proximal rod) and FlgG (distal rod), and that the flagellar system may be incomplete (Snyder et al., 2009). However, all species belonging to the genus Kineococcus are motile, and polar flagella have been observed in K. radiotolerans SRS30216 (Phillips et al., 2002). Similarly, several species belonging to the genera Nocardioidea and Leifsonia were observed to have motile cells and flagella (Cho et al., 2010; Madhaiyan et al., 2010). Interestingly, whole genome sequence data from A. cellulolyticus 11B revealed the presence of a flagellar system, even though this actinomycete species was previously reported to be non-motile (Barabote et al., 2009). In addition, genes for the flagellar system in Salmonispora tropica, CNB-440, which belongs to the family Micromonosporaceae, have not yet been identified (Udoway et al., 2007). Taken together, these findings indicate that the distribution and diversity of flagellar genes in actinomycetes is unclear (Snyder et al., 2009). This study therefore sought to characterize the flagellin-encoding gene in Actinoplanes species as a representative of motile actinomycetes.

In this article, we amplified, sequenced and analyzed flagellin gene sequences from selected Actinoplanes type strains. In addition, structural predictions were performed using the SWISS-MODEL server (Schwede et al., 2003), with a template from a known flagellin protein from Salmonella typhimurium (Maki-Yonekura et al., 2010). Finally, phylogenetic analysis based on the N-terminal region of the flagellin gene was conducted and the obtained phylogeny was discussed.

Materials and methods

DNA preparation from selected Actinoplanes strains

DNA from 21 Actinoplanes strains preserved at NITE Biological Resource Center (NBRC) was extracted for amplification and sequencing of the flagellin gene (Table 1). All of the tested strains were grown in YG broth (yeast extract, 10 g L\(^{-1}\); glucose 10 g L\(^{-1}\); pH 7.0) for 7 days at 30 °C. Cells were recovered by centrifugation (1600 g, 10 min) and washed twice with 0.5 M EDTA. Genomic DNA was extracted as described by Saito & Miura (1963) with minor modifications. Isolated DNAs were stored at −20 °C until analysis.

PCR amplification and sequencing of flagellin gene

To amplify the flagellin gene from Actinoplanes strains, the degenerate PCR primers 5F_Fla (5′-GTC TYC GCA TCA ACC AGA ACA TCG-3′) and 1219R_Fla (5′-GCA CGC CCT GCG RGG MCT GGT TCG CG-3′), corresponding to N- and C-terminal regions of the flagellin gene, respectively, were used. The primers were designed by comparing flagellin gene sequences derived from the genome sequence of Actinoplanes missouriensis NBRC 102363\(^T\) (AB600179), Nocardioidea sp. JS614 (CP000509 REGION: 814334..815251), and K. radiotolerans SR30216 (CP000750 REGION: 579397..581526). By mining the genome data of these species, the flagellin gene was only present in the genome as a single copy. Incidentally, the full-length sequence of the flagellin gene from A. missouriensis was determined by the A. missouriensis-sequencing team at the National Institute of Technology and Evaluation (NITE) and other research groups (the entire genome sequence will be published elsewhere).

The reaction mixture (50 μL) for amplification contained 0.5 μL of GC Buffer I (Takara Bio, Shiga, Japan), 2.5 mM of each dNTP, 0.2 μM of each of the two primers designed in this study, 100 ng of genomic DNA, and 1 U of Blend Taq polymerase (Toyobo, Osaka, Japan). Amplification was performed using a thermal cycler (TP600, Takara Bio) with an initial denaturation step of 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 1 min, and extension at 72 °C for 1.5 min. A final extension step was performed at 72 °C for 5 min before the temperature was cooled to 4 °C. PCR products were separated using horizontal gel electrophoresis on a 1% (w/v) Seakem GTG agarose gel (FMC Bioproducts, Rockland, Maine) containing 0.5 μg mL\(^{-1}\) ethidium bromide. Amplicon size was estimated by comparison with a 100 bp DNA size marker (Toyobo, Osaka, Japan).

PCR amplicons were purified using a MonoFas DNA purification kit (GL Sciences, Tokyo, Japan) and directly sequenced using an ABI Prism BigDye Terminator cycle sequencing kit (PE Applied Biosystems, Foster City, CA) and an automatic DNA sequencer (model 3730 Genetic Analyzer; PE Applied Biosystems) Sequencing primers 5F_Fla, 1219R_Fla, 226F_Fla (5′-CAG ACC GGT GAR GGT GCG-3′), and 1056R_Fla (5′-GGT GTG CTC GAA GGT GCG-3′).
MCG GTT CTG-3) were used, and the obtained flagellin gene sequences were registered in the DDBJ database under accession numbers AB640605 to AB640620.

**Homology modeling of flagellin structure**

The three-dimensional structure of flagellin was predicted using the SWISS-MODEL server (http://swissmodel.expasy.org/) (Schwede et al., 2003). The crystal structure of the L-type straight flagellar protein (PDB ID Code: 3a5x) was selected for use as the template structure, which showed amino acid sequence identities of 34% and 42% when compared with *A. missouriensis* and *Actinoplanes lobatus*, respectively. The structures were generated using PyMOL 0.99rc6 (http://pymol.sourceforge.net/).

**Characterization and phylogenetic analysis of flagellin gene sequences**

The flagellin gene sequences from 17 *Actinoplanes* species were translated into amino acid sequences using the European Bioinformatics Institute’s (EMBL-EBI) EMBoss ‘transeq’ program (http://www.ebi.ac.uk/Tools/emboss/transeq/index.html). These amino acid sequences were then aligned with known flagellin sequences stored in public databases using Clustal_W (Thompson et al., 1994). The number of gaps located in the central region of the flagellin sequences was identified by pairwise alignment with the flagellin sequence of *A. missouriensis*; gaps were counted manually. A highly conserved region (115 aa) located in the N-terminal region of the flagellin protein was used for further phylogenetic analysis. Phylogenetic tree generated using flagellin amino acid sequence data was constructed for 18 *Actinoplanes* spp., *K. radiotolerans* SRS30216 (YP_001361376), and *Nocardioides* sp. JS614 (YP_921978) using the maximum parsimony method implemented in the MEGA software package (Molecular Evolutionary Genetics Analysis) version 4 (Tamura et al., 2007). The resultant topologies were evaluated using bootstrap analysis (Felsenstein, 1985) with 1000 resamplings.

**Results**

**Flagellin gene polymorphisms**

The flagellin genes of 21 *Actinoplanes* strains were amplified and classified into two groups based on amplicon
size. Large PCR products were c. 1.2 kbp, and smaller products were c. 0.8 kbp. Most of the *Actinoplanes* strains, 17 of 21, had the larger flagellin, whereas the remaining four *Actinoplanes* strains had the smaller flagellin (Table 1). In this study, these two flagellin genes were referred to as type I (large amplicon) and type II (small amplicon). The PCR amplicons of all of the assayed *Actinoplanes* strains were directly sequenced, which yielded sequences from 17 strains that were of sufficient length. These sequences were aligned to identify gaps between sequences from 17 strains that were of sufficient length. An alignment revealed that the number of gaps was 414. These sequences were also found in *A. auranticolor*, *A. capillaceus*, and *A. lobatus*. The number of gaps was 414.

Homology modeling of flagellin from *Actinoplanes* strains

A flagellin protein model was constructed using the automatic homology modeling server SWISS-MODEL. The amino acid sequences of *A. missouriensis* and *A. lobatus* were considered to be representative of type I and II flagellins. These models of flagellin were constructed using the coordinates of the crystal structure of the L-type straight flagellar protein from *S. typhimurium* (PDB ID Code: 3a5x), which has a sequence identity with the representative type I and II flagellins of 34% and 43%, respectively. The three-dimensional structure model was successfully constructed for the type I and II flagellins in the two *Actinoplanes* strains (Fig. 2). Each of the predicted flagellin structures in the inner (D0 domain) and outer (D1 domain) core regions were well conserved. However, the D2 and D3 domains that form a knob-like projection on the surface of the flagellum are relatively quite different in terms of structure. According to the structural model of type I flagellin, the knob-like projection appeared to consist of four α-helixes and a double-stranded β-sheet, and had a total amino acid residue number of 151. The model of the type II flagellin was characterized as having a compact structure without a D3 domain, with only 26 amino acid residues in the D2/D3 domain. In addition, the number of solvent-exposed hydrophobic amino acids corresponding to the knob-like projection in the types I flagellin was 57 aa, and also the type II flagellin was 13 aa.

![Fig. 1. Comparison of flagellin amino acid sequence of *Actinoplanes missouriensis* NBRC 102363\(^T\) (AB600179) and *A. lobatus* NBRC 12513\(^T\) (AB640617). Identical and similar amino acids are shown by dark boxes. Gaps are shown as dashes.](image-url)
Phylogenetic analysis based on the conserved N-terminal region of flagellin

The phylogenetic tree generated based on the N-terminal flagellin amino acid sequences (115 aa) showed that almost all of Actinoplanes species could be divided into three subgroups (Fig. 3). Subgroup A consisted of six strains with type I flagellin amino acid sequences that had sequence similarities of 80.8–89.5%. The highest sequence similarity (89.5%) was observed between Actinoplanes liguriensis NBRC 13997T, Actinoplanes deccanensis NBRC 13994T, and Actinoplanes grobisporus NBRC 13912T. Subgroup B consisted of Actinoplanes consettensis NBRC 14913T and Actinoplanes humidus NBRC 14915T, which shared 100% similarities in flagellin amino acid sequences. On the other hand, subgroup C consisted of five type I flagellin sequences and three type II flagellin sequences, with similarity values in the range of 76.6–100%. Subgroup C contained sequences that were identical to those of Actinoplanes digitatis NBRC 12512T and A. missouriensis NBRC 102363T. Three of the Actinoplanes strains with the type II flagellin were phylogenetically closely related, with sequence similarity values in the range of 86.9–98.2%. However, A. auranticolor did not cluster with the other Actinoplanes species.

Discussion

Analysis of flagellin gene sequences

In this study, we developed new degenerate primers for assaying three phylogenetically distinct taxa belonging to order Actinomycetales. The primers successfully amplified the flagellin gene sequences of 21 Actinoplanes strains, as well as the flagellin gene sequences of other motile actinomycete strains (data not shown). Two flagellin gene polymorphisms were observed among the Actinoplanes species assayed; one of the PCR products was c. 1.2 kbp (type I), and other is c. 0.8 kbp (type II). The difference between type I and II flagellin gene sequences was revealed by alignment of nucleotide/amino acid sequences containing a large number of gaps in the central region of the sequence. Previously, Vesselinova & Ensign (1996) reported that Actinoplanes rectilineatus and Ampullariella pekinensis (currently Actinoplanes capilaceus) had distinct types of flagellin protein with molecular masses of 42 and 44 kDa.
32 kDa, respectively. The results of our study revealed that the size of the flagellin gene was correlated with the molecular mass of flagellin protein. Similarly, Hales et al. (1998) reported the existence of two differently sized flagellin gene sequences (1.4 and 1.0 kbp) in strains of B. cepacia. These authors also revealed that the strains containing the larger flagellin gene had flagella with relatively greater diameters.

**Structural prediction using homology modeling**

The three-dimensional structures of the type I and II flagellins in the Actinoplanes strains were predicted using SWISS-MODEL with a template structure (PDB ID Code: 3a5x). Interestingly, each of the structures differs most markedly at the region of the knob-like projection, which has been reported to contribute to stabilizing the conformation of the flagellin protein and flagellum fibers (Malapaka et al., 2007). In addition, the type II flagellin had a relatively compact structure that may decrease the mechanical stability of parts of flagellum. This study only observed the type II flagellin in four of the Actinoplanes strains tested, all of which are considered to be motile actinomycetes (Stackebrandt & Kroppenstedt, 1987; Matsumoto et al., 2000). In contrast, the structure of the inner and outer core regions were well conserved in both flagellin proteins. These conserved structures contained the N- and C-terminal regions.

**Flagellin-based phylogenetic analysis**

The phylogenetic analysis was performed using the N-terminal amino acid sequences (115 aa) of flagellin from 17 Actinoplanes species (Fig. 3). The phylogenetic tree showed that the Actinoplanes species formed three clusters, and that some of these relationships were well correlated with analyses obtained using 16S rRNA data. A. consetten-sis and A. humidus both had highly conserved N-terminal flagellin sequences (aa similarity of 99.3%), and the central and C-terminal regions were also similar. Furthermore, these two strains showed 100% similarity in the 16S rRNA analysis. On the other hand, both species were well characterized by a numerical taxonomical approach, and were established as distinct species without genotypic data. Until now, a polyphasic taxonomic approach including genotypic analysis is believed to determine the taxonomic position of prokaryotic microorganisms. Therefore, these two species should be clarifying the taxonomic position using genotypic characterization. The observation that A. auranticolor did not cluster with the other members of Actinoplanes in this study may have arisen due to horizontal transfer (Wassenaar et al., 1995; Liu & Ochman, 2007). The differences in the size of flagellin gene, and those revealed in the phylogenetic analysis imply that the N-terminal region of the flagellin gene was useful for phylogeny or evolutionary study in the genus Actinoplanes. However, it is not yet known why organisms with the type II flagellin have lost the knob-like projection on the flagella. In addition, further flagellin gene sequencing of non-Actinoplanes species are required to discuss the evolutionary distribution of flagellar gene in actinomycetes.

In conclusion, the amplification of flagellin genes using the PCR primers designed in this study was successful for all 21 of the Actinoplanes strains assayed. The resulting PCR amplicons consisted of two types, differing according to size. Comparative sequence analysis and structural prediction of the flagellin amino acid sequences revealed the presence of numerous large gaps in the D2/D3 domains, which located in flagellum surface. Phylogenetic analysis using partial N-terminal flagellin sequences revealed that the Actinoplanes species grouped into three subclusters. The diversity of flagellin gene provides us useful information to discuss the evolution of motile actinomycetes.

**Acknowledgement**

This study was supported in part by a research grant from the Institute for Fermentation, Osaka (IFO).

**References**


