Anaerobe 28 (2014) 182–198

Contents lists available at ScienceDirect

Anaerobe

journal homepage: www.elsevier.com/locate/anaerobe

Phylogeny and molecular signatures for the phylum Fusobacteria and its distinct subclades

Radhey S. Gupta*, Mohit Sethi

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada

ARTICLE INFO

Article history:
Received 12 March 2014
Received in revised form
10 June 2014
Accepted 12 June 2014
Available online 23 June 2014

Keywords:
Fusobacterium
Fusobacteriaceae
Leptotrichiaceae
Phylogenetic trees
Conserved signature indels
Molecular signatures

ABSTRACT

The members of the phylum Fusobacteria and its two families, *Fusobacteriaceae* and *Leptotrichiaceae*, are distinguished at present mainly on the basis of their branching in the 16S rRNA gene trees and analysis of the internal transcribed spacer sequences in the 16S-23S rDNA. However, no biochemical or molecular characteristics are known that are uniquely shared by all of most members of these groups of bacteria. We report here detailed phylogenetic and comparative analyses on 45 sequenced Fusobacteria genomes to examine their evolutionary relationships and to identify molecular markers that are specific for the members of this phylum. In phylogenetic trees based on 16S rRNA gene sequences or concatenated sequences for 17 conserved proteins, members of the families *Fusobacteriaceae* and *Leptotrichiaceae* formed strongly supported clades and were clearly distinguished. In these trees, the species from the genus *Fusobacterium* also formed a number of well-supported clades. In parallel, comparative analyses on Fusobacteria genomes have identified 44 conserved signature indels (CSIs) in proteins involved in a broad range of functions that are either specific for the phylum Fusobacteria or a number of distinct subclades within this phylum. Seven of these CSIs in important proteins are uniquely present in the protein homologs of all sequenced Fusobacteria and they provide potential molecular markers for this phylum. Six and three other CSIs in other protein sequences are specific for members of the families *Fusobacteriaceae* and *Leptotrichiaceae*, respectively, and they provide novel molecular means for distinguishing members of these two families. Fourteen additional CSIs in different proteins, which are specific for either members of the genera *Fusobacterium* or *Leptotrichia*, or a number of other well-supported clades of Fusobacteria at multiple phylogenetic levels, provide molecular markers for these groups and information regarding the evolutionary relationships among the members of this phylum. Lastly, the present work has also identified 14 CSIs in divergent proteins that are specific for three specific subclades of *Fusobacterium* species, which are also indicated to be distinct by phylogenetic analyses. The members of these three *Fusobacterium* subclades also differ significantly from each other in their whole genome average nucleotide identities values, suggesting that they are possible candidates for recognition as different genera. The molecular markers reported here provide novel means for the identification of members of the phylum Fusobacteria and for their classification.

1. Introduction

The phylum Fusobacteria is made up of Gram-negative, nonmotile, facultative aerobic to obligately anaerobic, fermentative, rod-shaped bacteria, which have generally fusiform (spindle-shaped) morphology [15,26,71]. Although members of the phylum Fusobacteria have been known for more than 100 years [17,22,38,47,55,75], they were only recently, on the basis of their branching in the 16S rRNA trees, grouped into a distinct phylum [71,72]. The bacteria from this phylum are commonly associated with the mucous membrane of humans and animals and they are significant constituents of the human oral cavity, playing an important role in periodontal disease [6,13,16,26,43]. They are also commonly present in the human and animal gastrointestinal tract, in the periurethral region and genitalia of women, and in necrotic lesions [15,16,26,39]. Additionally, several members from the genus *Fusobacterium*, particularly *Fusobacterium necrophorum*, play a central role in the causation of Lemierre's syndrome [45,51] and they are also enriched in the rectal mucosa of patients with colon carcinoma [36,49].

* Corresponding author. Tel.: +1 905 525 9140x22639; fax: +1 905 522 9033.
E-mail address: gupta@mcmaster.ca (R.S. Gupta).

http://dx.doi.org/10.1016/j.anaerobe.2014.06.007
1075-9964/

The phylum Fusobacteria contains a single class (*Fusobacteriia*) and a single order (*Fusobacteriales*) that is made up of two families, *Fusobacteriaceae* and *Leptotrichiaceae*, encompassing nine genera [70,71]. The members of this phylum and its two families are at presently distinguished from other bacteria primarily on the basis of their branching in the 16S rRNA gene trees and analysis of the internal transcribed spacer regions in the 16S-23S rDNA [11,71]. However, there is no biochemical or molecular characteristic known at present that is uniquely shared by all of most members of the phylum Fusobacteria, or its two families, which independently supports that the members of these groups are specifically related. In view of the clinical importance of the members of the phylum Fusobacteria, further studies on discovering characteristics that can more reliably distinguish these bacteria are of much interest.

Genome sequences provide a rich resource for the discovery of molecular markers (or signatures) that are specific for different groups and provide novel and independent means for demarcation of diverse bacterial taxa in more definitive molecular terms [5,19,34,61,65]. Of the molecular markers that are being identified, conserved signature inserts or deletions (i.e. indels) (CSIs) in protein sequences provide an important class of markers for identification of different groups of bacteria in molecular terms [2,3,4,60]. In recent years, genome sequences for large numbers of species from the phylum Fusobacteria have become publicly available (Table 1) [37,40,43,44,56,57,58,66]. Limited analyses of these genomes carried out thus far have not identified any characteristic that is unique to either all Fusobacteria or members of its different families/taxa [13,44,54].

In this work we have carried out detailed phylogenetic and comparative analyses on protein sequences from available Fusobacteria genomes to understand their evolutionary relationships and to identify molecular markers that are either specific for this

Table 1
Genome characteristics of the sequenced members of the phylum Fusobacteria.

Organism	Accession number	Size (Mb)	GC%	Genome source
Cetobacterium somerae WAL 14325	AXZF00000000	3.07	28.60%	Washington University
Fusobacterium gonidiaformans 3-1-5R	ACDD00000000	1.93	32.30%	Broad Institute[a]
Fusobacterium gonidiaformans ATCC 25563	ACET00000000	1.68	32.50%	Broad Institute[a]
Fusobacterium mortiferum ATCC 9817	ACDB00000000	2.67	29.10%	Broad Institute[a]
Fusobacterium necrophorum D12	ACDG00000000	1.96	35.20%	Broad Institute[a]
F. necrophorum subsp. *funduliforme* 1_1_36S	ADLZ00000000	2.3	34.60%	Broad Institute[a]
F. necrophorum subsp. *funduliforme* ATCC 51357	AJSY00000000	2.11	34.90%	JCVI[b]
F. necrophorum subsp. *funduliforme* Fnf 1007	ALKK00000000	2.17	35%	JCVI[b]
F. nucleatum subsp. *animalis* 11_3_2	ACUO00000000	2.72	26.90%	Broad Institute[a]
F. nucleatum subsp. *animalis* 21_1A	NZ_ADEE00000000.2	2.13	27.00%	Broad Institute[a]
F. nucleatum subsp. *animalis* 3_1_33	ACQE00000000	2.3	27.00%	Broad Institute[a]
F. nucleatum subsp. *animalis* 4_8	NC_021281.1	2.28	27.09%	Broad Institute[a]
F. nucleatum subsp. *animalis* 7_1	AKBT00000000	2.51	26.99%	[58]
F. nucleatum subsp. *animalis* ATCC 51191	AFQD00000000	2.27	27.50%	Baylor College[c]
F. nucleatum subsp. *animalis* D11	ACDS00000000	2.45	26.90%	Broad Institute[a]
Fusobacterium nucleatum subsp. *animalis* F0419	AGEH00000000	2.41	26.80%	Broad Institute[a]
F. nucleatum subsp. *fusiforme* ATCC 51190	AKXI00000000	1.84	27.20%	[57]
F. nucleatum subsp. *nucleatum* ATCC 23726	ADVK00000000	2.24	27%	Baylor College[c]
F. nucleatum subsp. *polymorphum* ATCC 10953	AARG00000000	2.44	26.79%	Baylor College[c]
F. nucleatum subsp. *polymorphum* F0401	ADDB00000000	2.46	26.80%	Broad Institute[a]
F. nucleatum subsp. *vincentii* 3_1_27	CP007064.1	2.19	27.08%	Broad Institute[a]
F. nucleatum subsp. *vincentii* 3_1_36A2	NC_022196	2.27	27.10%	Broad Institute[a]
F. nucleatum subsp. *vincentii* 4_1_13	ACDE00000000	2.27	26.80%	Broad Institute[a]
F. nucleatum subsp. *vincentii* ATCC 49256	AABF00000000	2.12	27.30%	[44]
Fusobacterium periodonticum 1_1_41FAA	ADGG00000000	2.48	27.90%	Broad Institute[a]
Fusobacterium periodonticum 2_1_31	ACDC00000000	2.48	28.00%	Broad Institute[a]
Fusobacterium periodonticum ATCC 33693	ACJY00000000	2.57	27.40%	WU-GSC[d]
Fusobacterium periodonticum D10	ACIF00000000	2.43	27.80%	Broad Institute[a]
Fusobacterium russii ATCC 25533	ARMK00000000	1.94	28.60%	DOE
Fusobacterium ulcerans 12-1B	AGWJ00000000	3.73	30.30%	Broad Institute[a]
Fusobacterium ulcerans ATCC 49185	ACDH00000000	3.44	30.30%	Broad Institute[a]
Fusobacterium varium ATCC 27725	ACIE00000000	3.3	29.20%	Broad Institute[a]
F. nucleatum subsp. *nucleatum* ATCC 25586	NC_003454	2.17	27.20%	[43]
Ilyobacter polytropus strain CuHbu1	NC_014632	3.13	34.36%	[66]
Leptotrichia hofstadii F0254	ACVB00000000	2.45	28.46%	WU-GSC[d]
Leptotrichia buccalis type strain C-1013-b	NC_013192	2.47	29.60%	[40]
Leptotrichia goodfellowii F0264	ADAD00000000	2.29	31.50%	JCVI[b]
Leptotrichia shahii DSM 19757	ARDD01000000	2.14	29.50%	DOE
Leptotrichia sp. oral taxon 215 str. W9775	AWVR00000000	2.31	31.40%	Washington University
Leptotrichia sp. oral taxon 879 str. F0557	AWVL00000000	2.42	29.60%	Washington University
Leptotrichia trevisanii DSM 22070	AXVL00000000	2.85	30.30%	DOE
Leptotrichia wadei DSM 19758	ARDS00000000	2.32	29.30%	DOE
Psychrilyobacter atlanticus HAW-EB21	AUFS00000000	3.54	31.90%	DOE
Sebaldella termitidis ATCC 33386	NC_013517	4.49	33.42%	[37]
Streptobacillus moniliformis DSM 12112	NC_013515	1.67	26.27%	[56]

DOE Joint Genome Institute (DOE).
The Genome Institute, Washington University School of Medicine (Washington University).
[a] The Broad Institute Genome Sequencing Platform (Broad Institute).
[b] The J. CraigVenter Institute (JCVI).
[c] Human Genome Sequencing Center, Baylor College of Medicine (Baylor College).
[d] Genome Sequencing Center, Washington University School of Medicine (WU-GSC).