Phylogenomic Analyses and Comparative Studies on Genomes of the Bifidobacteriales: Identification of Molecular Signatures Specific for the Order Bifidobacteriales and Its Different Subclades

Grace Zhang¹, Beile Gao², Mobolaji Adeolu¹, Bijendra Khadka¹ and Radhey S. Gupta¹*

¹ Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada, ² CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, Guangdong Key Laboratory of Marine Materia Medica, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

The order Bifidobacteriales comprises a diverse variety of species found in the gastrointestinal tract of humans and other animals, some of which are opportunistic pathogens, whereas a number of others exhibit health-promoting effects. However, currently very few biochemical or molecular characteristics are known which are specific for the order Bifidobacteriales, or specific clades within this order, which distinguish them from other bacteria. This study reports the results of detailed comparative genomic and phylogenetic studies on 62 genome-sequenced species/strains from the order Bifidobacteriales. In a robust phylogenetic tree for the Bifidobacteriales constructed based on 614 core proteins, a number of well-resolved clades were observed including a clade separating the Scardovia-related genera (Scardovia clade) from the genera Bifidobacterium and Gardnerella, as well as a number of previously reported clusters of Bifidobacterium spp. In parallel, our comparative analyses of protein sequences from the Bifidobacteriales genomes have identified numerous molecular markers that are specific for this group of bacteria. Of these markers, 32 conserved signature indels (CSIs) in widely distributed proteins and 10 signature proteins are distinctive characteristics of all sequenced Bifidobacteriales species and provide novel and highly specific means for distinguishing these bacteria. In addition, multiple other molecular signatures are specific for the following clades of Bifidobacteriales: (i) 5 CSIs specific for a clade comprising of the Scardovia-related genera; (ii) 3 CSIs and 2 CSPs specific for a clade consisting of the Bifidobacterium and Gardnerella spp.; (iii) multiple other signatures demarcating a number of clusters of the B. asteroides- and B. longum- related species. The described molecular markers provide novel and reliable means for distinguishing the Bifidobacteriales and a number of their clades in molecular terms and for the classification of these bacteria. The Bifidobacteriales-specific CSIs, found in important proteins, are predicted to play important roles in modifying the cellular functions of the affected proteins. Hence, biochemical studies on the cellular functions of these CSIs
INTRODUCTION

The order Bifidobacteriales contains a large collection of bacterial species, many of which are significant constituents of the gastrointestinal tract of humans, other mammals, birds and honey bees (Biavati et al., 2000; Biavati and Mattarelli, 2006; Turroni et al., 2009, 2011; Biavati, 2012; Milani et al., 2014). In addition to widely recognized health-promoting effects of bifidobacterial species (Leahy et al., 2005; Ventura et al., 2009a; Cronin et al., 2011), some members of the group found in human and animal oral cavities are implicated in the development of dental caries (Huys et al., 2007; Mantzourani et al., 2009; Ventura et al., 2009b). Additionally, Gardnerella vaginalis is indicated to play an important role in the pathogenesis of bacterial vaginosis and urinary tract infections (Smith et al., 1992; Bradshaw et al., 2006; Alves et al., 2014; Kenyon and Osbak, 2014). The order Bifidobacteriales is part of the phylum Actinobacteria (Ventura et al., 2007b; Zhi et al., 2009; Gao and Gupta, 2012) and it harbors a single family, Bifidobacteriaceae, containing >50 recognized species (Biavati, 2012; Lugli et al., 2014; Milani et al., 2014; Parte, 2014) that are grouped into eight genera: Aeriscardovia, Alloscardovia, Bifidobacterium, Gardnerella, Pseudoscardovia, Neoscardovia, Parascardovia, and Scardovia (Jian and Dong, 2002; Simpson et al., 2004; Huys et al., 2007; Biavati and Mattarelli, 2012; García-Aljaro et al., 2012; Storms and Vandamme, 2012; Keller et al., 2013). Of these, the genus Bifidobacterium, encompassing 39 species and 9 subspecies, forms the largest group and accounts for more than 75% of the described taxa within the order Bifidobacteriales (Biavati and Mattarelli, 2012; Milani et al., 2014).

Phylogenetic analyses based on 16S rRNA, as well as sequences for a number of housekeeping genes/proteins, are the main approaches used in the past to distinguish among the Bifidobacteriales species and genera (Miyake et al., 1998; Ventura and Zink, 2003; Ventura et al., 2004, 2006, 2007a; Biavati and Mattarelli, 2006; Yarza et al., 2008; Bottacini et al., 2010; Turroni et al., 2011; Mattarelli et al., 2014). In recent years, complete or draft genome sequences have become available for all currently recognized Bifidobacterium species and subspecies (Ventura et al., 2009b; Milani et al., 2014). Based on these sequences, a panel of multiplex PCR primers has been developed enabling rapid and specific identification of different Bifidobacterium species and subspecies (Ferrario et al., 2015). Based on genome sequences, two recent studies have also examined the evolutionary relationships among Bifidobacterium species employing large datasets of sequences comprising the core proteins of this genus (Lugli et al., 2014; Sun et al., 2015). The robust phylogenetic trees obtained in these studies provide important insights concerning the evolutionary relationships among the Bifidobacterium species and they strongly support the existence of 6–7 distinct clusters within this genus. These clusters are referred to as the B. asteroides, B. pseudolongum, B. longum, B. bifidum, B. adolescentis, B. pullorum, and B. boum groups (Lugli et al., 2014; Sun et al., 2015). Similar clusters are also observed in phylogenetic trees based on the 16S and 23S rRNA genes as well trees based on other gene/protein sequences. Comparative analyses of the Bifidobacteriales genomes are also providing useful insights concerning species-specific characteristics that are suggested to play important roles in the adaptation of particular species to either human or insect gut environment (Ventura et al., 2009b; Bottacini et al., 2010, 2012; Turroni et al., 2010).

Due to the health-promoting effects of bifidobacteria, it is of much interest to identify genetic and biochemical characteristics that are specific for the Bifidobacteriales or particular groups/clusters within this order of bacteria. Currently, very few such characteristics are known. One important class of genome sequence-based molecular markers, which have proven very useful for evolutionary, taxonomic and functional studies are conserved signature insertions or deletions (CSIs) that are uniquely present in the genes/proteins homologs from a defined group of organisms (Gao and Gupta, 2005, 2012; Gupta, 2010, 2014). Conserved signature proteins (CSPs), which are genes/proteins that are uniquely found within a monophyletic group of organisms, provide another class of useful molecular makers for evolutionary and functional studies (Gao et al., 2006; Ventura et al., 2007a; Gao and Gupta, 2012; Gupta, 2016a,b). Both these types of markers constitute highly reliable characteristics of specific groups of organisms and they have been extensively utilized for the identification/demarcation of prokaryotic taxa of different ranks in molecular terms (Gao and Gupta, 2012; Gupta et al., 2013a,b, 2016).

In the present work, we report detailed phylogenetic and comparative analyses on protein sequences from the sequenced members of the order Bifidobacteriales in order to identify CSIs and CSPs that are specific for different groups within this order. These studies have led to identification of 32 CSIs in widely distributed proteins and 10 CSPs that are uniquely found in all or most of the genome sequenced Bifidobacteriales species providing novel molecular markers that distinguish this order from all other bacteria. In addition, our work has also identified multiple other CSIs and CSPs that distinguish a number of clades of Bifidobacteriales, including a clade consisting of the Bifidobacterium and Gardnerella species, another clade consisting of the Scardovia-related genera, and
multiple signatures that are specific for different clusters of *B. asteroides* or *B. longum* related species. These signatures provide novel means for the identification and demarcation of the members of the described clades in molecular terms and for functional studies that could lead to discovery of novel biochemical and/or other novel properties of these bacteria.

**METHODS**

**Phylogenetic Analysis**

A phylogenetic tree for 62 genome-sequenced members from the order *Bifidobacteriales* was constructed based on concatenated sequences of 614 proteins. The protein families used in this phylogeny were identified using the UCLUST algorithm (Edgar, 2010) to identify proteins families present in at least 80% of the input genomes which shared at least 50% sequence identity and 50% sequence length. Each identified protein family was individually aligned using Clustal Omega (Sievers et al., 2011) and trimmed using Gblocks 0.91b (Castresana, 2000) with relaxed parameters (Talavera and Castresana, 2007). The concatenated dataset of the trimmed sequence alignments contained 197,777 aligned amino acid residues. A maximum-likelihood tree based on this alignment was constructed using FastTree 2 (Price et al., 2010) employing the Whelan and Goldman model of protein sequence evolution (Whelan et al., 2001) and RAxML 8 (Stamatakis, 2014) using the Le and Gascuel model of protein sequence evolution (Le and Gascuel, 2008). SH-like statistical support values (Guindon et al., 2010) for each branch node in the final phylogenetic tree were calculated using RAxML 8 (Stamatakis, 2014). This process was completed using an internally developed software pipeline.

In parallel, a phylogenetic tree based on the 16S rRNA gene sequences of type strains covering all described species within the order *Bifidobacteriales* was also constructed. The 16S rRNA sequences were retrieved from Ribosomal Database Project (Cole et al., 2014) and aligned using the SINA aligner (Pruesse et al., 2012) to form a multiple sequence alignment that was 1604 aligned nucleotides long with common gaps removed. A maximum-likelihood phylogenetic tree based on this multiple sequence alignment was created using MEGA 6 employing the General Time-Reversible model of sequence evolution with branch support based on 1000 bootstrap replicates (Tamura et al., 2013).

**Identification of Conserved Signature Indels**

Conserved signature indels (CSIs) were identified by the procedures described in detail recently (Gupta, 2014). Briefly, BLASTp (Altschul et al., 1997) searches were performed on each protein in the genome of *Bifidobacterium adolescentis* ATCC 15703 (Accession number AP009256.1) against all available sequences in the GenBank non-redundant database. Multiple sequence alignments were then created using ClustalX (Jeanmougin et al., 1998) for proteins that returned high scoring matches from *Bifidobacteriales* and other prokaryotes. The alignments were then visually inspected for the presence of insertions or deletions that were flanked on both sides by at least 5-6 conserved amino acid residues in the neighboring 30–40 amino acids. Detailed BLASTp searches were then carried out on short sequence segments containing the indel and the flanking conserved regions (60–100 amino acids long) to determine the specificity of the indels. SIG_CREATE and SIG_STYLE (available on Gleans.net) were then used to create Signature files for CSIs that were specific for the *Bifidobacteriales* order or its subgroups as described in earlier work (Gupta et al., 2013a; Gupta, 2014). Due to space limitations, sequence information for all *Bifidobacterium* species, particularly for different subspecies of *B. longum*, *B. animalis*, *B. pseudolongum*, and *B. thermodacidophilum*, is not shown in the presented alignment files. However, unless otherwise noted, all of the described CSIs are specific for the indicated groups (i.e., similar CSIs were not present in the protein homologs from other bacteria in the top 500 Blast hits).

It should be noted that significant blast hits for a number of CSIs and CSPs described here are also observed for one of the following three *Chlamydia trachomatis* strains (SwabB1, H1 IMS, and H17 IMS) deposited by the Sanger Institute. We suspect that these anomalous results are due to cross contamination of the sequenced cultures from the above *Chlamydia trachomatis* strains by a *Gardnerella vaginalis* strain. We have communicated our concern with the supporting evidence to the Sanger Institute.

**Identification of Conserved Signature Proteins**

BLASTp searches were carried out to examine the specificity of some previously described conserved signature proteins (CSPs), which were indicated to be specific for the order *Bifidobacteriales* (Gao and Gupta, 2012). Additionally, limited work to identify CSPs for the *B. asteroides* group of species was carried out by conducting BLASTp searches on all proteins from the genomes of *Bifidobacterium asteroides* (Bottacini et al., 2012) as query sequences. BLASTp searches were performed against all available sequences in the GenBank non-redundant sequence database and the results of these searches were then manually inspected, as described in earlier work (Gao et al., 2006; Gao and Gupta, 2012), for proteins for which all significant hits were from the *B. asteroides* group of species.

**Homology Modeling of Elongation Factor Tu from *Bifidobacterium longum***

Homology models of EF-Tu homolog from *Bifidobacterium longum* were built using the solved crystallographic structure of EF-Tu from *Escherichia coli* (PDB ID: 3U6K) as the template. Initially, 200 models were generated using MODELER v9.14 (Sali and Blundell, 1993) and ranked/selected using assigned discrete optimized potential (DOPE) scores (Shen and Sali, 2006). The model with the highest DOPE score was then submitted to the ModRefiner program to obtain atomic-level energy minimization and to obtain a model with reliable stereochemistry quality (Xu and Zhang, 2011).
RESULTS
Phylogenetic Analysis of the Species from the Order Bifidobacteriales

Phylogenomic analyses of members of the genus *Bifidobacterium* have been previously reported based on core protein sequences from 45 and 48 described species from this genus (Lugli et al., 2014; Sun et al., 2015). However, these studies did not include the other members of the order Bifidobacteriales such as Gardnerella, Scardavia, Alloscardovia, and Parascardovia, as well as several unnamed *Bifidobacterium* spp. (viz. strains A11, 7101, AGR2158, MSTE12, 12.1.47FAAA) whose genomes have been sequenced. Additionally, the genome sequence of a recently described species *B. aesculapii* is also now available (Toh et al., 2015). To comprehensively examine the evolutionary relationships among different members of the order Bifidobacteriales, a phylogenetic tree was constructed for all 62 genome sequenced members of the family which included 54 *Bifidobacterium* species/strains, 5 species from Scardovia and related genera (viz. Alloscardovia and Parascardovia) and three strains of Gardnerella vaginalis. The tree was constructed based on the concatenated sequences of 614 universally or nearly universally present core proteins for which sequence information could be obtained from the 62 sequenced genomes. A maximum-likelihood tree based on these sequences, which represents the most comprehensive phylogenetic analysis of the order Bifidobacteriales to date, is presented in Figure 1.

In the tree shown, members of the order Bifidobacteriales, at the highest level, form two main clusters. One of these clusters referred to as the Scardovia cluster groups together the genera Scardovia, Parascardovia, and Alloscardovia, whereas the second cluster is comprised of members of the genus Bifidobacterium and Gardnerella. Importantly in this tree, as well as in an earlier study in a phylogenetic tree based on concatenated sequences for RpoB, RpoC, and GyrB proteins, different strains of *Gardnerella vaginalis* were found to branch in between the *Bifidobacterium* species (Gao and Gupta, 2012), making the genus *Bifidobacterium* polyphyletic. Earlier phylogenetic studies on members of the genus *Bifidobacterium* have identified a number of different clusters, which are referred to as the *B. asteroides*, *B. pseudolongum*, *B. longum*, *B. bifidum*, *B. adolescentis*, *B. pullorum*, and *B. boulum* groups (Ventura et al., 2006; Turroni et al., 2011; Lugli et al., 2014; Sun et al., 2015). The existence of these groups/clusters is also confirmed and supported by the tree shown in Figure 1. Of these clusters, the species-related to *B. asteroides* cluster exhibited the deepest branching within the genus *Bifidobacterium*, as also observed in earlier work (Lugli et al., 2014; Sun et al., 2015). The *B. asteroides* clade is generally demarcated as being comprised of the *B. asteroides*, *B. indicum*, *B. coryneforme*, and *B. actinocoloniiforme* species (marked as cluster III in Figure 1). However, as discussed later, a number of clusters, marked I, II, and IV, which are either part of the *B. asteroides* clade or are related to this clade are also distinguished in phylogenetic trees and by the CSIs identified in this work.

We have also created a phylogenetic tree based on 16S rRNA gene sequences of all named *Bifidobacteriales* species (Supplementary Figure S1). The overall branching pattern in the 16S rRNA tree is similar to that observed in the concatenated protein tree with *Scardovia* and related genera forming the deepest branches in the tree and the genera *Scardovia*, *Alloscardovia*, and *Parascardovia* were part of one of the deepest branching clusters. The different clusters of the *Bifidobacterium* spp. that are observed in the concatenated protein tree were also supported by the 16S rRNA tree and *G. vaginalis* was found to branch in between these clusters. The polyphyletic nature of the genus *Bifidobacterium* in 16S rRNA gene based phylogenies is also observed in earlier work (Yilmaz et al., 2014).

Identification of Molecular Markers That Are Specific for the Order Bifidobacteriales

The main focus of this work is the identification of molecular characteristics that are specific for the *Bifidobacteriales* species and could be used for their identification as well as functional studies. As noted earlier, conserved inserts and deletions (i.e., indels or CSIs) in genes/proteins and conserved signature proteins that are uniquely found in a phylogenetically coherent group of organisms provide very useful molecular markers for such purposes. The indels that provide useful molecular markers are of defined size and they are flanked on both sides by conserved regions to ensure that they are reliable characteristics (Gupta, 1998; Gupta and Griffiths, 2002; Ajawatanawong and Baldauf, 2013). These conserved indels in gene/protein sequences result from highly specific and rare genetic changes, hence when such an indel is uniquely found in a phylogenetically coherent group of species, its simplest explanation is that the genetic change responsible for it occurred once in a common ancestor of the indicated group and then the change was passed on to various descendants (Gupta, 1998, 2014; Rokas and Holland, 2000; Gao and Gupta, 2005). Based upon the presence or absence of a conserved indel in outgroup species, it is also possible to determine whether a given indel represents an insert or a deletion (Gupta, 1998; Gao and Gupta, 2012).

Comparative analyses of protein sequence alignments from bifidobacteria species carried out in this work have led to the identification of 32 CSIs in a broad range of highly conserved proteins, which are specifically found in different *Bifidobacteriales* taxa (see Table 1). One example of a CSI that is specific for all members of the order Bifidobacteriales is shown in Figure 2. In this case, a 4 amino acid (aa) insertion is present in a highly conserved region of the protein synthesis elongation factor EF-Tu, which is commonly shared by all sequenced bifidobacteria species, but it is not found in any other bacteria in the top 500 BLAST hits. The protein EF-Tu is a highly conserved protein, which is universally present in all organisms (Harris et al., 2003) and the 4 aa CSI in this protein is a distinctive characteristic of homologs from all sequenced *Bifidobacteriales* species. Sequence information for 31 other CSIs, which are also specifically shared by members of the order Bifidobacteriales, and which are present in proteins involved in different other functions, is provided in Supplementary Figures S2–S32 and some of their
FIGURE 1 | A maximum-likelihood tree based on concatenated sequences of 614 core proteins from 62 sequenced genome-sequenced members of the order Bifidobacteriales. The tree was rooted at the midpoint and SH-like support values are indicated at nodes. A number of different clades/clusters that are consistently observed in phylogenetic trees are marked.
TABLE 1 | Characteristics of conserved signature indels that are Specific for the order Bifidobacteriales.

| Protein name                              | GI number          | Figure no.         | Indel size | Indel region
|-------------------------------------------|--------------------|--------------------|------------|----------------|
| Elongation factor Tu                      | 386068895          | Figure 2           | 4 aa ins   | 106–144        |
| DNA topoisomerase I                       | 489904111          | Supplementary Figure S2 | 1 aa del  | 31–80          |
| DNA polymerase sliding clamp subunit      | 408500301          | Supplementary Figure S3 | 1 aa ins  | 79–118         |
| Beta-galactosidase                        | 504834401          | Supplementary Figure S4 | 1–2 aa ins | 371–423        |
| Ketol-acid reductoisomerase               | 651881972          | Supplementary Figure S5 | 2 aa del  | 242–284        |
| Serine-pyruvate aminotransferase         | 489903803          | Supplementary Figure S6 | 2 aa ins  | 74–119         |
| 50S ribosomal protein L21                | 489922190          | Supplementary Figure S7 | 1 aa ins  | 42–82          |
| Methionine aminopeptidase                 | 547078960          | Supplementary Figure S8 | 1 aa ins  | 34–70          |
| Bifunctional acetaldehyde-CoA/alcohol dehydrogenase | 500062906 | Supplementary Figure S9 | 1 aa ins  | 534–574        |
| Bifunctional acetaldehyde-CoA/alcohol dehydrogenase | 500062906 | Supplementary Figure S10 | 1 aa ins  | 809–845        |
| Formate acetyltransferase                 | 500063439          | Supplementary Figure S11 | 2 aa ins  | 367–416        |
| ATP synthase FO subunit A                | 547078870          | Supplementary Figure S12 | 1 aa ins  | 131–163        |
| Peptide chain release factor 1           | 489924412          | Supplementary Figure S13 | 2 aa ins  | 197–237        |
| Arginine ABC transporter ATP-binding protein | 489905014          | Supplementary Figure S14 | 1 aa del  | 224–280        |
| Transketolase                            | 489905793          | Supplementary Figure S15 | 4 aa ins  | 338–388        |
| Histidine kinase                         | 547084095          | Supplementary Figure S16 | 1 aa ins  | 362–405        |
| DNA repair ATPase                        | 489906284          | Supplementary Figure S17 | 3 aa ins  | 353–394        |
| n-acetyl-gamma-glutamyl-phosphate reductase | 547072106          | Supplementary Figure S18 | 1 aa ins  | 10–60          |
| Arginine biosynthesis bifunctional protein ArgJ | 547072098          | Supplementary Figure S19 | 1 aa ins  | 1–42          |
| Excinuclease ABC subunit C               | 494111998          | Supplementary Figure S20 | 1 aa ins  | 103–150        |
| Cysteine desulfurase                     | 500063210          | Supplementary Figure S21 | 4 aa ins  | 54–105         |
| 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase | 489906135 | Supplementary Figure S22 | 1 aa ins  | 58–81          |
| Argininosuccinate lyase                   | 547072080          | Supplementary Figure S23 | 5 aa ins  | 405–454        |
| CarD family transcriptional regulator    | 500063173          | Supplementary Figure S24 | 1 aa ins  | 30–79          |
| Acetyltransferase GNAT family            | 547074268          | Supplementary Figure S25 | 1 aa ins  | 112–152        |
| Acetyltransferase GNAT family            | 547074268          | Supplementary Figure S25 | 2 aa ins  | 112–152        |
| Signal recognition particle protein       | 489904236          | Supplementary Figure S26 | 1 aa ins  | 70–110         |
| 50S ribosomal protein L13                | 489923970          | Supplementary Figure S27 | 1 aa del  | 51–90          |
| DNA gyrase B subunit protein              | 547082727          | Supplementary Figure S28 | 2 aa del  | 637–696        |
| Hemolysin III                            | 489923478          | Supplementary Figure S29 | 1 aa ins  | 171–216        |
| Pseudouridine synthase                    | 547071034          | Supplementary Figure S30 | 1 aa ins  | 56–95          |
| Guanylate kinase                         | 500063064          | Supplementary Figure S31 | 4 aa ins  | 85–124         |
| D-alanine–D-alanine ligase               | 493336643          | Supplementary Figure S32 | 2–7 aa ins | 202–244        |

*The indel region indicates the region of the protein where the described CSI is present.

characteristics are summarized in Table 1. Barring an isolated exception, all of the CSIs listed in Table 1 are specifically found in different members of the order Bifidobacteriales and are not present in the protein homologs from other bacteria. Due to their specific presence in bifidobacteria species, the described CSIs provide novel molecular markers for distinguishing and demarcating members of the order Bifidobacteriales from all other bacteria. We have previously described 14 CSPs, whose homologs were specifically found in the 13 different sequenced bifidobacteria species that were available at the time (Gao and Gupta, 2012). Updated BLASTp searches on the sequences of these CSPs confirm that 10 of these CSPs, information for whom is provided in Table 2, are still distinctive characteristics of members of the order Bifidobacteriales and they provide additional molecular markers for identification and functional studies on bifidobacteria.

Molecular Signatures for Some of the Subclades of Bifidobacteriales

In the phylogenetic tree based on concatenated protein sequences, Bifidobacteriales species form a number of different clusters. At the deepest level, of the two main clusters that are observed, one consists of the genus Scardovia and related genera, whereas the other is comprised of species from the genera Bifidobacterium and Gardnerella. In our analyses, we have also identified a number of CSIs and CSPs which distinguish these two clades of the Bifidobacteriales. Figure 3 shows one example of a CSI consisting of a 1 aa insertion in the DNA polymerase IV protein that is specifically found in different Bifidobacterium species and Gardnerella, but which is not found in any of the sequenced Scardovia-related genera of the Bifidobacteriales. Two other CSIs in the ribosomal RNA small subunit methyltransferase E protein and GTP-binding
FIGURE 2 | Partial sequence alignment of the protein synthesis elongation factor-Tu showing a 4 aa insertion in a conserved region that is specific for members of the order Bifidobacteriales. The dashes in this alignment as well as all other alignments show identity with the amino acid on the top line. The
(Continued)
A number of distinct clusters within the genus *Bifidobacterium* are consistently observed in different phylogenetic studies including in the phylogenetic trees constructed in this work (Figure 1). A number of CSIs identified in our work serve to distinguish some of the *Bifidobacterium* clusters. Three of the identified CSIs are specific for the *B. longum* group and sequence information for one of these CSIs, consisting of a 1 aa insertion in the phosphogluconate dehydrogenase, is shown in Figure 5. Sequence information for the other 2 CSIs that are also specific for a subgroup of species from the *B. longum* clade are presented in Supplementary Figures S40, S41 and their characteristics are summarized in Table 3. One additional CSI consisting of a 1 aa insertion in transketolase protein is specifically shared by members of the *B. longum*, *B. bifidum*, and *B. adolescentis* clades. Members of these clusters group together in phylogenetic trees and the shared presence of this CSI supports the view that that the members of these taxa are more closely and specifically related to each other.

The members of the *B. asteroides* cluster forms the deepest branching group within the genus *Bifidobacterium*. A number of CSIs identified in this study are specific for group of species, which are either part of the *B. asteroides* clade or related to this clade. The *B. asteroides* clade is demarcated as being made up of the species *B. asteroides*, *B. indicum*, *B. coryneforme*, and *B. actinocoliniforme* species (marked cluster III in Figure 1) (Lugli et al., 2014; Sun et al., 2015). Surprisingly, in our work no CSI was identified that was commonly shared by all of the species from this clade. However, our work identified four CSIs for a cluster (cluster II) comprising of all of other species from the *B. asteroides* clade, except *B. actinocoliniforme*, which shows the deepest branching within this clade. One example of a CSI specific for members of the *B. asteroides* cluster II consisting of 1 aa insertion in the purine biosynthesis protein purH is shown in Figure 6A. Sequence information for three other CSIs that are also specific for the *B. asteroides* group is presented in Supplementary Figures S43–S45. In our phylogenetic trees as well as in different identified signatures, two *Bifidobacterium* spp. strains A11 and 7101, isolated from honey bee guts (Anderson et al., 2013), also consistently group with the *B. asteroides*. Two CSIs identified in our work are specifically shared by *B. asteroides* and the *Bifidobacterium* sp. A11 and *Bifidobacterium* sp.7101 (referred to as *B. asteroides* cluster I) providing additional evidence of the close relationship of these *Bifidobacterium* strains to the *B. asteroides*. Sequence information for these CSIs is presented in Supplementary Figure S46. Lastly, one additional CSI identified in this work, consisting of a 3 aa insertion in a conserved region of the protein 5'-methylthioadenosine nucleosidase, is commonly shared by all the members of the *B. asteroides* cluster II providing additional evidence of the close relationship of these *Bifidobacterium* strains to the *B. asteroides*.
Zhang et al. Molecular Signatures for the Bifidobacteriales

FIGURE 3 | Partial sequence alignment of DNA polymerase IV showing a 1 aa insertion that is specific for the Bifidobacterium and Gardnerella species, but not found in any other Bifidobacteriales. Information for other CSIs specific for this clade is presented in Table 3 and Supplementary Figures S33–S35.

| Bifidobacterium adolescentsis            | 489904866  | FLPVDDDYVYFMMPMBEI F | TEFVDRVTDRIEQVSDVCEY | 88 |
| Bifidobacterium angulatum               | 489923981  | K---R----Q------M    | RD----LT-----F--------G-- |
| Bifidobacterium animals                 | 490389093  | H----K---S--V----RE- | HT----RE----------I--|
| Bifidobacterium asteroides              | 504835263  | HT-----SIT-------I- | T7------I-NQ--------|
| Bifidobacterium bifidum                 | 489911054  | M---M--------V     | E-R----Q-------F-----A-|
| Bifidobacterium breve                   | 651390579  | M-----M--------M   | K-LI-TQ-------Q-----G--|
| Bifidobacterium cametulatum             | 49932716     | K---M---A-------H- | K-LI-I-G------Q-----G--|
| Bifidobacterium choerinum               | 639202583  | H-SA----QI-------D- | GTI------I-------I--|
| Bifidobacterium dentium                 | 489930408  | M-----------Q      | JG--------K--------|
| Bifidobacterium gallicum                | 493383339  | M----------Q---I-- | Q-------Q---------|
| Bifidobacterium longum                  | 494112237  | M---M--------M     | K-LI-LT-------Q-----G--|
| Bifidobacterium magnus                  | 651390243  | M-----------Q      | JG--------K--------|
| Bifidobacterium minimum                 | 551240467  | M-----------Q      | JG--------K--------|
| Bifidobacterium moulakalense            | 575769369  | M-----------Q      | JG--------K--------|
| Bifidobacterium pseudocatenulatum       | 49033627    | M-----------Q      | JG--------K--------|
| Bifidobacterium pseudolongum            | 651883820  | M-----------Q      | JG--------K--------|
| Bifidobacterium ruminantium             | 651886874  | M-----------Q      | JG--------K--------|
| Bifidobacterium subtile                 | 639065366  | M-----------Q      | JG--------K--------|
| Bifidobacterium thermadophilum          | 657871235  | M-----------Q      | JG--------K--------|
| Bifidobacterium thermophilum            | 505263981  | M-----------Q      | JG--------K--------|
| Bifidobacterium thermuiensae            | 651881823  | M-----------Q      | JG--------K--------|
| Bifidobacterium pullorum                | 705444526  | M-----------Q      | JG--------K--------|
| Bifidobacterium bivatii                 | 705395711  | M-----------Q      | JG--------K--------|
| Bifidobacterium scardovii               | 705445973  | M-----------Q      | JG--------K--------|
| Bifidobacterium crudiactis              | 736119830  | M-----------Q      | JG--------K--------|
| Bifidobacterium mongolensis             | 705435860  | M-----------Q      | JG--------K--------|
| Bifidobacterium psychraecophilum        | 705397424  | M-----------Q      | JG--------K--------|
| Bifidobacterium kashiwanenhense         | 74613641   | M-----------Q      | JG--------K--------|
| Bifidobacterium actinomoliniforme       | 705420974  | M-----------Q      | JG--------K--------|
| Bifidobacterium bohemicum               | 91751172   | M-----------Q      | JG--------K--------|
| Bifidobacterium stellensboscense        | 73611112   | M-----------Q      | JG--------K--------|
| Bifidobacterium cuniculi                | 705443047  | M-----------Q      | JG--------K--------|
| Bifidobacterium indicum                 | 705389070  | M-----------Q      | JG--------K--------|
| Bifidobacterium coryneforme             | 759123986  | M-----------Q      | JG--------K--------|
| Bifidobacterium bombi                   | 763215354  | M-----------Q      | JG--------K--------|
| Bifidobacterium aesculapi                | 943597392  | M-----------Q      | JG--------K--------|
| Bifidobacterium saguinii                | 672997532  | M-----------Q      | JG--------K--------|
| Bifidobacterium reuteri                 | 705426260  | M-----------Q      | JG--------K--------|
| Bifidobacterium calliclithicos           | 759442531  | M-----------Q      | JG--------K--------|
| Bifidobacterium merycicum               | 705458204  | M-----------Q      | JG--------K--------|
| Bifidobacterium gallinarum              | 672964903  | M-----------Q      | JG--------K--------|
| Bifidobacterium stercoris               | 673001421  | M-----------Q      | JG--------K--------|
| Gardnerella vaginallis                  | 490207126  | M-----------Q      | JG--------K--------|
| Alloscardovia omnicolens                | 545376575  | M-----------Q      | JG--------K--------|
| Alloscardovia cricetti                  | 51607855    | M-----------Q      | JG--------K--------|
| Parascardovia denticolens               | 493331646  | M-----------Q      | JG--------K--------|
| Scardovia inopinata                     | 493335694  | M-----------Q      | JG--------K--------|
| Scardovia wiggiae                       | 494250503  | M-----------Q      | JG--------K--------|
| Aciromices phytoaeuilu                  | 518600991  | M-----------Q      | JG--------K--------|
| Alkaliphilus metalldigreens             | 501012359  | M-----------Q      | JG--------K--------|
| Caldistrix abysii                       | 493987017  | M-----------Q      | JG--------K--------|
| Clostridium arbusi                      | 497914541  | M-----------Q      | JG--------K--------|
| Eikenella corrodens                     | 565108605  | M-----------Q      | JG--------K--------|
| Enterococcus raffinosus                 | 498440001  | M-----------Q      | JG--------K--------|
| Eubacterium ramulus                     | 545613317  | M-----------Q      | JG--------K--------|
| Fackleria ignava                        | 493752444  | M-----------Q      | JG--------K--------|
| Gottschalkia acidurici                  | 504780063  | M-----------Q      | JG--------K--------|
| Haemophilus parasuis                    | 491998641  | M-----------Q      | JG--------K--------|
| Lactobacillus fructivorans              | 497707820  | M-----------Q      | JG--------K--------|
| Lactococcus lactis                     | 497687618  | M-----------Q      | JG--------K--------|
| Longisporae ulidica                     | 511762071  | M-----------Q      | JG--------K--------|
| Methanococcus bourgensis                | 504680316  | M-----------Q      | JG--------K--------|
| Pasteurella multocida                   | 492120851  | M-----------Q      | JG--------K--------|
| Prevotella bryantii                     | 493325432  | M-----------Q      | JG--------K--------|
| Sporomusa ovata                        | 544739093  | M-----------Q      | JG--------K--------|
| Streptococcus pororis                   | 493116418  | M-----------Q      | JG--------K--------|

FIGURE 3 | Partial sequence alignment of DNA polymerase IV showing a 1 aa insertion that is specific for the Bifidobacterium and Gardnerella species, but not found in any other Bifidobacteriales. Information for other CSIs specific for this clade is presented in Table 3 and Supplementary Figures S33–S35.
**DISCUSSION**

Members of the order *Bifidobacteriales* are one of the main groups within bacteria where several members exhibit health-promoting probiotic effects on humans (Biavati et al., 2000; Biavati and Mattarelli, 2006; Ventura et al., 2007b, 2009a; Cronin et al., 2011; Turrioni et al., 2011). Other *Bifidobacteriales* species are also responsible for implicated in the development of dental caries as well as bacterial vaginosis and urinary tract infections (Bradshaw et al., 2006; Mantzourani et al., 2009; Ventura et al., 2009b; Kenyon and Osbak, 2014). However, very little is known at present concerning the genetic or biochemical characteristics of these bacteria that mediate their beneficial or pathogenic effects. In the present work, we have carried out detailed phylogenetic and comparative analyses of protein sequences from the genomes of *Bifidobacteriales* species to examine in depth their evolutionary relationships and also to identify molecular markers that are unique to these bacteria at multiple phylogenetic levels. Based on a robust and comprehensive phylogenetic tree for the *Bifidobacteriales* species based on 614 core proteins from the sequenced genomes, the following inferences regarding the evolutionary relationships among the *Bifidobacteriales* species could be made. (i) The sequenced *Bifidobacteriales* species appear to form two main clusters, a deeper clade consisting of the *Scardovia*-related genera (viz. *Scardovia, Parascardovia* and *Alloscardovia*) and another cluster grouping together *Bifidobacterium* and *Gardnerella* genera. (ii) *Gardnerella vaginalis* rather than branching separately is found to consistently branch in between different *Bifidobacterium* species.

---

**Table 3**

| Protein name | GI number | Figure number | Indel size | Indel position | Specificity |
|--------------|-----------|---------------|------------|----------------|-------------|
| DNA polymerase IV | 489904486 | Figure 3       | 1 aa ins   | 88–125         | *Bifidobacterium*-Gardnerella |
| Ribosomal RNA small subunit methyltransferase E | 540781721 | Supplementary Figure S33 | 3 aa del | 118–160 | *Bifidobacterium*-Gardnerella |
| GTP-binding protein YchF | 54075080 | Supplementary Figure S34 | 1 aa ins | 309–354 | *Bifidobacterium*-Gardnerella |
| Cytochrome C | 500682679 | Supplementary Figure S35 | 3 aa del | 730–765 | *Bifidobacterium* |
| Triosephosphate isomerase | 651360171 | Figure 4 | 1 aa ins | 251–286 | Scardovia clade |
| FHA domain protein | 49335662 | Supplementary Figure S36 | 2 aa ins | 37–67 | Scardovia clade |
| Phosphate ABC transporter substrate-binding protein | 294458767 | Supplementary Figure S37 | 2 aa ins | 32–77 | Scardovia clade |
| Glycosyl transferase | 493335662 | Supplementary Figure S38 | 2 aa ins | 167–206 | Scardovia clade |
| Phosphogluconate dehydrogenase | 49776884 | Figure 5 | 1 aa ins | 360–401 | *B. longum* cluster |
| PhoU family transcriptional regulator | 489926631 | Supplementary Figure S40 | 2 aa del | 159–190 | *B. longum* cluster |
| Cystathionine gamma-synthase | 494112910 | Supplementary Figure S41 | 2 aa ins | 262–302 | *B. longum* cluster |
| Transketolase | 489905793 | Supplementary Figure S42 | 1 aa ins | 234–274 | *B. longum, B. bifidum and B. adolescens* clade |
| Purine biosynthesis protein purH | 658453400 | Figure 6A | 1 aa ins | 247–278 | *B. asteroides* cluster II # |
| Shikimate dehydrogenase | 658453363 | Supplementary Figure S43 | 1 aa ins | 264–301 | *B. asteroides* cluster II # |
| 5-methyltetrahydropteroyltriglutamate–homocysteine methyltransferase | 504834759 | Supplementary Figure S44 | 1 aa ins | 336–369 | *B. asteroides* cluster II # |
| ABC transporter substrate-binding protein | 504835116 | Supplementary Figure S45 | 1 aa del | 253–286 | *B. asteroides* cluster II # |
| 5’-methylthioadenosine nucleosidase | 504835309 | Figure 6B | 3 aa ins | 1–33 | *B. asteroides*-related cluster IV # |
| Peptide ABC transporter ATP-binding protein | 504834913 | Supplementary Figure S46A | 20 aa ins | 76–127 | *B. asteroides* cluster IV # |
| N-acetyl-gamma-glutamyl-phosphate reductase | 504834965 | Supplementary Figure S46B | 1 aa ins | 34–74 | *B. asteroides* cluster IV # |

# The *B. asteroides*-related cluster I, II, and IV are demarcated in Figure 1.
**Scardovia-related genera**

| Genus                  | Accession Number | Sequence Description |
|------------------------|------------------|----------------------|
| Alloscardovia omnolens | 651360171        | YGGEVTAQANLIMNEA     |
| Metascardovia crioeti  | 648490205        | --N---L---          |
| Parascardovia denticolens | 493334606     | --S--DL--Q--H-      |
| Scardovia inopinata   | 535476484        | --S--MDL--Q--H-     |
| Scardovia wiggiae     | 515714829        | --S--MDL--DQ-H-     |
| Bifidobacterium angulatum | 489922926     | --S--K--AEILIA-P-   |
| Bifidobacterium asterolidae | 504834734    | --S--K--SGM1TQP-P-  |
| Bifidobacterium bifidum | 489909858        | --S--K--HLIG-P-     |
| Bifidobacterium bomu   | 651391054        | --S--K--DQLRQP-P-   |
| Bifidobacterium breve  | 585130875        | --S--K--AEILISQP-P- |
| Bifidobacterium longum | 547239381        | --S--K--AEILISQP-P- |
| Bifidobacterium magnun | 651390093        | --S--K--VELIQLQ-P-  |
| Bifidobacterium minimum | 651387732       | --S--K--SRWQK1LQP-P |
| Bifidobacterium sp. 71D1 | 658453366       | --S--K--SRM1TQP-P-  |
| Bifidobacterium sp. All | 639450006        | --S--K--SM1TQP-P-   |
| Bifidobacterium subtile | 639063916        | --S--G--AEILIE-P-   |
| Bifidobacterium thermacidophilum | 657872129 | --S--K--SGL1TQP-P-  |
| Bifidobacterium thermophilum | 505623497 | --S--K--SGL1TQP-P-  |
| Bifidobacterium tourismeneae | 651882159 | --S--K--SGQLS-P-    |
| Bifidobacterium psychraeophilum | 705400252 | --S--B--AQL1Q--P-   |
| Bifidobacterium bohemiacum | 705453892 | --S--S--RELIS-P-    |
| Bifidobacterium gallinarum | 705423798 | --S--S--VQL1TQP-P-  |
| Bifidobacterium crudilactis | 736120758 | --S--S--AQL1Q-R--   |
| Bifidobacterium coryneforme | 799123296 | --S--K--SQ--Q-H--   |
| Bifidobacterium pullorum | 705442364 | --S--S--VR1L1QP-P-  |
| Bifidobacterium mongoliense | 705436912 | --S--K--OQ--Q-IA-P- |
| Bifidobacterium reuteri | 763215760 | --S--S--K--AEILISQP-P |
| Bifidobacterium merycium | 705458800 | --S--S--K--AEILIE-P- |
| Bifidobacterium saguinii | 727803668 | --S--S--K--AEILISQP-P |
| Bifidobacterium bombii | 763213935 | --S--S--K--VET1DQN-P |
| Bifidobacterium kashiwanoense | 705419368 | --S--S--K--VET1DQN-P |
| Bifidobacterium caldichinos | 672961096 | --S--S--K--VELIG-P-  |
| Bifidobacterium pseudocatulatum | 524473885 | --S--S--K--VELIG-P-  |
| Bifidobacterium sp. AGR2158 | 651885675 | --S--S--K--VELIS-P-  |
| Bifidobacterium choerinum | 639202923 | --S--S--K--VELIS-P-  |
| Bifidobacterium gallicum | 762077799 | --S--S--K--VELIG-P-  |
| Bifidobacterium adolescentis | 489057779 | --S--S--K--VELIG-P-  |
| Bifidobacterium actinocoloniiforme | 705421242 | --S--S--K--VQL1TQP-P |
| Bifidobacterium cuniculi | 705455566 | --S--S--K--VELIG-P-  |
| Bifidobacterium bisavili | 705383331 | --S--S--K--VELIG-P-  |
| Bifidobacterium catenulatum | 489931487 | --S--S--K--VELIG-P-  |
| Bifidobacterium stenohemence | 736510025 | --S--S--K--VELIG-P-  |
| Bifidobacterium mokalabense | 736975813 | --S--S--K--VELIG-P-  |
| Bifidobacterium rumianum | 651886380 | --S--S--K--VELIG-P-  |
| Bifidobacterium dentium | 171278564 | --S--S--K--VELIG-P-  |
| Bifidobacterium asculapii | 943597283 | --S--S--K--VELIG-P-  |
| Bifidobacterium scardovil | 672995889 | --S--S--K--VELIG-P-  |
| Bifidobacterium sp. MRE112 | 57042516 | --S--S--K--VELIG-P-  |
| Bifidobacterium stercoris | 673003375 | --S--S--K--VELIG-P-  |
| Bifidobacterium seculare | 672990889 | --S--S--K--VQL1TQP-P |
| Bifidobacterium indicum | 65535248 | --S--S--K--SM1TQP-P- |
| Bifidobacterium animalis | 490326467 | --S--S--K--CVE1L1QP-P |
| Bifidobacterium pseudoalum | 746105858 | --S--S--K--VEGL1TQP-P |
| Bifidobacterium longum | 524120693 | --S--S--K--AEILISQP-P |
| Gardnerella vaginalis | 490207959 | --S--S--SGL1TQP-P-  |
| Buchnera aphidicola | 499991861 | --K--KGL11QK-P-     |
| Cellulosononas fim | 503536946 | --K--S--VGL1TQP-P-  |
| Acidimicrobium ferrooxidans | 506278511 | --D--D--VAF1L1QK-P- |
| Persephonella marina | 501943652 | --K--N--ROL1KQ-P-   |
| Verrucomicrobiobacterium | 640163520 | --K--EQL1AQQP-P-    |
| Waddlia chondrophila | 502947249 | --K--P--R1L1QKP-P-  |
| Xylanomonas cellulosilytica | 502641898 | --K--K--S111QKP-P-  |

**Other Bifidobacteria**

**Other Bacteria**

FIGURE 4 | Example of 1 aa conserved signature indel in the protein triosephosphate isomerase that is specific for the Scardovia clade comprising of the genera Scardovia, Parascardovia, Metascardovia, and Alloscardovia. Information for other CSIs specific for this clade is presented in Table 3 and Supplementary Figures S36–S39.
Within *Bifidobacterium* species, a number of distinct clusters, referred to as the *B. asteroides*, *B. pseudolongum*, *B. longum*, *B. bifidum*, *B. adolescentis*, *B. pullorum*, and *B. boum* groups, are observed as described in earlier work (Lugli et al., 2014; Sun et al., 2015). Of these clusters, the *B. asteroides* group forms the deepest branching lineage within the *Bifidobacterium* (Bottacini et al., 2012; Lugli et al., 2014; Sun et al., 2015).

The present work also identified large number of novel molecular signatures in the forms of CSIs and CSPs, which are specific characteristics of the members of the order *Bifidobacteriales* at multiple phylogenetic levels. Of these signatures, 32 CSIs and 10 CSPs are specific for the entire order *Bifidobacteriales*. The identified *Bifidobacteriales*-specific CSIs are present in assorted widely distributed proteins carrying out wide variety of cellular functions. All of the 10 *Bifidobacteriales*-specific CSPs are proteins of unknown functions. Given the specificity of these CSIs and CSPs for the *Bifidobacteriales*, the genetic changes leading to these molecular characteristics have likely occurred in a common ancestor of the *Bifidobacteriales* (Gao and Gupta, 2005, 2012). Additionally, our analyses have also identified many other molecular signatures (CSIs and CSPs), which independently support the existence of a number of clades of bifidobacteria that are consistently observed in phylogenetic trees. The clades identified by these molecular signatures include, (i) a clade encompassing the genera *Scardovia*, *Parascardovia* and *Alloscardovia*, (ii) signatures that are commonly shared by
**Figure 6** | Conserved signature indels that are specific for the *B. asteroides*-related clades of the *Bifidobacteriales*. (A) Partial sequence alignment of the purine biosynthesis protein purH showing a 1 aa insertion which is specific for the *B. asteroides* cluster II species in the protein tree (*Figure 1*); (B) Excerpt from sequence alignment of the protein 5′-methylthioadenosine nucleosidase showing a 3 aa insertion that is specific for the *B. asteroides*-related cluster IV in the protein tree.
Bifidobacterium and Gardnerella species to the exclusion of other bifidobacteria, and (iii) signatures demarcating specific clusters of B. asteroides- or B. longum- related species.

The order Bifidobacteriales presently contains a single family, Bifidobacteriaceae. Based upon the results of phylogonomic studies and identified molecular signatures, it appears that the members of this order could be divided into two family-level groups, one comprising of the Scardovia-related genera (viz. Scarodivia, Parascardovia, and Alloscardovia) and the other consisting of the genera Bifidobacterium and Gardnerella. However, genome sequence information for members of several newly described Scardovia-related genera (viz. Aeriscardovia, Neoscardovia, and Pseudoscardovia), is lacking at present (Simpson et al., 2004; García-Aljaro et al., 2012; Killer et al., 2013). In future studies, depending upon whether the species from these genera branch with the Scardovia-clade and their sharing of the molecular signatures specific for this clade, the possibility of dividing the order Bifidobacteriales into two or more families could be considered.

The genus Bifidobacterium, which is comprised of 49 species and subspecies, contains most of the recognized taxa within the order Bifidobacteriales. Although earlier phylogenetic studies have consistently observed 6–7 distinct clusters of Bifidobacterium species (Ventura et al., 2006, 2007b; Turroni et al., 2011; Lugli et al., 2014; Sun et al., 2015), due to lack of any other distinguishing characteristics, no attempt has been made to formally recognize any of these clusters. In our work, we have identified a number of molecular signatures that are either completely or largely specific for the members of two of these clusters (viz. the B. asteroides and B. longum groups). Of these clusters, the distinctness of the B. asteroides group (comprising of the species B. asteroides, B. indicum, B. coryneforme, B. actinocoloniforme, B. sp. A11, and B. sp. 7101) which forms the deepest branching lineage within the Bifidobacterium, is supported by 2 CSIs and 4 CSPs that are uniquely shared by most of the members of this clade. Further, most of the species which are part of the B. asteroides clade have been isolated from the gastrointestinal tract of honey bees, and unlike other bifidobacteria, they are also capable of carrying out respiratory metabolism (Killer et al., 2010, 2011; Bottacini et al., 2012; Lugli et al., 2014; Sun et al., 2015). All of these characteristics indicate that the members of the B. asteroides clade are a good candidate for recognition as a distinct genus level taxon within the order Bifidobacteriales.

The molecular markers for the order Bifidobacteriales and some of its clades, in addition to their utility for taxonomy and diagnostic studies (Ahmod et al., 2011; Gupta, 2014; Wong et al., 2014), also provide important new tools for genetic and biochemical studies. Earlier work on a number of CSIs in the Hsp60 and Hsp70 proteins has established that both large and small CSIs in conserved proteins are essential for the group of organisms in which they are found (Singh and Gupta, 2009; Gupta, 2016b). Removal of these CSIs, or any significant change in them, was shown to be incompatible with the cellular growth of the CSI-containing organisms. Thus, the identified CSIs are predicted to play essential role in the organisms in which they are found. Structural studies on several studied CSIs show that the sequences corresponding to them are present in the surface loops of the proteins (Singh and Gupta, 2009; Gupta and Khadka, 2016). Limited structural work on some of the Bifidobacteriales-specific CSIs that we have carried out also shows that these CSIs are located in the surface loops of the proteins. One example of the structural location of a Bifidobacteriales-specific CSI is illustrated in Figure 7. In this case, a homology model of protein synthesis elongation factor Tu from B. longum was created to determine the location of the 4 aa Bifidobacteriales-specific CSI found in this protein. A structural comparison of the EF-Tu from B. longum and E. coli shown in Figure 7 reveals that the CSI in the B. longum homolog is present in the protein surface loop within the GTPase domain of EF-Tu. The surface loops in proteins play important role in mediating protein-protein or protein-ligand interactions and it is expected that the identified CSIs are involved in mediating novel interactions that are specific and essential for the CSI-containing organisms (Akiva et al., 2008; Hashimoto and Panchenko, 2010). Similar to the CSIs in the EF-Tu protein, our work has identified numerous other CSIs in different essential proteins, which are specific for the Bifidobacteriales species. Functional studies on proteins harboring these CSIs provide an important means for discovering novel biochemical characteristics that are unique to either all Bifidobacteriales or specific clades of these bacteria, and which could possibly also provide useful insights into the growth-promoting as well as pathogenic effects of some of these bacteria.

AUTHOR CONTRIBUTIONS

GZ, BG, MA, BK carried out comparative analyses of the bifidobacteriales genomes to identify signatures reported here. ZG and MA constructed phylogenetic trees and BK carried out homology modeling of the protein sequences. BG, MA, and RG were responsible for the writing and editing of the manuscript. All of the work was carried out under the direction of RG.
Gupta, R. S., Naushad, S., Fabros, R., and Adeolu, M. (2016). A phylogenomic
Jian, W. Y., and Dong, X. Z. (2002). Transfer of
Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement
Frontiers in Microbiology | www.frontiersin.org
16
Zhang et al. Molecular Signatures for the
Huys, G., Vancanneyt, M., D’Haene, K., Falsen, E., Wauters, G., and Vandamme,
Mantzourani, M., Fenlon, M., and Beighton, D. (2009). Association between
Miyake, T., Watanabe, K., Watanabe, T., and Oyaizu, H. (1998). Phylogenetic
Killer, J., Kopecný, J., Mrázek, J., Koppová, I., Havlík, J., Benada, O., et al. (2011).
Lugli, G. A., Milani, C., Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J. H., Zomer, A., et al. (2014). Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. Proc. Natl. Acad. Sci. U.S.A. 107, 19514–19519. doi: 10.1073/pnas.1011100107
Turroni, F., van Sinderen, D., and Ventura, M. (2011). Genomics and ecological overview of the genus Bifidobacterium. Int. J. Food Microbiol. 149, 37–44. doi: 10.1016/j.ijfoodmicro.2012.09.003
Turroni, F., van Sinderen, D., and Ventura, M. (2009). Bifidobacteria: from ecology to genomics. Front. Biosci. 14, 4673–4684. doi: 10.2741/3559
Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J. H., Zomer, A., et al. (2010). Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. Proc. Natl. Acad. Sci. U.S.A. 107, 19514–19519. doi: 10.1073/pnas.1011100107
Turroni, F., van Sinderen, D., and Ventura, M. (2011). Genomics and ecological overview of the genus Bifidobacterium. Int. J. Food Microbiol. 149, 37–44. doi: 10.1016/j.ijfoodmicro.2012.09.003
Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J. H., Zomer, A., et al. (2010). Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. Proc. Natl. Acad. Sci. U.S.A. 107, 19514–19519. doi: 10.1073/pnas.1011100107
history of an ancient phylum. Microbiol. Mol. Biol. Rev. 71, 495–548. doi: 10.1128/MMBR.00005-07
Ventura, M., Canchaya, C., Zink, R., Fitzgerald, G. F., and van Sinderen, D. (2004). Characterization of the groEL and groES loci in Bifidobacterium breve UCC 2003: genetic, transcriptional, and phylogenetic analyses. Appl. Environ. Microbiol. 70, 6197–6209. doi: 10.1128/AEM.70.10.6197-6209.2004
Ventura, M., O’Flaherty, S., Claesson, M. J., Turrini, F., Kleenhammer, T. R., van Sinderen, D., et al. (2009a). Genome-scale analyses of health-promoting bacteria: probiogenomics. Nat. Rev. Microbiol. 7, 61–71. doi: 10.1038/nrmicro2047
Ventura, M., Turrini, F., Zomer, A., Foroni, E., Giubellini, V., Bottacini, F., et al. (2009b). The Bifidobacterium dentium Bd1 genome sequence reflects its genetic adaptation to the human oral cavity. PLoS Genet. 5:e1000785. doi: 10.1371/journal.pgen.1000785
Ventura, M., and Zink, R. (2003). Comparative sequence analysis of the tuf and recA genes and restriction fragment length polymorphism of the internal transcribed spacer region sequences supply additional tools for discriminating Bifidobacterium lactis from Bifidobacterium animalis. Appl. Environ. Microbiol. 69, 7517–7522. doi: 10.1128/AEM.69.12.7517-752.2003
Whelan, S., Liò, P., and Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet. 17, 262–272. doi: 10.1016/S0168-9525(01)02272-7
Wong, S. Y., Paschos, A., Gupta, R. S., and Schellhorn, H. E. (2014). Insertion/deletion-based approach for the detection of Escherichia coli O157:H7 in freshwater environments. Environ. Sci. Technol. 48, 11462–11470. doi: 10.1021/es502794h
Xu, D., and Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J. 101, 2525–2534. doi: 10.1016/j.bpj.2011.10.024
Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K. H., et al. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst. Appl. Microbiol. 31, 241–250. doi: 10.1016/j.syapm.2008.07.001
Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res. 42, D643–D648. doi: 10.1093/nar/gkt1209
Zhi, X. Y., Li, W. J., and Stackebrandt, E. (2009). An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. Int. J. Syst. Evol. Microbiol. 59, 589–608. doi: 10.1099/ijs.0.05780-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhang, Gao, Adeolu, Khadka and Gupta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.