African Journal of Microbiology Research Vol. 7(8), pp. 701-703, 19 February, 2013
Available online at http://www.academicjournals.org/AJMR
DOI: 10.5897/AJMR12.2159
ISSN 1996-0808 ©2013 Academic Journals

Short Communication

Flagellin complete sequence as an inter-specific molecular phylogenetic marker among bacteria

Ernesto Borrayo[1,2], M. D. Martinez-Pena[1], J.A. Morales-Valencia[2], M. Cortes-Cruz[1] and R. Arteaga-Garibay[1]*

[1]Centro Nacional de Recursos Genéticos, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias.
[2]Departamento de Ciencias Computacionales, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara.

Accepted 15 February, 2013

Flagellin, a flagelar subunit protein, presents four structural domains, two of which are highly conserved among bacteria, while the other two tend to be variable. Flagellins have been used as phylogenetic molecular markers, mainly intra-specifically. In this work, a marker comparison for bacterial phylogenetic reconstruction was performed by means of five different DNA sequences: the conserved and complete sequence of flagellins, standard 16S rRNA and recA molecular markers, as well as coding S16 sequence. Results showed that flagellin complete sequence may be an accurate phylogenetic marker for phylogenetic reconstruction among bacteria.

Key words: Flagellin, molecular phylogenetic marker, tree comparison.

INTRODUCTION

Bacterial flagellum is an extracellular organelle that provides mobility to bacteria. In addition to cellular mobility, flagella are related to other biological functions of bacteria, such as adherence, colonization and virulence. The bacterial flagellar filament is a tubular structure composed by several thousand subunits of a protein called flagellin, which are translocated to the central channel of the filament. The bacterial flagellin molecule presents four structural domains called D0, D1, D2 and D3. The chain starts at the N-terminal portion included in the D0 domain and continues with domains D1 to D3, then the chain returns in opposite direction and include the C-terminal within the D0 domain. Approximately, 50 amino acids in both N- and C-terminals constituting the D0 domain are rich in hydrophobic residues, allowing such domain to interact hydrophobically with other flagellin subunits in the filament. These structural characteristics are present in most bacterial flagellins (Samatey et al., 2001; Yonekura et al., 2003).

Hypervariable region of the flagellar protein (Beatson et al., 2006; Schoenhals et al., 1993), domains D2 and D3, constitute the structural elements in contact with the extracellular environment (Beatson et al., 2006, and gives rise to serotype-specific epitopes (Schoenhals et al., 1993).

This fact has made coding regions of flagellins to been targeted as phylogenetic markers, especially in intra-species analysis (Hales et al., 1998; Overly, 2003; Schoenhals et al., 1993; Sun et al., 2006), becoming popular proper markers to identify individuals within closely related organisms, yet their effectiveness as inter-species marker remains unexplored.

The presence of two different evolutionary rates -one highly conserved and the other highly variable- among coding regions of flagellins, leads to the hypothesis that these elements constitute a potential tool for phylogenetic analysis, not only as an intraspecies-level marker, but also as an accurate interspecies one.

This work presents a preliminary study where clusters obtained by traditional bacterial phylogenies markers are

*Corresponding author. E-mail: arteaga.ramon@inifap.gob.mx. Tel: +52 (378) 106-5020. Ext. 5108.

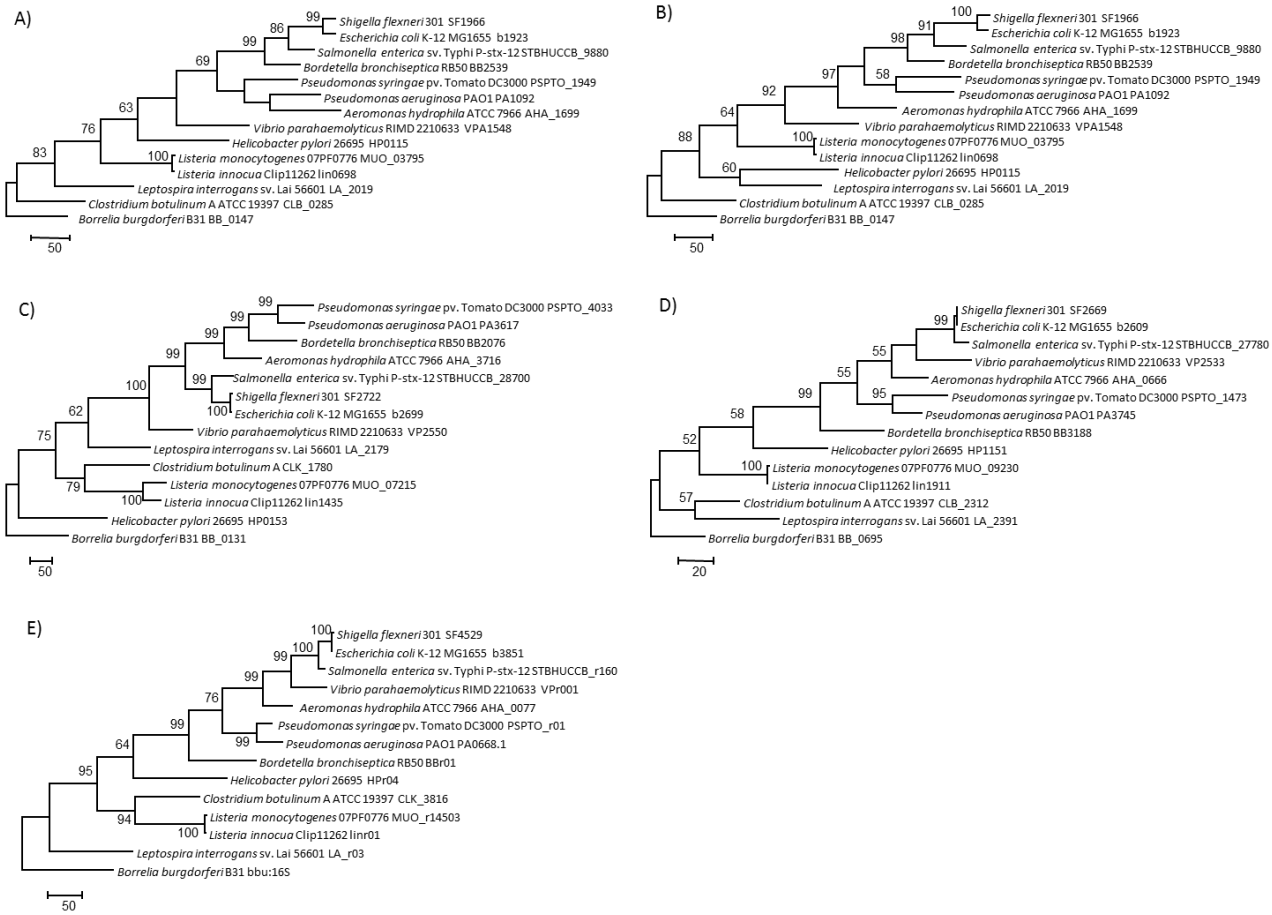

Figure 1. Calculated phylogenetic trees. The phylogenetic reconstruction with analyzed methods: A) flagellin conserved N-terminal domain, B) flagellin complete sequence, C) recA sequence, D) S16 coding sequence, and E) 16S rRNA sequence. The trees were generated using the Neighbour-Joining method and Tajima-Nei substitution matrix. The branch lengths are proportional to the evolutionary distance, and the numbers shown at the branch points indicate the bootstrap values above 50. The data set was subjected to 10,000 bootstrap replicates. Sequences from members of the bacterial genus isolated from gastrointestinal tract. The reference sequences were obtained from the KEEG database. The accession numbers are detailed next to each organism name.

compared to those obtained by flagellin phylogenies analysis of same taxa in representative elements of bacterial groups.

MATERIALS AND METHODS

The phylogenetic analyses were performed using four phylogenetic markers: 16S rRNA, recA, S16 protein and flagellin. Nucleotide sequences of the four markers used in this study were retrieved from the Kyoto Encyclopedia of Genes and Genomes database (http://www.genome.jp/kegg). The selection of bacterial sequences was representative groups of bacterial also tightly related to the relevant biological role of flagella as virulence factors and/or environmental adaptation. A selection was made for sequences of each molecular marker for the fourteen bacteria analyzed in this study. Figure 1 summarizes sequences accession numbers and microorganisms employed in this work.

The sequences were aligned by the multialign tool with default parameters in MATLAB® R2010b, according to description of the MATLAB® Bioinformatics Toolbox® v3.6 (http://www.mathworks.com). Resulting alignments were hand curated and then phylogenetic reconstruction was performed. Neighbour-Joining based phylograms were constructed by MEGA v5.0 (Tamura et al., 2011) with Tajima-Nei substitution matrix and default parameters except for the inclusion of 10,000 bootstraps.

Resulting trees were compared by a Cluster Overlapping comparison. Initially based on the splitting method by Robinson and Foulds (Robinson and Foulds, 1981; Farris, 1989), the method was modified and implemented to be able to compare phylograms regardless of their scale (Borrayo et al., submitted 2012). Cluster overlapping similarity score is calculated by the arithmetic mean of all leaf occurrences for every branch. When comparing a tree against itself, the score obtained is 100%.

RESULTS AND DISCUSSION

To determine the efficiency of the flagellin sequence as a phylogenetic marker, an analysis of 14-selected organisms

Table 1. Comparison of phylogenetic trees.

	Flagellin C- (%)	**Flagelin N- (**%)	**RecA (**%)	**S16 (**%)	**16S rRNA (**%)
Flagellin C-	100	98.75	94.70	96.54	96.00
Flagelin N-	-	100	96.18	96.81	96.36
RecA	-	-	100	95.08	95.90
S16	-	-	-	100	99.11

Cluster overlapping score is presented in similarity percentage. Flagellin C, Flagellin complete sequence; Flagellin N, Flagellin N-terminal conserved domain.

was performed. The aim was to establish an accurate phylogenetic tree with well-defined known clusters, where five different phylogenetic trees were computed. The first two trees are the result of the widely employed 16S rRNA and recA sequences, respectively. The third tree is the product of the *tmrD* protein-coding-sequence for S16, a single strain DNase that has been reported as essential for ribosomal assembly. The fourth and fifth trees are the result of the experimental flagellin sequences, where one is the product of the N-terminal conserved region while the other, is the product of the complete flagellin sequence analysis. The five resulting trees are presented in Figure 1. The comparison results are presented in Table 1. The comparisons showed similar clustering for all trees, all being above 94.5%.

As expected, the compared trees with the best cluster overlapping score were the N-terminal and complete sequence of flagellin, as the latter follows the first, as well as S16 and 16S rRNA, both highly conserved 30S ribosomal components with similar evolutionary rate.

The conserved flagellin analysis provides accurate clustering of Proteobacteria, with efficient differentiation of this clade from the Spirochaetes and Firmicutes. The complete sequence analysis of flagellin did not adequately cluster ε-proteobacteria (Bern and Goldberg, 2005). Both conserved and complete approaches of flagellins had inaccurate clustering of Firmicutes and Spirochaetes.

Despite the fact that phylogenetic reconstruction was not exactly the same as the golden standard 16S rRNA, flagellin analysis provided a closer tree to 16S rRNA and S16 than the widely used recA.

Until now, phylogenetic reconstruction based on flagellin coding sequence was limited to intra-specific analysis, were its highly variable regions have proven effective for bacterial serotype discrimination, but the use of this sequence as a phylogenetic marker in distant related groups remained unexplored.

The results suggested that flagellins may be used as phylogenetic markers in an inter-species approach, not only by its N-terminal conserved domains, but also by a complete sequence analysis.

The implementation of flagellin sequence as a bacterial phylogenetic marker might provide a protein-coding marker unique to bacteria, which would have both highly variable and highly conserved regions, allowing different simultaneous molecular clock analysis.

A flagellin complete sequence approach may provide accurate taxonomical classification when bacterial organisms: both closely and distantly related, are compared, in which case flagellin would prove a wide-range marker for bacterial phylogenetic reconstruction.

REFERENCES

Beatson SA, Minamino T, Pallen MJ (2006). Variation in bacterial flagellins: from sequence to structure. Trends Microbiol. 14(4):151-155.

Bern M, Goldberg D (2005). Automatic selection of representative proteins for bacterial phylogeny. BMC Evol. Biol. 5:34.

Farris JS (1989). The retention Index and the rescaled Consistency Index. Cladistics 5(4):417-419.

Hales BA, Morgan JA, Hart CA, Winstanley C (1998). Variation in flagellin genes and proteins of *Burkholderiacepacia*. J. Bacteriol. 180(5):1110-1118.

Overly K (2003). Phylogenetic analysis of *Borrelia* species detected in small mammals and ticks in Florida. Osprey J. Ideas Inquiry 3:38-47.

Robinson DF, Foulds LR (1981). Comparison of phylogenetic trees. Math. Biosci. 53(1-2):131-147.

Samatey FA, Imada K, Nagashima S, Vonderviszt F, Kumasaka T, Yamamoto M, Namba Kl (2001). Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. Nature 410(6826):331-337.

Schoenhals G, Whitfield C (1993). Comparative analysis of flagellin sequences from *Escherichia coli* strains possessing serologically distinct flagellar filaments with a shared complex surface pattern. J. Bacteriol. 175(17):5395-5402.

Sun W, Dunning FM, Pfund C, Weingarten R, Bent AF (2006). Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of *Arabidopsis* FLAGELLIN SENSING2-dependent defenses. Plant Cell 18(3):764-779.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28(10):2731-2739.

Yonekura K, Maki-Yonekura S, Namba K (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature. 424(6949):643-650.