Systematic and Applied Microbiology 36 (2013) 376–382

Contents lists available at SciVerse ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

A phylogenetic analysis of the phylum *Fibrobacteres*

Kelsea A. Jewell, Jarrod J. Scott[1], Sandra M. Adams, Garret Suen*

Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706, United States

ARTICLE INFO

Article history:
Received 10 December 2012
Received in revised form 3 April 2013
Accepted 19 April 2013

Keywords:
Fibrobacteres
Fibrobacter
Phylogeny
Host range
Rumen
Phylogeny

ABSTRACT

Members of the phylum *Fibrobacteres* are highly efficient cellulolytic bacteria, best known for their role in rumen function and as potential sources of novel enzymes for bioenergy applications. Despite being key members of ruminants and other digestive microbial communities, our knowledge of this phylum remains incomplete, as much of our understanding is focused on two recognized species, *Fibrobacter succinogenes* and *F. intestinalis*. As a result, we lack insights regarding the environmental niche, host range, and phylogenetic organization of this phylum. Here, we analyzed over 1000 16S rRNA *Fibrobacteres* sequences available from public databases to establish a phylogenetic framework for this phylum. We identify both species- and genus-level clades that are suggestive of previously unknown taxonomic relationships between *Fibrobacteres* in addition to their putative lifestyles as host-associated or free-living. Our results shed light on this poorly understood phylum and will be useful for elucidating the function, distribution, and diversity of these bacteria in their niches.

Introduction

The ability of herbivores to convert plant biomass into usable nutrients is predicated on symbiotic associations with diverse microbial communities [13]. A key example is ruminants, which use a consortium of microbes to degrade and ferment recalcitrant forms of cellulosic biomass into short-chain, host-available volatile fatty acids (for a review, see [49]). Among these important microbes is the bacterium *Fibrobacter succinogenes* [17], a prolific cellulose degrader [4,53] that produces succinic acid as its major fermentation product and lesser amounts of acetic and formic acids. The recently completed genome sequence for the type strain, *F. succinogenes* S85, highlighted the metabolic and cellulolytic specialization of this bacterium in the rumen [45]. *F. succinogenes* belongs to the phylum *Fibrobacteres*, and work within this phylum has focused primarily on this species and *F. intestinalis*, which is typically found associated with non-ruminant mammalian guts [32]. As a result, our knowledge of the *Fibrobacteres*' host range, environmental niche, and species diversity is based almost entirely on *F. succinogenes* and (to a much lesser extent) *F. intestinalis*. A better understanding of this phylum, and its member species, will be useful for guiding research in areas such as agriculture and biofuel production.

The genus *Fibrobacter* was previously classified within the *Bacteroidetes*, but was elevated to its own phylum based on 16S rRNA sequence analysis and physiological differences from other members of the *Bacteroidetes* [32]. The *Fibrobacteres* are currently defined as anaerobic, Gram-negative, non-spore forming, cellulolytic, non-motile rods [32]. However, only two *Fibrobacter* species have been cultured and formally described (*F. succinogenes* and *F. intestinalis*) [32], despite 16S rRNA sequence data that strongly suggests that cryptic species exist [3,18,24,55]. At present, our understanding of *Fibrobacteres* physiology is based entirely upon *F. succinogenes* and *F. intestinalis* isolates associated with animals [18–20,46,55]. These studies do not reflect the wide diversity of 16S rRNA sequences identified as *Fibrobacteres* that have been reported from surveys of environments as disparate as landfills [30], freshwater lakes [29], the ocean [51], limestone cave sulfidic waters [26], termite hindguts [15,52], and a wide variety of animal feces [23,28]. Given the paucity of information available for the *Fibrobacteres*, and the highly desirable cellulolytic properties of its currently described members, it is important to establish a knowledgebase that documents the diversity of species, distribution, and host associations that exist within this phylum.

Here, we capitalize on extensive public databases of existing 16S rRNA sequence data to create and analyze a *Fibrobacteres*-specific phylogeny. This dataset includes many sequences of environmental origin, in addition to host-associated sequences, and we use this data to estimate the diversity present in non-ruminant and non-fecal *Fibrobacteres*. Our analysis reveals a diverse phylogeny that we use to estimate both the species structure and environmental distribution of this poorly understood phylum.

* Corresponding author. Tel.: +1 608 890 3971; fax: +1 608 262 9865.
E-mail address: gsuen@wisc.edu (G. Suen).
[1] Current address: Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, P.O. Box 380, East Boothbay, ME 04544, United States.

0723-2020/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.syapm.2013.04.002

Materials and methods

Sequence data collection, screening, and phylogenetic analysis

All 16S rRNA sequences available through the National Center for Biotechnological Information's (NCBI) nucleotide database marked with the search terms "*Fibrobacteres*," "Fibrobacter," or "Fibrobact*" (where "*" indicates a wild-card search term, accessed: 09/17/2012) were used to construct an initial sequence library. Roche 454-based pyrosequence libraries were not included, as their short average read length complicates the ability to generate usable alignments for downstream analyses [12]. Literature searches were conducted to identify other 16S rRNA sequences likely belonging to the phyla "*Fibrobacteres*" or "*Fibrobacteres*/*Acidobacteria*" but not marked as such in NCBI. The *Fibrobacteres* sequences present in the GreenGenes [11], Ribosomal Database Project [9] and Silva [37] sequence repositories were also included. The total sequence set was annotated to include sequence source and location from both GenBank deposit information and the original publication, if such existed. The complete dataset is presented in Table S1.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.syapm.2013.04.002.

The following phylogenetic analysis was performed on the compiled sequence dataset. All sequences were imported into ARB [25] and a full alignment was created against the current Silva 16S/18S rRNA non-redundant sequence database (SSU Ref NR; release 102; 262,092 total sequences); sequences with closer affinity to known *Fibrobacteres* than to any other phyla were considered as belonging to the phylum *Fibrobacteres*. The *Fibrobacteres*-associated sequences were processed in MOTHUR (v.1.26.0, commands used in the following description denoted in italics) [42]. Sequences >900 bp with 5 or fewer ambiguous nucleotides and nine or fewer homopolymers were retained (*screen.seqs*), duplicate sequences were removed (*unique.seqs*), and a preliminary alignment was created (*align.seqs*, Needleman–Wunsch pairwise alignment method, gap extension penalty = −1, gap opening penalty = −1, and match = +1, mismatch penalty = −1, *k* size = 7). Chimera detection (*chimera.uchime*) used the Silva 16S/18S rRNA non-redundant sequence database (SSU Ref NR, accessed 09/2012). Aligned sequences and were filtered (*filter.seqs*, trump=., vertical = T, soft = 50) and used to create a distance matrix (*dist.seqs*, calc = onegap, cutoff = 0.2). The distance matrix used to calculate the estimated number of operational taxonomic units (OTUs) present at 90%, 95%, and 97% similarity cut-offs (*cluster*, nearest neighbor algorithm). Representative sequences were chosen for each OTU (*get.oturep*) for use in constructing a tree in MrBayes (v3.1) [16,39] (ngen = 10,000,000, chain = 4), with the resulting tree visualized using FigTree (v1.3.1) [38]. In addition, a Neighbor-Joining tree was generated from all sequences [40] using the Maximum Composite Likelihood method [47] in MEGA5 [48] and visualized using the Interactive Tree of Life project [22]. The complete 16S rRNA sequence for *Bacteroides fragilis* NCTC9343 was included in our phylogenetic analyses as an outgroup (GenBank genome accession number: NC_003228.3).

A second Neighbor-Joining tree was constructed for sequences of 450 bp or greater in length using these same methods. Some modifications to the sequence manipulations were required due to this shortened minimal sequence length, and were as follows: no ambiguous nucleotides were allowed, and the alignment filter did not include *trump=*. After alignment and filtering the dataset was used to create a p-distance pair-wise distance matrix in MEGA5, and those sequences for which it was not possible to estimate evolutionary distances were removed.

Results

Distribution and definition of the Fibrobacteres

From an initial database of 1166 putative *Fibrobacteres* sequences we generated a database of 863 confirmed *Fibrobacteres* sequences of >900 bp (henceforth, "long sequence database") and 1095 sequences >450 bp (henceforth, "short sequence database") after all filtering steps were performed (Table S1 and Fig. S1). The 900 bp length cut-off was chosen to retain a maximum number (at least 70%) of the original sequences while minimizing the effects of reduced alignment quality (number of columns removed during alignment filtering). The shorter *Fibrobacteres* database was constructed for comparative purposes. The results presented here, except when stated otherwise, refer to analyses performed using the long sequence database.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.syapm.2013.04.002.

An analysis of our sequences revealed that they were generated from both animal [2–4,15,20,23,24,31,43,52,55] and environmental [5,26,30,36,51] sources, including a wide range of locations spanning the globe (Fig. 1). This range of sampling locations likely reflects the distribution of research groups and interests rather than the dominant natural reservoirs for *Fibrobacteres* bacteria. The two largest groups of sequences belonged to either those associated with mammals (17%, of which 9% were specifically from ruminants) or termites (81%), with the remaining sequences isolated from non host-associated environments (2%) and a single isolate from the water flea *Ventiella sulfuris*. The rumen samples were nearly evenly split between sheep (40 sequences) and cows (32 sequences), with the remaining two sequences from goats. Among the non-ruminants there was a high degree of diversity (Fig. 1 and Table S1), with the most represented non-ruminant animal hosts being Somali wild asses (23 sequences), Yunnan snub-nosed monkeys (12 sequences), and the Eastern black and white colobus (7 sequences). Of the non-host associated environments, 3 were from marine sources, 7 from freshwater sources, 2 from soil, and the remaining 8 from landfills. Among the freshwater sources, surface water, acid-impacted lakes, and sulfidic cave waters were all represented.

Further analysis of the 245 *F. succinogenes* sequences only present in our short sequence database revealed that they were generated primarily from free-living environments, the bulk of which were from landfills (24%), soil (16%), and aquatic (38%) sources. The 43 host-associated short sequences were dominated by rumen (10%) and ostrich cecum (7%) sources. Specific locations unique to the short sequence set included oil sands, yak and buffalo rumens, hot springs, ostrich cecae, and a swine anaerobic lagoon.

An initial Fibrobacteres phylogeny

To gain an initial understanding of the phylogenetic relationship between our identified *Fibrobacteres* sequences, we generated Neighbor-Joining (NJ) trees using the combined long and short sequence databases (Fig. S2). In general, we found that *Fibrobacteres* sequences grouped according to their host or environmental association. For example, the higher and lower termites formed clades distinct from other host-associated sequences, while there was very little mixing of environmental (aquatic or soil) and host-associated sequences within terminal clades. Furthermore, we found a single branch point that includes almost all of the host-associated sequences, with the *F. succinogenes* and *F. intestinalis* sequences forming distinct clades from that common branch point.