Prophages of the genus *Bifidobacterium* as modulating agents of the infant gut microbiota

Running title: Bifidobacterial prophages and infant gut microbiota

Key words: Bifidobacteria, bifidophages, infant gut microbiota, microbiomes, comparative genomics, phylogenomics

Gabriele Andrea Lugli[1], Christian Milani[1], Francesca Turroni[1], Denise Tremblay[2], Chiara Ferrario[1], Leonardo Mancabelli[1], Sabrina Duranti[1], Doyle V. Ward[3], Maria Cristina Ossiprandi[4], Sylvain Moineau[2], Douwe van Sinderen[5] and Marco Ventura[1]

Laboratory of Probiogenomics, Department of Life Sciences, University of Parma, Parma, Italy[1]; Département de Biochimie, Microbiologie et Bio-Informatique & PROTEO, Faculté des Sciences et de Génie, Félix d'Hérelle Reference Center for Bacterial Viruses & GREB, Faculté de Médecine Dentaire, Université Laval, Québec City, Québec, Canada[2]; Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA[3]; Department of Medical-Veterinary Science, University of Parma, Parma, Italy[4]; APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland[5];

Corresponding author. Mailing address for Marco Ventura Department of Life Sciences, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: marco.ventura@unipr.it

Abstract

Phage predation is one of the key forces that shape genetic diversity in bacterial genomes. Phages are also believed to act as modulators of the microbiota composition and, consequently, as driving agents of bacterial speciation in complex bacterial communities. Very little is known about the occurrence and genetic variability of (pro)phages within the *Bifidobacterium* genus, a dominant bacterial group of the human infant microbiota. Here, we performed cataloguing of the predicted prophages sequences from the currently available bifidobacterial genomes. We analysed their genetic diversity and deduced evolutionary development, thereby highlighting an intriguing origin. Furthermore, we assessed infant gut microbiomes for the presence of (pro)phage sequences and found compelling evidence that these viral elements influence the composition of bifidobacterial communities in the infant gut microbiota.

INTRODUCTION

It has been estimated that 25 % of phage genomes on earth correspond to prophages (1, 2). Prophage-associated sequences make up a sizable part of the mobilome of bacterial genomes, and these accessory genes contribute significantly to the bacterial inter-strain genetic variability (3, 4). The advent of the genomic era has highlighted the apparent importance of prophages as a catalyst for lateral gene transfer between bacteria, and as a selective force that shapes the population structure of a bacterial species (5). Moreover, prophages are far from being passive residents as they can modify existing or confer new properties to their host (6), thereby increasing its fitness (2).

In environmental samples such as the human gut, the collective genome content of (pro)phages, known as the gut virome, constitutes a substantial proportion of the encountered genetic biodiversity (7). In fact, the human gut is probably the richest concentration of biological entities (8). The in depth study of human viral communities is only in its infancy, though has in recent times enjoyed significant progress due to advancements in sequencing technologies and associated data handling abilities (9). For example, a previously unidentified bacteriophage present in the majority of published human faecal metagenomes has recently been described (10). The majority of the deduced proteins specified by this novel phage do not match known sequences in the database, and explains why it had remained undetected. In this context, there is a growing awareness of the key contribution of the virome not only in terms of overall genetic diversity, but also as agents that are capable of modulating the gut microbiota composition (11). A prophage-host network of the human gut has recently been deduced, including the elucidation of numerous novel host-phage associations (12).

Bifidobacteria represent one of the dominant microbial groups that occur in the mammalian gut, as well as in the digestive tract of birds and social insects (13, 14). As members of the human gut microbiota, they reach a particularly high relative abundance in infants (15, 16). Among other microbial members of the gut microbiota, *Bifidobacterium* represent an important commensal genus whose presence is often associated with health-promoting effects.

Among the high G+C Gram positive bacteria, prophages of bifidobacteria have only very recently been investigated. Genome analyses of 12 bifidobacterial genomes from human gut species have provided convincing evidence that phage infections do play a role in the genetic make-up of this genus (17-19). Interestingly, although the genetic signs for the existence of phages infecting bifidobacteria are clear, there is only fragmentary information about virion identification from this bacterial group (20). Furthermore, the precise extent of prophage distribution in bifidobacterial genomes and their biological role is still unknown.

In this study, we performed an extensive *in silico* survey of prophages in infant microbiome datasets and in the 48 current publicly available genomes of bifidobacterial type strains (NCBI source), representing all currently known (sub)species belonging to the genus *Bifidobacterium* (21, 22). Phage particles were also isolated and morphologically characterized from bifidobacterial taxa following a prophage induction protocol. Furthermore, we explored the contribution of bifidobacterial (pro)phages, designated here as bifido(pro)phages, as regulators of bifidobacterial population dynamics within the infant gut microbiota.

MATERIALS AND METHODS

Bifidobacterial and phages genome sequences. We retrieved the genome sequences of the 48 *Bifidobacterium* type strains from the National Center for Biotechnology Information (NCBI) public database (Table 1). We also analysed the genome sequences of 1260 dsDNA bacteriophages, similarly available in the NCBI database.

Bifidoprophages identification. The pan-genome of the genus *Bifidobacterium* was screened for prophage sequences using as identification markers genes encoding integrases, portal proteins and endolysins (17, 18). A manually examination of the DNA region surrounding a putative phage-encoding gene was then performed, followed by a second screening through BLAST analysis (23) (E-value cut-off of $1e^{-5}$) involving all putative prophage genes collected. These screenings allowed us to identify integrases, portal proteins and endolysins that did flank a cluster of phage genes, and to then define the boundaries of such a putative prophage genome. The resulting identified prophages are listed in Table 1.

Analysis of bifidoprophages genome sequences. For each bifidoprophage pair, a value of nucleotide similarity, based on localized sequence alignments of the genomes, was calculated using the software LAST (24). The results were used to build a dotplot matrix representing the genomic synteny of the two prophages, and to generate a clustering tree, which was reconstructed using the FigTree software (http://tree.bio.ed.ac.uk/software/figtree/). Identified bifidoprophage genes were further screened against multiple databases such as NCBI, PHAST

(25) and Pfam (26). The resulting genes were, where possible, assigned to functional phage modules following detailed manual annotation.

Network of similarities between phages. For all bifidoprophages and phages retrieved from NCBI, a cluster of orthologous genes (COGs) calculation was performed using the pan-genome analysis pipeline (27) (PGAP). The open reading frame (ORF) content of the examined genomes was organized in functional gene clusters using gene family (GF) method involving comparisons of each protein against all other proteins using BLAST analysis (employing an E-value cut-off of $1e^{-10}$, coupled to at least 50 % identity across at least 50 % of each of the two protein sequences). Sequences were then clustered into protein families using a graph theory-based Markov clustering algorithm (28) (MCL). Each set of orthologous proteins was used to create a network through Cytoscape software (29), where each line represents the correlation identified between a COG family and a phage.

Meta-genomic/transcriptomic analyses. Faecal metagenomic and corresponding metatranscriptomic data sets of healthy infants were retrieved from the NCBI public database (BioProject ID: PRJNA63661 and ID: PRJNA218186, respectively). Individual data sets were filtered to improve dataset quality as follows: a preliminary filtering step was performed to obtain only high quality reads (minimum mean quality score 20, window-size 5, quality threshold 25, minimum length 80) using the fastq-mcf script (http://code.google.com/p/ea-utils), and a second filtering process was performed to remove human reads, using the human genome sequence as template. Furthermore, the metatranscriptomic data were processed to remove rRNA-encompassing reads. With these collected filtered reads, for every sample, we identified

bifidoprophage-associated reads within the data samples. To evaluate the abundance of bifidobacterial and bifidoprophage reads, we reconstructed 48 bifidobacterial type strain genomes lacking deduced bifidoprophages sequences. Furthermore, we performed a normalization of the relative read counts based on the sequencing output and length of the genomes. The resulting rpkm counts, i.e., read mapping to the genome per kilobase of transcript per million reads sequenced (30), were used to investigate the abundance of every genome within the filtered metagenomic data sets and the abundance of every gene within the metatranscriptomic data sets. The Burrows-Wheeler Aligner program was used for the alignment of the read Aligner (31), while the software employed to calculate read counts corresponding to either bifidoprophage or bifidobacterial genomes was HTSeq (32).

Bifidobacteriophage induction. *B. bifidum* LMG 11041, *B. boum* LMG 10736, *B. choerinum*, LMG 10510, *B. longum subsp. suis* LMG 21814, *B. moukalabense* DSM 27321, *B. ruminantium* LMG 21811, and *B. saeculare* LMG 14934 were grown in de Man-Rogosa-Sharp (MRS) supplemented with 0.05 % (w/v) L-cysteine hydrochloride and incubated at 37°C in an anaerobic chamber. Strains were grown until the optical density (600 nm) reached 0.1 to 0.3 and mitomycin C was added at various concentrations, from 0.1 to 5 µg/ml. Hydrogen peroxide was also used following a previously described protocol (17). The cultures were incubated at 37^{o}C for 18 hours, centrifuged, and filtered (0.45 µm). Lysates were observed under a transmission electron microscope (TEM) as described previously (33). Capsid size as well as tail length and width were determined by measuring at least 15 phage specimens, except for *B. boum* for which five specimens were examined.

RESULTS AND DISCUSSION

Identification of prophage-like elements in bifidobacteria. Recently, draft genome sequences of all 48 currently known type strains encompassed by the *Bifidobacterium* genus have been determined (21), providing valuable genetic data to characterize their prophage content. Integrase- and *cI* repressor-encoding genes are considered key markers for the identification of prophages in bacterial genomes (17, 18). In addition to these phage genes, screening for genes that encode putative portal proteins or endolysins in the above mentioned bifidobacterial genome sequences revealed the presence of 60 predicted bifidoprophage genomes (Table 1). Furthermore, an additional 30 prophage remnants (i.e., bifidoprophages that exhibit obvious genome degeneration, showing less than 20 ORFs or a genome sequence length lower than 10Kb) were retrieved from the 48 bifidobacterial type strains, suggesting phage infection and genome integration, followed by genomic decay, which is a common evolutionary trend in prophage genomes (34). Of note, we identified homologous sequences between phage-like genes and spacer sequences in the CRISPR loci (Clustered Regularly Interspaced Short Palindromic Repeats) of the decoded bifidobacterial genomes (35), reminiscent of an on-going phage-host arms race.

Overall, predicted prophage gene clusters (representing both remnants and apparently complete genomes) were retrieved from 38 bifidobacterial type strain genomes, indicated that ten type strains appear to lack identifiable prophage content, i.e., *B. adolescentis*, *B. angulatum*, *B. asteroides*, *B. gallinarum*, *B. longum* subsp. *longum*, *B. minimum*, *B. pseudocatenulatum*, *B. pullorum*, *B. thermacidophilum* subsp. *porcinum* and *B. tsurumiense*. Interestingly, five type strain genomes contain a substantial amount of prophage regions, i.e., *B. biavatii*, *B. cuniculi*, *B. longum* subsp. *infantis*, *B. moukalabense* and *B. scardovii*, with five or more putative prophage-related gene clusters. The identified prophage-like elements are integrated in various positions in the genome of the investigated

bifidobacterial strains (Table 1 and see below) and represent 3.2 % of the bifidobacterial pangenome content (Fig. 1). We then focused our analyses on the 60 prophage-like sequences that appear to represent complete phage genomes.

A comparative study was undertaken to determinate putative orthology between the 60 bifidoprophage-derived gene sequences, which resulted in the identification of 1,804 ProCOGs (prophage-specific clusters of orthologous genes). An indication as to how much bifidoprophages contribute to genome variability among bifidobacterial taxa was calculated by comparing the number of ProCOGs with the previously identified 18,435 BifCOGs (22) (bifidobacterial-specific cluster of orthologous genes). This analysis showed that the genomic makeup of the bifidoprophages corresponds to about one tenth of the BifCOGs (Fig. 1), therefore representing a sizable portion of the overall genetic variability detected in the genus *Bifidobacterium* (21, 22), particularly when considering that only those prophage sequences were included that were presumed to represent complete phage genomes.

Prophages are known to contribute to the individuality of bacterial strains, as they often contain many truly unique genes (36) (TUG). The bifidoprophages identified here contributed from 0.3 % to 35.4 % of the TUG detected in individual members of this genus (Fig. 1), confirming that they represent major contributors to genetic diversity of *Bifidobacterium*.

The screening for direct repeats (DR) surrounding the identified prophage-like element sequences allowed the identification of putative phage attachment sites for 37 out of the 60 analyzed bifidoprophages (Table 2). These putative attachment sites were retrieved in 26 bifidobacterial taxa, containing one bifidoprophage per genome, up to four in *B. scardovii* (Bsca1, Bsca2, Bsca3 and Bsca4). Notably, half of the identified attachment sites overlap with tRNA genes for methionine, leucine, and lysine (Table 2). These finding corroborate the previously described attachment sites

within a tRNA gene for methionine in some bifidobacterial prophages (17), which had been considered to be inherently unsuitable for phage genome integration (37).

Genomic organization of bifidoprophages. Database matches allowed a tentative subdivision of the 60 bifidoprophages into functional modules resembling a typical lambdoid phage genome organization, including modules that encode functions involved in lysogeny, DNA replication, DNA packaging, head and tail morphogenesis, and host lysis (38). Overall, the genomic structure and predicted functions suggest that all 60 analyzed bifidoprophages represent members of the *Siphoviridae* family.

Among the identified genomic modules, we looked at conserved genes between the analyzed bifidoprophages (Fig. 2). The most conserved module was the DNA packaging module, in which a consistent level of conservation of the gene constellation was noticed, i.e., terminase (81 %), portal (79 %) and capsid (78 %). In contrast, the most variable functional module was the DNA replication region, containing less conserved genes between genomes. When we focussed on individual genes, the most conserved one was the integrse-encoding gene, which belongs to the lysogeny module and could be retrieved in 90 % of the bifidoprophage genomes. The next most conserved genes were those encoding the endolysin (86 %) found in the host lysis module, and the genes encoding the tape measure protein (83 %) of the tail morphogenesis module (Fig. 2). However, our comparative analyses demonstrated the absence of genes shared between all these mobile elements. In fact, the high heterogeneity between the bifidoprophages is reproduced within the COGs. Even when we focused on the highly conserved genes such as those coding for the integrase or the tail proteins, we still observed a high level of genetic variability (Fig. 1).

Of note, in the predicted lysogeny module of 18 bifidoprophages, we observed the presence of genes encoding putative toxin-antitoxin family proteins, which may be crucial to retain prophage genomes in daughter cells (39) or be involved in defense mechanisms against other phages (40, 41).

Bifidoprophages morphology. To support our metagenomics-based results, we attempted to isolate phage particles from a number of bifidobacterial cultures (a selection of strains whose genomes were predicted to contain intact prophages). Strains representing seven *Bifidobacterium* species (*B. bifidum* LMG 11041, *B. boum* LMG 10736, *B. choerinum*, LMG 10510, *B. longum subsp. suis* LMG 21814, *B. moukalabense* DSM 27321, *B. ruminantium* LMG 21811, and *B. saeculare* LMG 14934) were treated with different concentrations of mitomycin C or hydrogen peroxide. However, for the latter treatment we did not achieve any effect (see below). Reduced growth (suggesting prophage induction) was observed with these seven strains when treated with mitomycin C. The supernatant of these induced cultures was analyzed by transmission electron microscopy (TEM). Complete phage particles were observed for *B. boum*, *B. choerinum*, and *B. moukalabense* (Fig. 3), with significantly higher numbers of released virions for *B. choerinum*, and *B. moukalabense.* All of the observed phage particles exhibit the typical morphology of phages belonging to the *Siphoviridae* family, with small isometric capsids and long, non-contractile tails. Of note, the tail decorations visible for the induced *B. boum* prophages are reminiscent of those found on *Lactococcus lactis* virulent phage 1358 (42, 43). Such tail decorations have previously been postulated to participate in non-specific adsorption of the phage to its host surface polysaccharides (43, 44).

Phylogenetic analysis of bifidoprophages. A systematic dot plot analysis of the 60 *Bifidobacterium* prophage genome sequences was undertaken to highlight possible synteny between them (Fig. S1). The examined bifidobacterial prophage sequences exhibited sequence homology among each other in a patch-wise fashion across the non-structural gene modules. However, with respect to DNA sequences encoding the predicted structural components of the phage particles, five homology groups were distinguished (Fig. 4). Group 1 encompassed seven bifidoprophages (Bsaec1/*B. saeculare*, Bbia1/*B.*

biavatii, Bcat1/*B. catenulatum*, Bdent2/*B. dentium*, Bmou2/*B. moukalabense*, Breu1/*B. reuteri*, and Bsag3/*B. saguini*). These seven prophages represent the most conserved group of bifidoprophages with respect to the synteny of their nucleotide sequences, all containing an identical 16 kb DNA fragment and being integrated at an identical position in a methionine tRNA gene. Interestingly, Group 1 members possess a GC content that is 3.2 % higher than that of their hosts.

Group 2 bifidoprophages are represented by ten members, i.e., Bcun3/*B. cuniculi*, Bcho1/*B. choerinum*, Bcru1/*B. crudilactis*, Bcun4/*B. cuniculi*, Bcun5/*B. cuniculi*, Binf3/*B. longum* subsp. *infantis*, Bmagn2/*B. magnum*, Bpsy1/*B. psychraerophilum*, Bstell2/*B. stellenboschense*, and Bster4/*B. stercoris*. In contrast to Group 1, the genomes of Group 2 bifidoprophages possess an average GC content that is 2.8 % lower than that of their hosts, possess genomes of variable length up to 55,883 base pairs, while also being integrated at different positions. Interestingly, the majority of prophages retrieved in members of the *B. pseudolongum* phylogenetic group (22, 45) are part of Group 2, which encompass *Bifidobacterium* species isolated from rabbit and pig faeces.

Group 3 bifidoprophages are represented by Bact2/*B. actinocoloniiforme*, Bbia2/*B. biavatii*, Bbif1/*B. bifidum*, Bbomb1/*B. bombi*, Bcor1/*B. coryneforme*, Bindi1/*B. indicum*, Binf1/*B. longum* subsp. *infantis*, Binf2/*B. longum* subsp. *infantis*, Bmon1/*B. mongoliense*, Bmou5/*B. moukalabense*, Bsca1/*B. scardovii*, Bsca2/*B. scardovii*, Bster1/*B. stercoris*, and Bsuis2/*B. longum* subsp. *suis*, and include many prophages identified in genomes of *Bifidobacterium* species isolated form insects.

Group 4 and 5 bifidoprophages are heterogeneous with respect to the taxonomic allocation of the corresponding hosts and genome length, although they are co-branching in the clustering tree (Fig. 4).

In addition, we compared the phylogeny of these bifidoprophage sequences with those of their respective bacterial hosts. A phylogenomic analysis based on the core-genome sequences of the genus *Bifidobacterium* (22) revealed significant discrepancies when compared to the evolutionary

development of the prophage-like element sequences (Fig. 4). In fact, our cluster tree of bifidoprophages highlighted a distribution of the analyzed prophages that only partially corresponds to the evolutionary development of their hosts. These findings confirm previous observations indicating that evolution of phages is not subject to the same pressures and mechanisms/limitations as those of their hosts (18), and suggest that bifidoprophage genomes are an example of genetic mosaicism arising from non-homologous recombination events between ancestral sequences following a web-like, rather than a tree-like phylogeny.

Genetic-positioning of bifidoprophages in the phageome. Mosaicism appears to represent a universal feature of phage genomes. Many attempts have been made for virus classification based on sequence data (46, 47). One of these approaches, based on reticulate relationships and displayed as a weighted graph where nodes represent phages (48), has previously been applied to try to capture evolutionary developments among bifidoprophages (18).

Here, we performed a reticulate representation of the evolutionary development of phage gene sequences using the database of 1260 phage genomes retrieved from NCBI as well as the bifidoprophage genomes identified here. We performed the reticulate representation at different levels using COGs as nodes. A genus representation level was obtained by allocating phage COG members that belong to a particular bacterial host genus in a particular group (Fig. 5), followed by a specific phage representation level whereby genes of a given phage were considered as a separate entity (Fig. S2). With respect to the genus representation, we observed that the majority of COGs shared with bifidoprophages belonged to phages infecting the genera *Mycobacterium*, *Burkholderia*, *Enterococcus*, *Lactobacillus*, *Microbacterium* or *Streptococcus* (Fig. 5). Notably, these connections are established with heterogeneous members of the bacterial domain, including both high G+C Gram positive bacteria from the *Actinobacteria* phylum (such as *Mycobacterium* and *Microbacterium*), as well as low G+C

Gram positive bacteria of the *Firmicutes* phylum (such as *Enterococcus*, *Lactobacillus* and *Streptococcus*) and Gram-negative bacteria including members of the *Proteobacteria* phylum (*Burkholderia*). As shown in Figure 5, the bifidoprophages do not cluster within a specific group of phages infecting a characteristic phylum of bacteria. These prophages seem to form a bridge between low and high G+C genome content bacteria.

Modulation in the infant gut by bifidophages. We were interested in exploring the overall occurrence of bifidoprophages in human gut microbiomes and their potential impact on bifidobacterial communities. We used the data of the project entitled "Impact of Antibiotic Administration on the Establishment and Development of Infant Gut Flora" (https://olive.broadinstitute.org/projects/infant_gut_flora_and_antibiotics), which aims to explore the biodiversity of the human gut microbiome of infants. It consists of 3579 microbiomic data sets obtained from 173 infants (with an average of 1.5 Million of paired-end reads per sample), including both healthy and preterm infants affected by necrotizing enterocolitis (NEC). Specifically, we focused on those infant gut microbiomes displaying the highest densities in bifidobacteria, i.e., those from breast-fed healthy subjects (15, 16, 49).

Firstly, we confirmed the presence of *Bifidobacterium* sequences in the datasets of 20 healthy infant gut microbiomes, identifying a high bifidobacterial abundance, which ranged from 0.13 % to 78.65 % of the total filtered microbiome sample-reads (Fig. 6). When we investigated the bifidobacterial (sub)species occurrence, we found an abundance of *B. breve* (37 %), *B. longum* spp. (30 %) and *B. bifidum* (11 %) (Fig. 6), confirming previously identified infant-type bifidobacterial taxa (13).

The filtered microbiome datasets corresponding to these 20 healthy infants were screened against the 60 bifidoprophage genomes and we found their presence, ranging from 0.001 % to 6.65 %

per individual data set (Fig. 6). The majority of the identified bifidoprophage reads appear to belong to specific bifidobacterial taxa that colonize the infant gut (15, 16), such as *B. longum* subsp. *infantis* (32 %) and *B. bifidum* (40 %). In addition, we identified sequences with homology to predicted bifidoprophage genomes found in poorly characterized bifidobacterial (sub)species, such as *B. longum* subsp. *suis* (12 %) and *B. scardovii* (4 %) (Fig. 6).

The availability of microbiomes from 18 infant faecal samples collected at different time points (ranging from five days to 14 months) allowed us to evaluate the dynamics of the bifidobacterial populations and their associated prophages in these microbiomes. Interestingly, we observed that a high abundance of a specific bifidobacterial taxon, e.g. *B. scardovii* or *B. longum* subsp. *suis*, at a particular time point is followed at the next sampling point by a considerably reduction ($\geq$ 41 fold; $p \leq 0.001$) of its relative abundance (Fig. 7). In contrast, the relative abundance of bifidophages Bsca3 and Bsuis2 followed the opposite trend (Fig. 7). These findings suggest that bifidoprophages may enter the lytic life cycle to generate infective phage particles, which then cause alteration of the relative abundance of a specific taxon in the gut microbiota. Our observations are consistent with the "kill the winner" hypothesis, which proposes that when a specific bacterial strain becomes dominant in a particular ecological niche, it will be targeted by phage infection, which will then (re-)establish a microbial equilibrium (50).

Taken together, the observed correlation between the disappearance of a bifidobacterial species and the corresponding increase of specific phages supports the notion that the infant gut microbiota is modulated by phage predation.

Evaluation of the bifidoprophages induced genes in the infant gut microbiomes. Because of the results obtained from the metagenomic datasets we choose to perform a more in depth analysis. To confirm the notion that these bifidoprophages were in fact replicating within the infant gut, we analysed

a number of corresponding metatranscriptome datasets (Fig. 8). This analysis revealed that sequences corresponding to transcribed bifidoprophage genes represent up to 0.2 % of the total filtered reads, and showed that genes belonging to putative DNA packaging modules (15.9 % of bifidoprophage assigned reads) and tail modules (6.6 % of bifidoprophage assigned reads) exhibited the highest level of transcription (Fig. 8).

We focused our analyses in particular on samples belonging to subject 30101, where the highest abundance of a *B. bifidum*-specific bifidoprophage, designated here as phage Bbif1/30101, was detected in the metagenomic analyses (Fig. 8). Thanks to the high frequency of Bbif1/30101-associated metagenomic reads, we were able to reconstruct its genome. In fact, the majority of the transcribed bifidoprophage genes from certain 30101-derived metatranscriptomic data sets correspond to phage Bbif1. Interestingly, the Bbif1/30101 prophage modules exhibiting the highest transcription level were those involved in DNA packaging, head/tail morphogenesis and host lysis. Comparing the transcription levels of *B. bifidum* phage and host genes showed a significant reduction in the latter (5.8 fold reduction from day 19 to 117) at a time point that coincided with the presumed proliferation of the Bbif1/30101 prophage (71.8 fold induction from day 19 to 117), (Fig. 8). These data clearly support the notion that the Bbif/30101 phage was multiplying in the infant gut at the cost of its *B. bifidum* host.

Conclusions

Prophages are considered to be one of the key drivers of genomic diversity and consequently speciation processes in bacteria (34). Prior to the current study, their presence in the genus *Bifidobacterium* appeared to be limited to a small number of genomes (17-19). Here, we have provided an exhaustive catalogue of prophages identified in 48 bifidobacterial type strain taxa belonging to the genus *Bifidobacterium* (21), highlighting a siphophage-specific genomic organization (51). TEM

analyses from supernatants of induced bifidobacterial cultures confirmed the presence of bifidoprophage virions, thus presenting new avenues in phage biology research in the genus *Bifidobacterium*. Notably, we revealed the existence of five bifidoprophage homology groups based on their genome sequences, exhibiting highly conserved prophage genomes as well as groups that are more heterogeneous. The evolutionary development analyses of bifidoprophage gene sequences within the phageome revealed relatedness to phages infecting other *Actinobacteria*, *Firmicutes* as well as Gram negative bacteria. In addition, through an 'omics' survey of infant samples, we imply the existence of replicating bifidophages in this human niche, which participate in the modulation of bifidobacterial populations in the infant gut ecosystem.

Acknowledgments

We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. This work was also financially supported by a FEMS Jensen Award to FT, and by a Ph.D. fellowship (Spinner 2013, Regione Emilia Romagna) to S.D. DvS and FT are members of The APC Microbiome Institute funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant no. SFI/12/RC/2273). DvS is supported by an SFI-funded Investigator award (Grant no. 13/IA/1953). S.M. holds a Tier 1 Canada Research Chair in Bacteriophages. The authors declare that no competing interests exist.

References

1. **Casjens SR.** 2005. Comparative genomics and evolution of the tailed-bacteriophages. Curr Opin Microbiol **8:**451-458.
2. **Bondy-Denomy J, Davidson AR.** 2014. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. J Microbiol **52:**235-242.
3. **Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H.** 2003. Phage as agents of lateral gene transfer. Curr Opin Microbiol **6:**417-424.
4. **Frost LS, Leplae R, Summers AO, Toussaint A.** 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol **3:**722-732.

5. **Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, Mira A.** 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol **7:**828-836.
6. **Menouni R, Hutinet G, Petit MA, Ansaldi M.** 2015. Bacterial genome remodeling through bacteriophage recombination. FEMS Microbiol Lett **362:**1-10.
7. **Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD.** 2013. Rapid evolution of the human gut virome. Proc Natl Acad Sci U S A **110:**12450-12455.
8. **Dalmasso M, Hill C, Ross RP.** 2014. Exploiting gut bacteriophages for human health. Trends Microbiol **22:**399-405.
9. **Abeles SR, Pride DT.** 2014. Molecular bases and role of viruses in the human microbiome. J Mol Biol **426:**3892-3906.
10. **Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA.** 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun **5:**4498.
11. **De Paepe M, Leclerc M, Tinsley CR, Petit MA.** 2014. Bacteriophages: an underestimated role in human and animal health? Front Cell Infect Microbiol **4:**39.
12. **Waller AS, Yamada T, Kristensen DM, Kultima JR, Sunagawa S, Koonin EV, Bork P.** 2014. Classification and quantification of bacteriophage taxa in human gut metagenomes. ISME J **8:**1391-1402.
13. **Ventura M, Turroni F, Motherway MO, MacSharry J, van Sinderen D.** 2012. Host-microbe interactions that facilitate gut colonization by commensal bifidobacteria. Trends Microbiol **20:**467-476.
14. **Bergey DH GM, Whitman WB, Parte AC.** 2012. The actinobacteria. Bergey's manual of systematic bacteriology **5**.
15. **Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D, Marchesi JR, Ventura M.** 2012. Diversity of bifidobacteria within the infant gut microbiota. PLoS One **7:**e36957.
16. **Turroni F, Foroni E, Pizzetti P, Giubellini V, Ribbera A, Merusi P, Cagnasso P, Bizzarri B, de'Angelis GL, Shanahan F, van Sinderen D, Ventura M.** 2009. Exploring the diversity of the bifidobacterial population in the human intestinal tract. Appl Environ Microbiol **75:**1534-1545.
17. **Ventura M, Lee JH, Canchaya C, Zink R, Leahy S, Moreno-Munoz JA, O'Connell-Motherway M, Higgins D, Fitzgerald GF, O'Sullivan DJ, van Sinderen D.** 2005. Prophage-like elements in bifidobacteria: insights from genomics, transcription, integration, distribution, and phylogenetic analysis. Appl Environ Microbiol **71:**8692-8705.
18. **Ventura M, Turroni F, Lima-Mendez G, Foroni E, Zomer A, Duranti S, Giubellini V, Bottacini F, Horvath P, Barrangou R, Sela DA, Mills DA, van Sinderen D.** 2009. Comparative analyses of prophage-like elements present in bifidobacterial genomes. Appl Environ Microbiol **75:**6929-6936.
19. **Ventura M, Turroni F, Foroni E, Duranti S, Giubellini V, Bottacini F, van Sinderen D.** 2010. Analyses of bifidobacterial prophage-like sequences. Antonie Van Leeuwenhoek **98:**39-50.
20. **Sgorbati B, Smiley MB, Sozzi T.** 1983. Plasmids and phages in Bifidobacterium longum. Microbiologica **6:**169-173.

21. **Milani C, Lugli GA, Duranti S, Turroni F, Bottacini F, Mangifesta M, Sanchez B, Viappiani A, Mancabelli L, Taminiau B, Delcenserie V, Barrangou R, Margolles A, van Sinderen D, Ventura M.** 2014. Genomic encyclopedia of type strains of the genus Bifidobacterium. Appl Environ Microbiol **80:**6290-6302.
22. **Lugli GA, Milani C, Turroni F, Duranti S, Ferrario C, Viappiani A, Mancabelli L, Mangifesta M, Taminiau B, Delcenserie V, van Sinderen D, Ventura M.** 2014. Investigation of the evolutionary development of the genus Bifidobacterium by comparative genomics. Appl Environ Microbiol **80:**6383-6394.
23. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.
24. **Kielbasa SM, Wan R, Sato K, Horton P, Frith MC.** 2011. Adaptive seeds tame genomic sequence comparison. Genome Res **21:**487-493.
25. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res **39:**W347-352.
26. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res **42:**D222-230.
27. **Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J.** 2012. PGAP: pan-genomes analysis pipeline. Bioinformatics **28:**416-418.
28. **Enright AJ, Van Dongen S, Ouzounis CA.** 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res **30:**1575-1584.
29. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.** 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13:**2498-2504.
30. **Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B.** 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods **5:**621-628.
31. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754-1760.
32. **Anders S, Pyl PT, Huber W.** 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31:**166-169.
33. **Fortier LC, Moineau S.** 2007. Morphological and genetic diversity of temperate phages in Clostridium difficile. Appl Environ Microbiol **73:**7358-7366.
34. **Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H.** 2003. Prophage genomics. Microbiol Mol Biol Rev **67:**238-276, table of contents.
35. **Briner A LG, Milani C, Duranti S, Turroni F, Gueimonde M, Margolles A, van Sinderen D, Ventura M, Barrangou R.** 2015. Occurrence and diversity of CRISPR-Cas systems in the genus Bifidobacterium. PLoS One **[Epub ahead of print]**.
36. **Bottacini F, Medini D, Pavesi A, Turroni F, Foroni E, Riley D, Giubellini V, Tettelin H, van Sinderen D, Ventura M.** 2010. Comparative genomics of the genus Bifidobacterium. Microbiology **156:**3243-3254.
37. **Williams KP.** 2002. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res **30:**866-875.
38. **Botstein D.** 1980. A theory of modular evolution for bacteriophages. Ann N Y Acad Sci **354:**484-490.
39. **Dziewit L, Jazurek M, Drewniak L, Baj J, Bartosik D.** 2007. The SXT conjugative element and linear prophage N15 encode toxin-antitoxin-stabilizing systems homologous to the tad-ata module of the Paracoccus aminophilus plasmid pAMI2. J Bacteriol **189:**1983-1997.

40. **Labrie SJ, Samson JE, Moineau S.** 2010. Bacteriophage resistance mechanisms. Nat Rev Microbiol **8:**317-327.
41. **Samson JE, Spinelli S, Cambillau C, Moineau S.** 2013. Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system. Mol Microbiol **87:**756-768.
42. **Dupuis ME, Moineau S.** 2010. Genome organization and characterization of the virulent lactococcal phage 1358 and its similarities to Listeria phages. Appl Environ Microbiol **76:**1623-1632.
43. **Spinelli S, Bebeacua C, Orlov I, Tremblay D, Klaholz BP, Moineau S, Cambillau C.** 2014. Cryo-electron microscopy structure of lactococcal siphophage 1358 virion. J Virol **88:**8900-8910.
44. **Auzat I, Droge A, Weise F, Lurz R, Tavares P.** 2008. Origin and function of the two major tail proteins of bacteriophage SPP1. Mol Microbiol **70:**557-569.
45. **Ventura M, Canchaya C, Del Casale A, Dellaglio F, Neviani E, Fitzgerald GF, van Sinderen D.** 2006. Analysis of bifidobacterial evolution using a multilocus approach. Int J Syst Evol Microbiol **56:**2783-2792.
46. **Rohwer F, Edwards R.** 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol **184:**4529-4535.
47. **Pride DT, Wassenaar TM, Ghose C, Blaser MJ.** 2006. Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. BMC Genomics **7:**8.
48. **Lima-Mendez G, Van Helden J, Toussaint A, Leplae R.** 2008. Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol **25:**762-777.
49. **Turroni F, Ventura M, Butto LF, Duranti S, O'Toole PW, Motherway MO, van Sinderen D.** 2014. Molecular dialogue between the human gut microbiota and the host: a Lactobacillus and Bifidobacterium perspective. Cell Mol Life Sci **71:**183-203.
50. **Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Martin-Cuadrado AB, Mira A, Nulton J, Pasic L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F.** 2010. Viral and microbial community dynamics in four aquatic environments. ISME J **4:**739-751.
51. **Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A **96:**2192-2197.

Figure Legends

Figure 1. Pangenome and Cluster of Orthologous Groups (COGs) comparison between bifidoprophages and *Bifidobacterium* genus. Panel **a** shows the comparison between the number of genes from the *Bifidobacterium* genus and bifidoprophages. Panel **b** displays the comparison between the number of bifidobacterial COGs (BifCOGs) and bifidoprophages COGs (ProCOGs). Panel **c** exhibits the percentage of bifidoprophage-associated truly unique genes (TUG) within the bifidobacterial taxa. The numbers indicated on the x-axis correspond to the bifidobacterial taxa as listed in Table 1. Panel **d** illustrates the abundance of the ProCOGs with an identical predicted function. The pie chart on the left shows the percentage of ProCOGs with or without (i.e. hypothetical proteins) function, while the pie chart on the right displays the number of ProCOGs with the same predicted function, emphasizing on the distribution of the conserved genes within the bifidoprophages genomes (as illustrated in Fig. 2). The colors used to identify the ProCOGs reflect the same pattern established for the prophage modules in Fig. 2.

Figure 2. Consensus of the bifidoprophage genome based on the genomic modules identified. Panel **a** shows the consensus genome subdivided in five modules supported by a heat map of the identified genes for every bifidoprophages. The names of the bifidoprophages are indicated on the left margin of the heat map, gene names are displayed on top, while the number of modules identified within each bifidoprophages are indicated on the right-hand margin. Panel **b** displays the abundance of individual functions identified within the bifidoprophages. The first column shows the number of bifidoprophages that encode a particular function listed in the second column, while the third column highlights the relative percentage.

Figure 3. Electron micrograph of prophages induced from bifidobacterial species. Panel **a** displays *B. choerinum* LMG 10510-induced prophage; Panel **b** shows *B. moukalabense* DSM 27321-induced prophage; Panel **c** illustrates *B. boum* LMG 10736- induced prophage. Panel **d** contains the measured dimensions of the identified prophages.

Figure 4. Phylogenetic tree of the bifidoprophages and *Bifidobacterium* genus. Panel **a** shows the genomic alignment-base clustering of the 60 prophages identified within bifidobacterial type strain genomes. The five different groups are highlighted with different colors. Panel **b** displays the supertree of the *Bifidobacterium* genus based on the concatenation of the 411 core-BifCOGs amino acid sequences (22). Each colored dot represents at least one prophage identified in a specific bifidobacterial type strain genome. The colors of the dots correspond to the associated bifidoprophages groups. Whereas, the colors of the branches reflect the *Bifidobacterium* phylogenetic groups (red: *B. asteroides* group, bright green: *B. pseudolongum* group, azure: *B. longum* group, blue: *B. bifidum* group, violet: *B. adolescentis* group, yellow: *B. pullorum* group and dark green: *B. boum* group).

Figure 5. Reticulate representation of evolutionary development of phage gene sequences within *Bifidobacterium* genus. Panel **a** shows the complete map. Panel **b** displays an enlargement related to the bifidoprophages. Each white dot represents a COG, every oval represents all the genes of phages infecting a specific bacterial genus and the lines connect the COGs with the associated genus where the phage genes were identified by the analysis. The green ovals represent the whole phages of the *Bifidobacterium* genus, the red ovals designate the phages that belong to a genus that shared a high number of COGs with the bifidoprophages and the blue ovals denote the phages belonging to the remaining genus.

Figure 6. Relative abundance of bifidoprophages and *Bifidobacterium* (sub)species within the infant gut microbiome. Panel **a** shows the overall abundance of bifidoprophages and *Bifidobacterium* (sub)species reads within the filtered infant gut microbiome samples. On the y axes are represented the percentage of reads identified, while on the x axes are reported the number of the samples (horizontal numbers) and the corresponding time sampling expressed as collecting days (vertical numbers). The bifidoprophages percentage is represented as red-colored bars, while the *Bifidobacterium* (sub)species are colored in blue. Panel **b** exhibits the abundance of prophages belonging to the bifidobacterial type-strain identified and the abundance of the *Bifidobacterium* (sub)species (only the data over 1% are shown).

Figure 7. Abundances of specific bifidoprophages and their host. Panel **a** shows the abundance of phage Bsca3 with respect to the *B. scardovii* host. Panel **b** displays the abundance of phage Bsuis2 in respect of the *B. longum* subsp. *suis* host. The y axes represent the rpkm (reads mapping to the genome per kilobase of transcript per million reads sequenced) of bifidobacterial and prophages reads, while the x axes report the corresponding data of the collecting days.

Figure 8. Bifidoprophages genes expression within the infant gut microbiome. Panel **a** displays the overall expression rate of the bifidoprophages genes in the samples analysed, while panel **b** exhibits the percentage of the genes distribution within the identified reads. Panel **c** shows the reconstructed phage genome of Bbif1/30101, surrounded by pillars representing the number of metetranscriptomics reads matches for every genes. Panel **d** illustrates the rpkm counts of Bbif1/30101 and *B. bifidum* genes within the metatranscriptomics and metagenomics time points of the sample 30101.

Figure S1. Dot plot comparison based on genomic sequence alignments of the 60 identified bifidoprophages.

Figure S2. Reticulate representation of evolutionary development of bifidophage gene sequences within the total phages genomes retrieved from NCBI. Panel **a** shows the complete map, while panel **b** depicts an enlargement related to the bifidoprophages group 1 (Fig. 4). Each white dot represent a COG, every oval represents all the genes of a phage and the lines connect the COGs with the related phage. The green ovals denote the bifidoprophages, while the blue ovals indicate the 1260 phages retrieved from NCBI database.

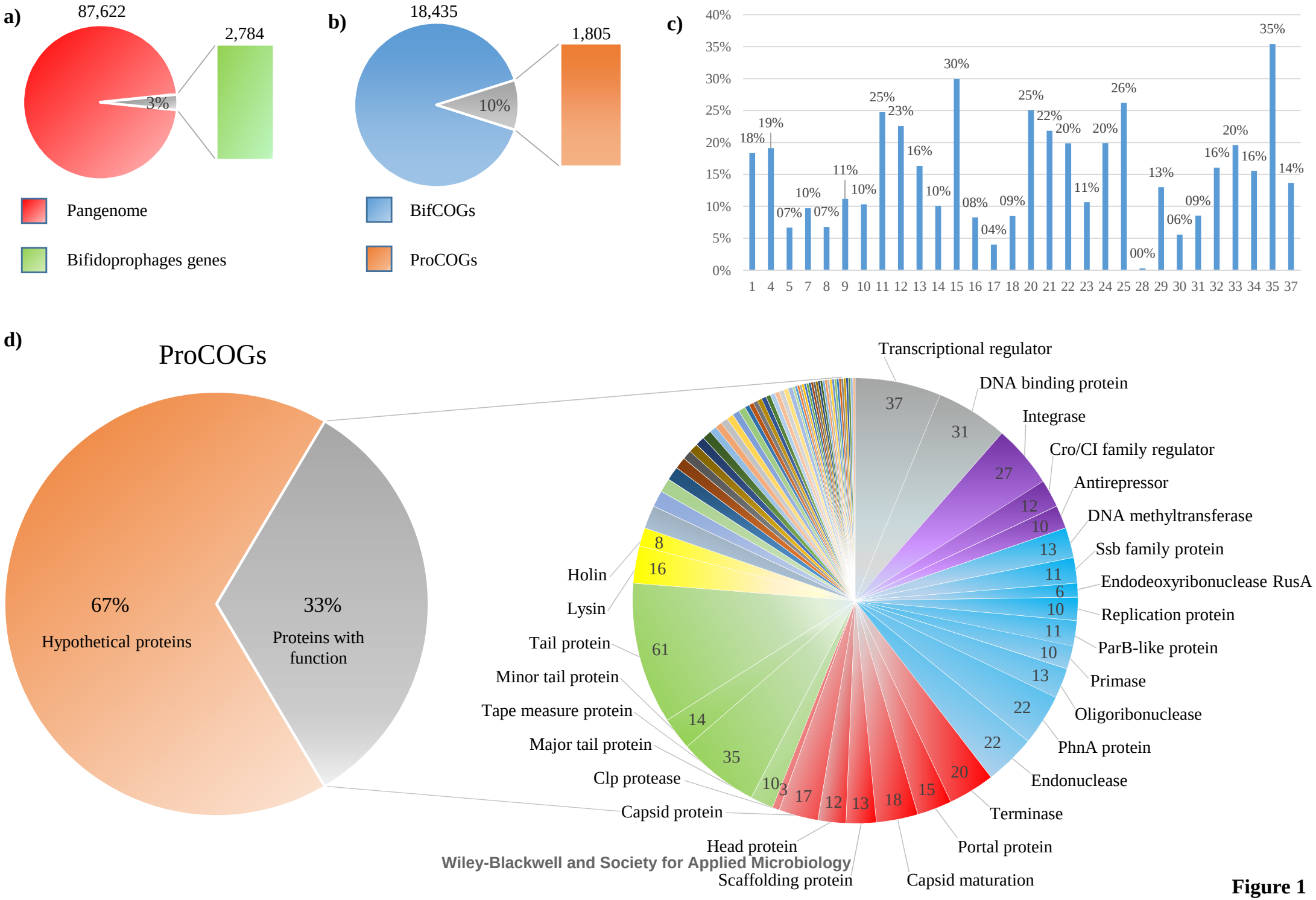
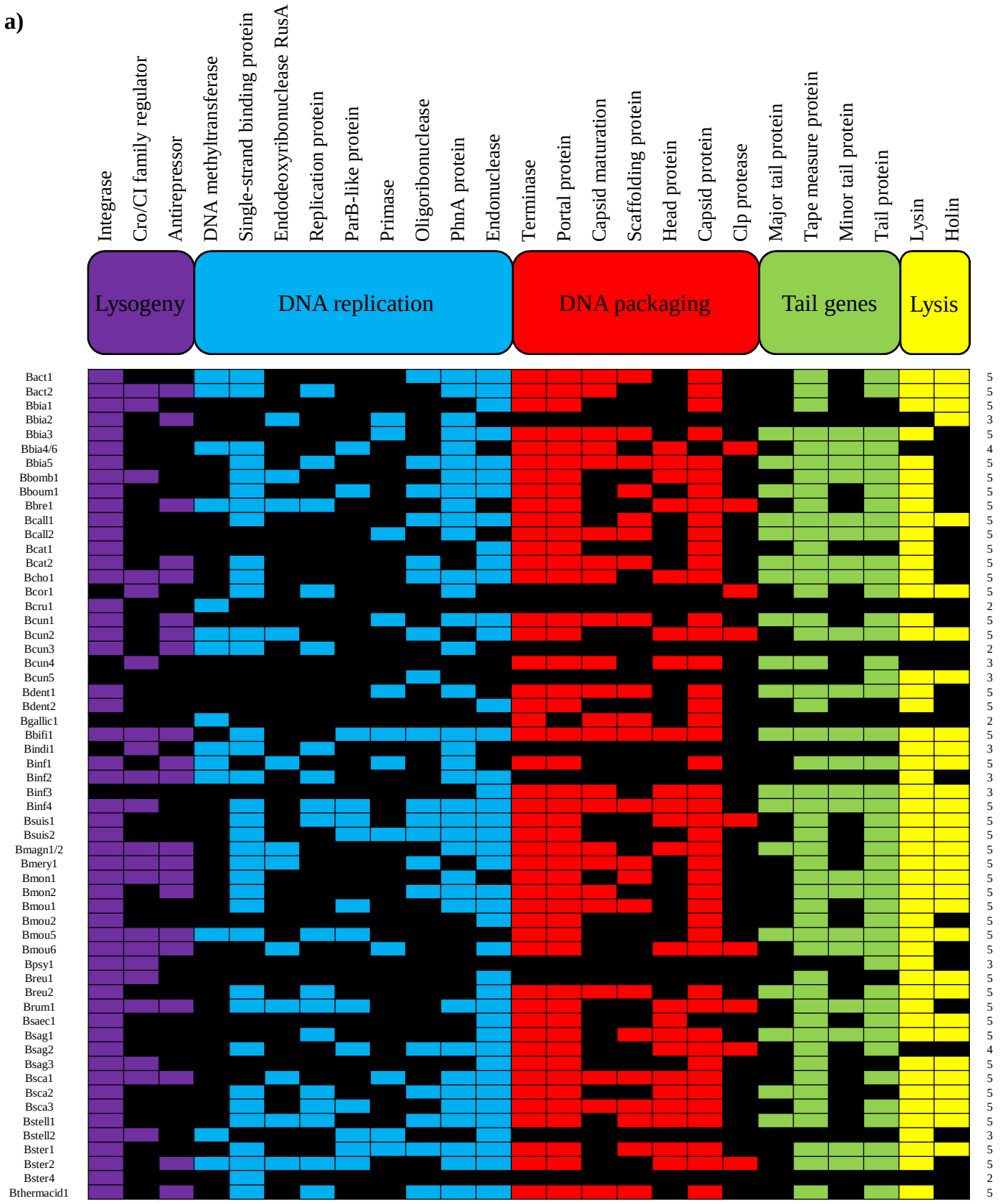


Figure 1

a)

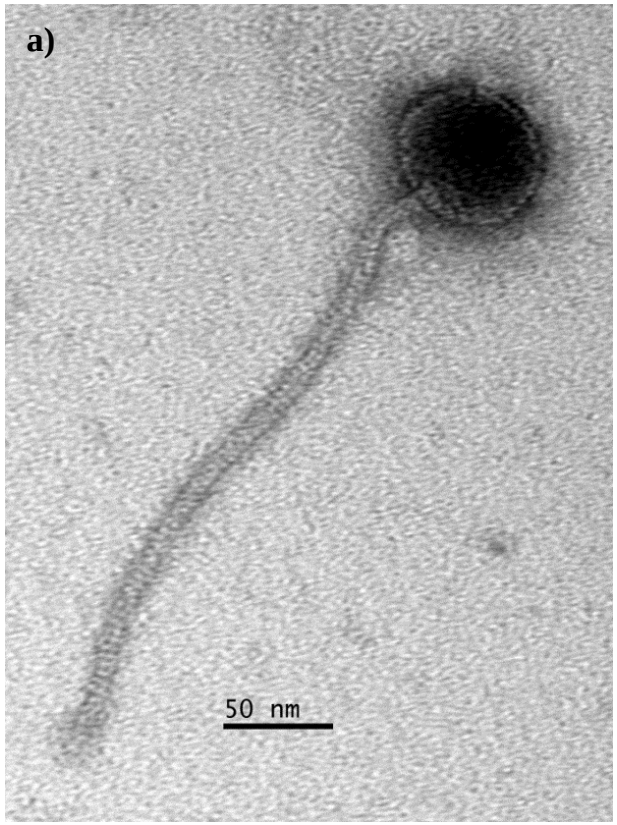


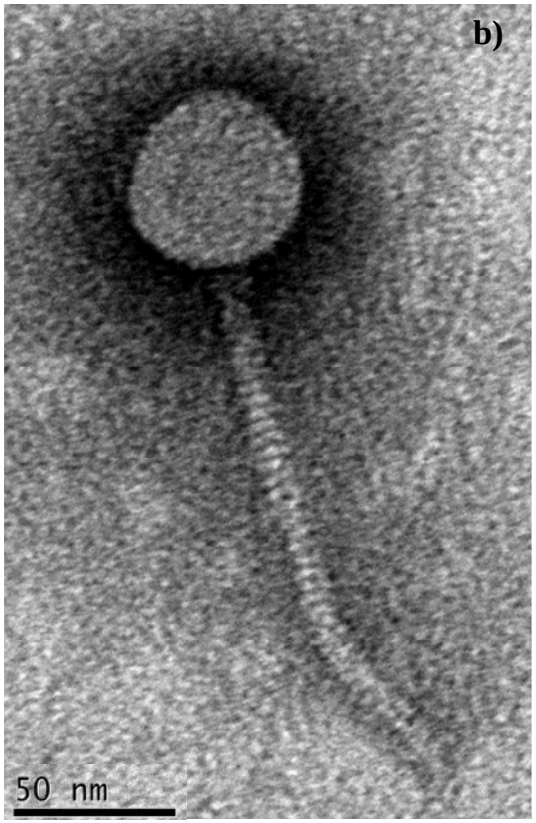
b)

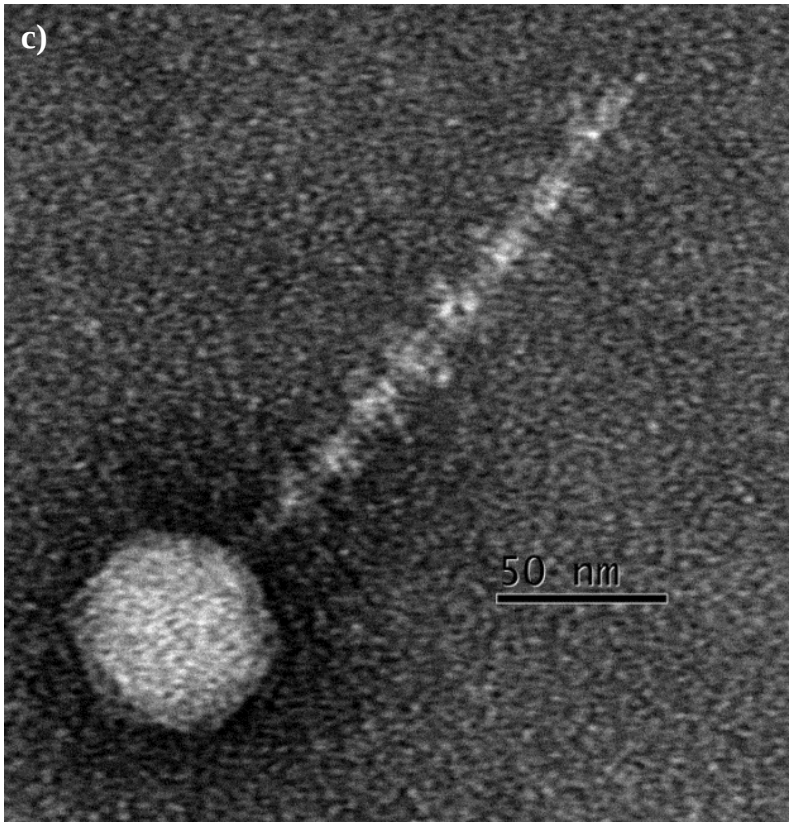
52	Integrase	90%
50	Lysin	86%
48	Tape measure protein	83%
47	Terminase	81%
46	Portal protein	79%
45	Capsid protein	78%
44	Tail protein	76%
41	Endonuclease	71%
36	PhnA protein	62%
35	Single-strand binding family protein	60%
31	Holin	53%
23	Capsid maturation	40%
23	Head protein	40%

22	Scaffolding protein	38%
22	Minor tail protein	38%
21	Cro/CI family transcriptional regulator	36%
21	Antirepressor	36%
19	Oligoribonuclease	33%
19	Major tail protein	33%
18	Replication protein	31%
14	DNA methyltransferase	24%
14	ParB-like protein	24%
12	Crossover junction endodeoxyribonuclease RusA	21%
12	Primase	21%
9	Clp protease	16%

Figure 2







d)

Phage	Capsid (nm)	Tail length (nm)	Tail width (nm)
B. choerinum	71.9 +/- 4.0	352.7 +/- 17.9	10.6 +/- 1.2
B. moukalabense	65.8 +/- 1.1	171.7 +/- 3.0	10.3 +/- 0.5
B. boum	63.5 +/- 0.5	176.9 +/- 3.7	21.7 +/- 1.1

Figure 3

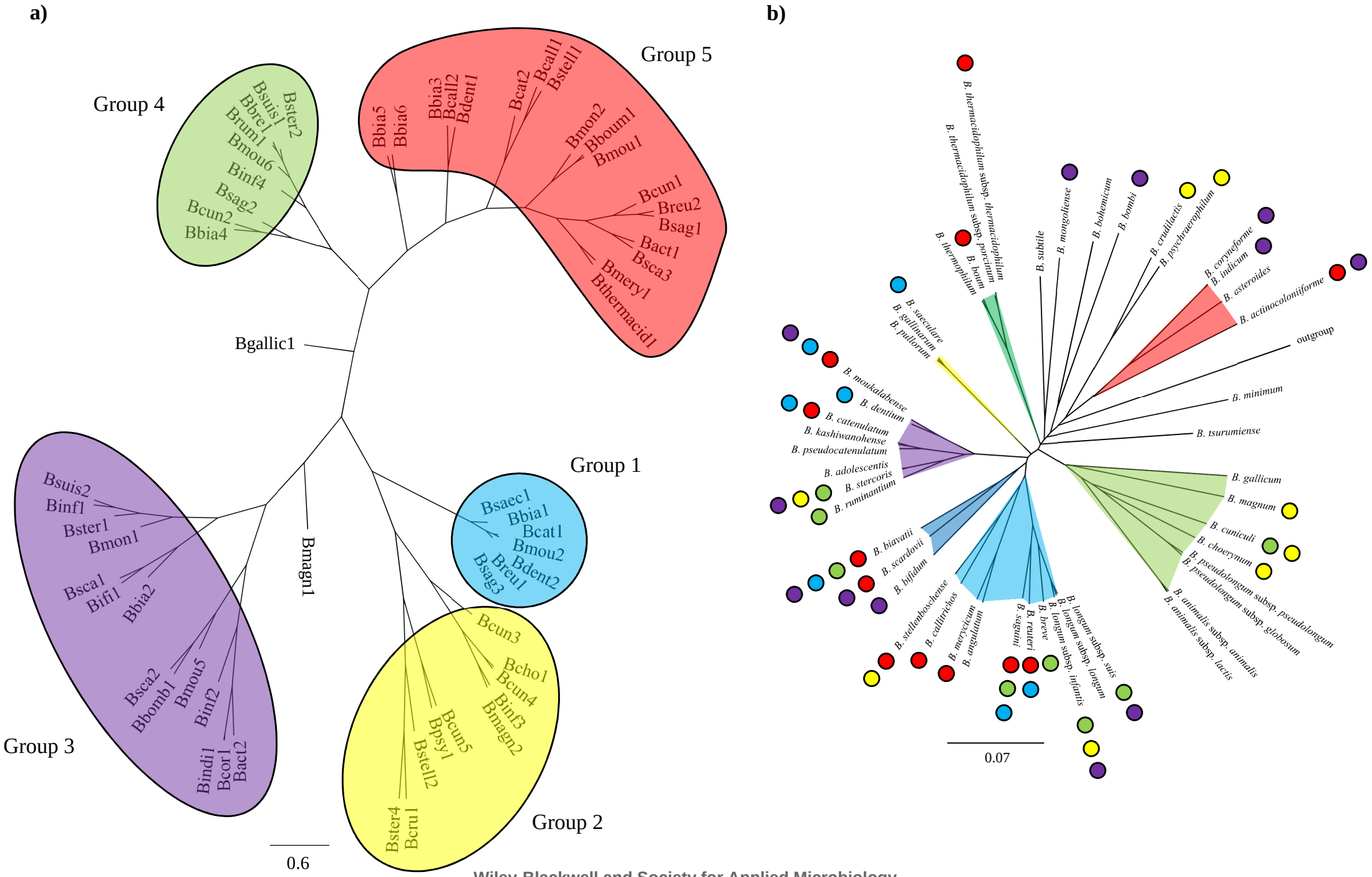

a)
Group 4
Group 5
Group 1
Group 3
Group 2
Bgallic1
Bmagn1
Bsuis1
Bbre1
Brum1
Bmou6
Binf4
Bsag2
Bcun2
Bbia4
Bster2
Bbia5
Bbia6
Bbia3
Bcall2
Bdent1
Bcat2
Bcall1
Bstell1
Bmon2
Bboum1
Bmou1
Bcun1
Breu2
Bsag1
Bact1
Bsca3
Bmery1
Bthermacid1
Bsuis2
Binf1
Bster1
Bmon1
Bsca1
Bifi1
Bbia2
Bsca2
Bbomb1
Bmou5
Binf2
Bindi1
Bcor1
Bact2
Bsaec1
Bbia1
Bcat1
Bmou2
Bdent2
Breu1
Bsag3
Bcun3
Bcho1
Bcun4
Binf3
Bmagn2
Bcun5
Bpsy1
Bstell2
Bster4
Bcru1
0.6
b)
B. thermacidophilum subsp. thermacidophilum
B. thermacidophilum subsp. porcinum
B. thermophilum
B. boum
B. subtile
B. mongoliense
B. bohemicum
B. bombi
B. cradilactis
B. psychraerophilum
B. coryneforme
B. indicum
B. asteroides
B. actinocolonitiforme
outgroup
B. minimum
B. tsurumiense
B. saeculare
B. gallinarum
B. pullorum
B. moukalabense
B. dentium
B. catenulatum
B. kashiwanohense
B. pseudocatenulatum
B. adolescentis
B. stercoris
B. ruminantium
B. biavatii
B. scardovii
B. bifidum
B. stellenboschense
B. callitrichos
B. merycicum
B. angulatum
B. saguini
B. reuteri
B. breve
B. longum subsp. suis
B. longum subsp. longum
B. longum subsp. infantis
B. gallicum
B. magnum
B. cuniculi
B. choerynum
B. pseudolongum subsp. pseudolongum
B. pseudolongum subsp. globosum
B. animalis subsp. animalis
B. animalis subsp. lactis
0.07

Figure 4

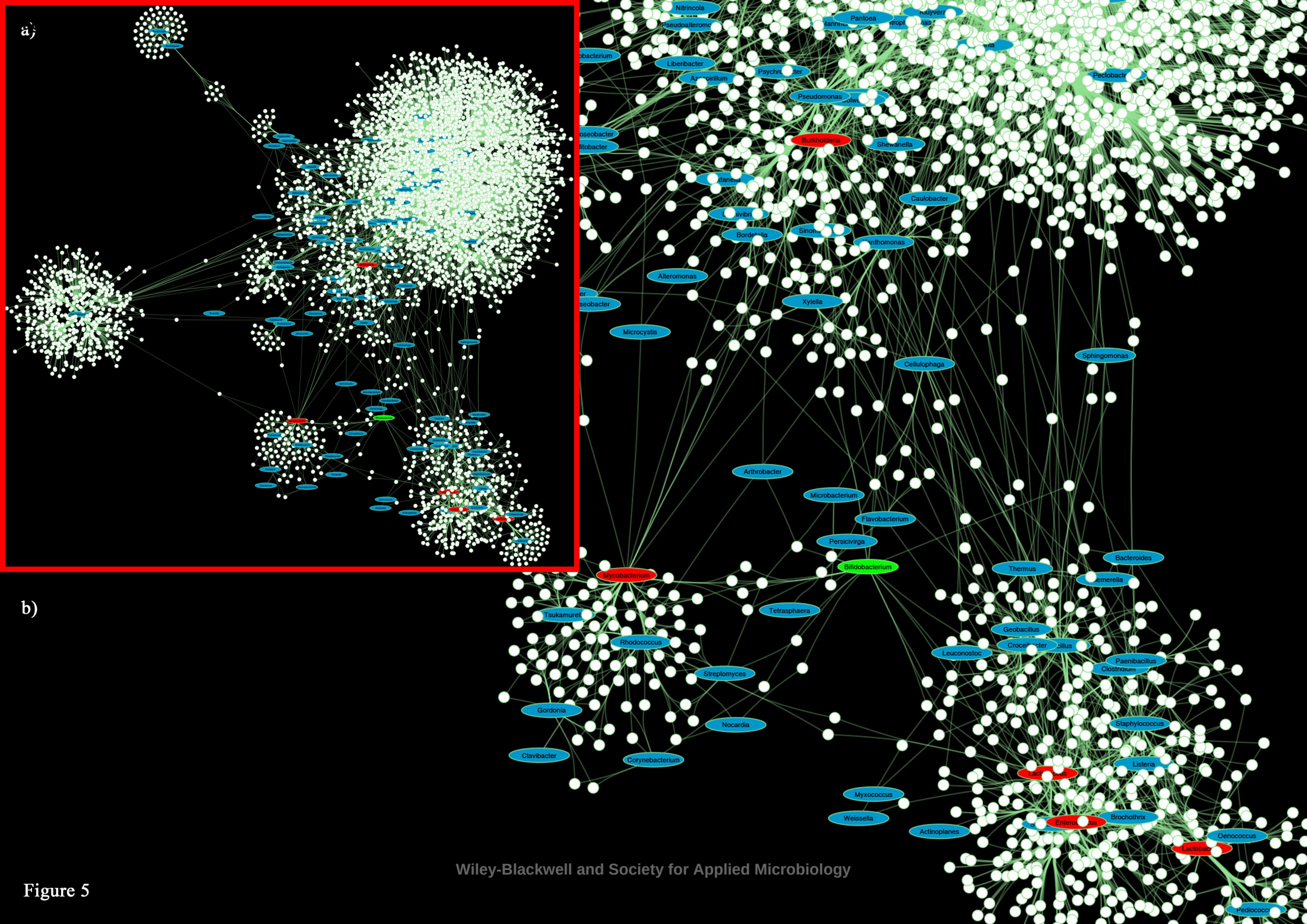


Figure 5

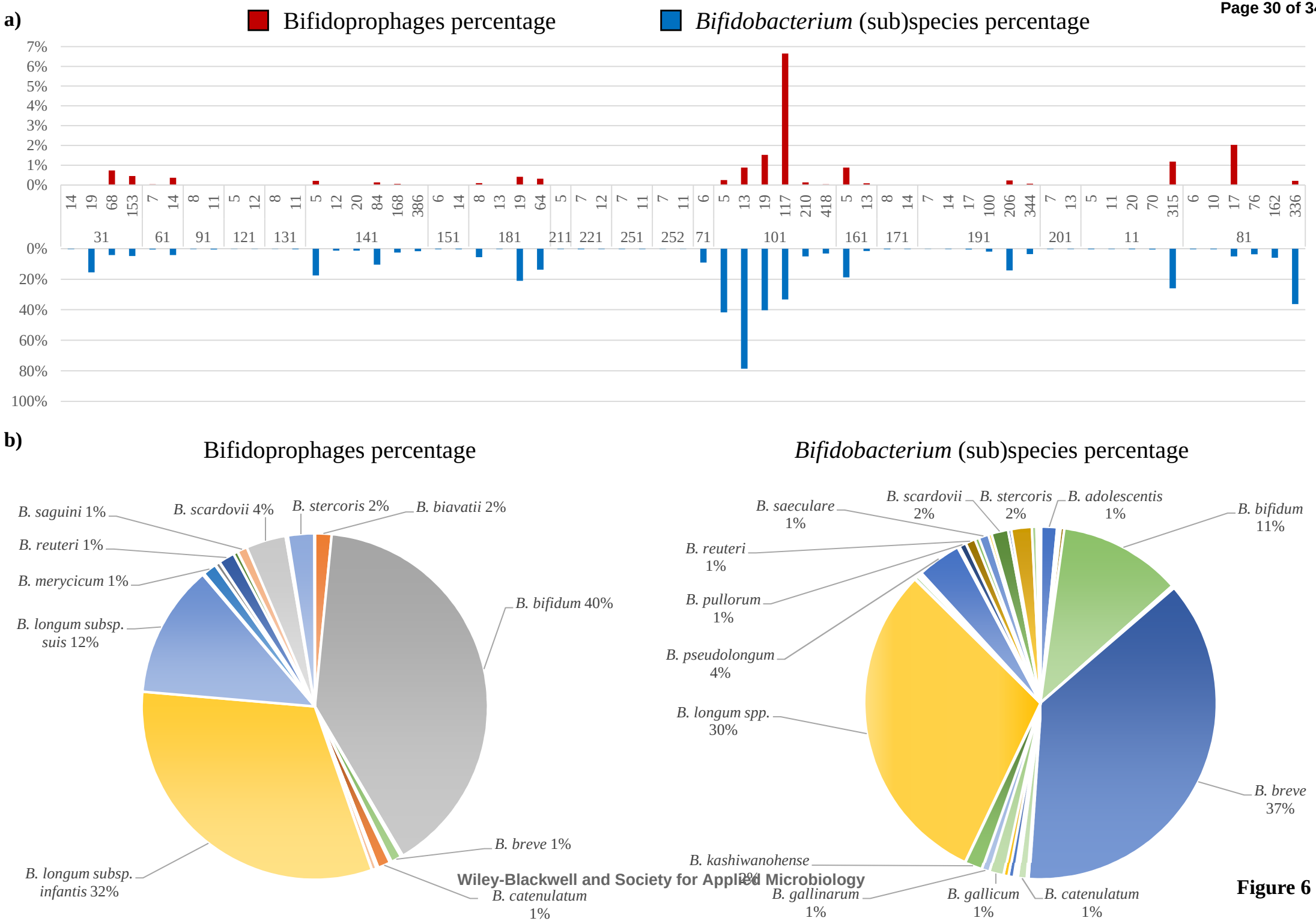
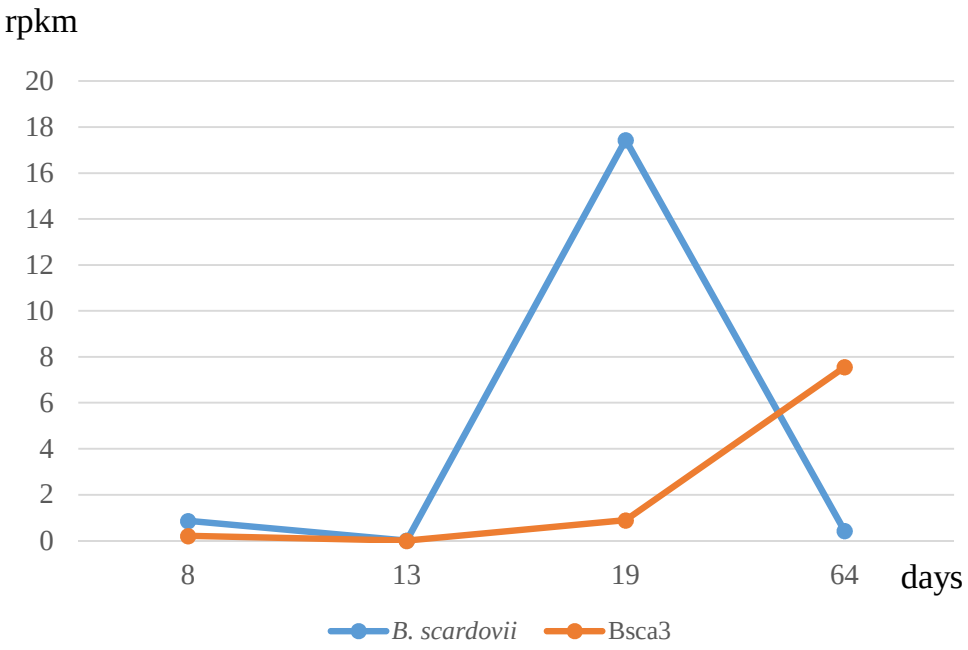

a)
Bifidoprophages percentage
Bifidobacterium (sub)species percentage
7%
6%
5%
4%
3%
2%
1%
0%
0%
20%
40%
60%
80%
100%
14 19 68 153 7 14 8 11 5 12 8 11 5 12 20 84 168 386 6 14 8 13 19 64 5 7 12 7 11 7 11 6 5 13 19 117 210 418 5 13 8 14 7 14 17 100 206 344 7 13 5 11 20 70 315 6 10 17 76 162 336
31 61 91 121 131 141 151 181 211 221 251 252 71 101 161 171 191 201 11 81
b)
Bifidoprophages percentage
B. saguini 1%
B. scardovii 4%
B. stercoris 2%
B. biavatii 2%
B. reuteri 1%
B. merycicum 1%
B. bifidum 40%
B. longum subsp. suis 12%
B. breve 1%
B. longum subsp. infantis 32%
B. catenulatum 1%
Bifidobacterium (sub)species percentage
B. saeculare 1%
B. scardovii 2%
B. stercoris 2%
B. adolescentis 1%
B. bifidum 11%
B. reuteri 1%
B. pullorum 1%
B. pseudolongum 4%
B. longum spp. 30%
B. breve 37%
B. kashiwanohense 2%
B. gallinarum 1%
B. gallicum 1%
B. catenulatum 1%

Figure 6

a)


rpkm
20
18
16
14
12
10
8
6
4
2
0
8
13
19
64
days
B. scardovii
Bsca3

b)

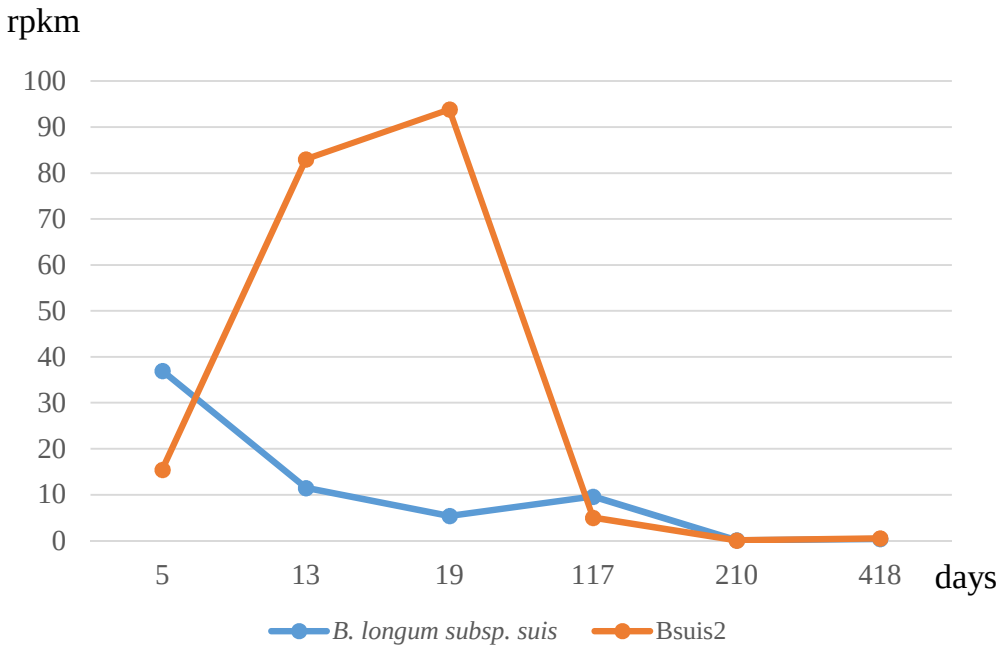
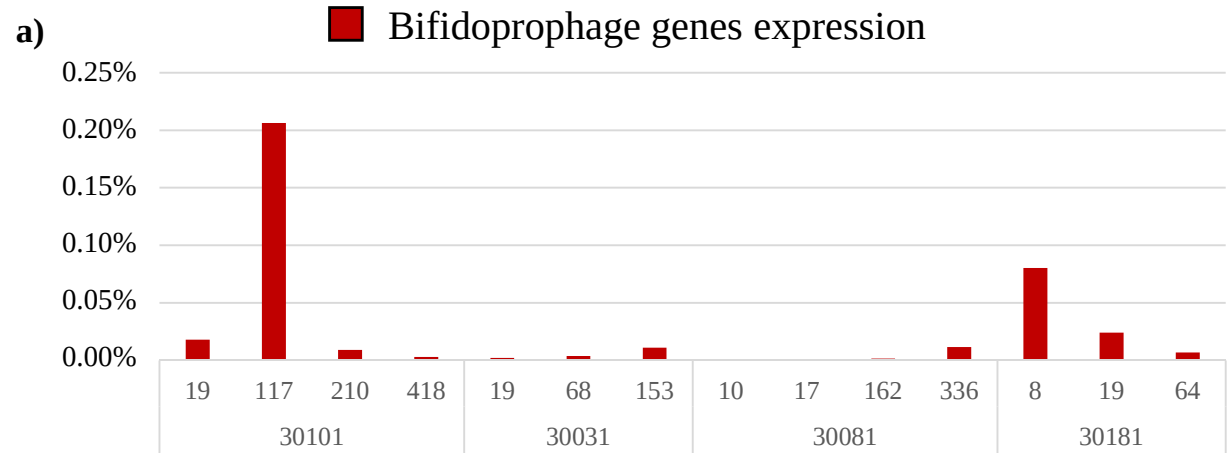
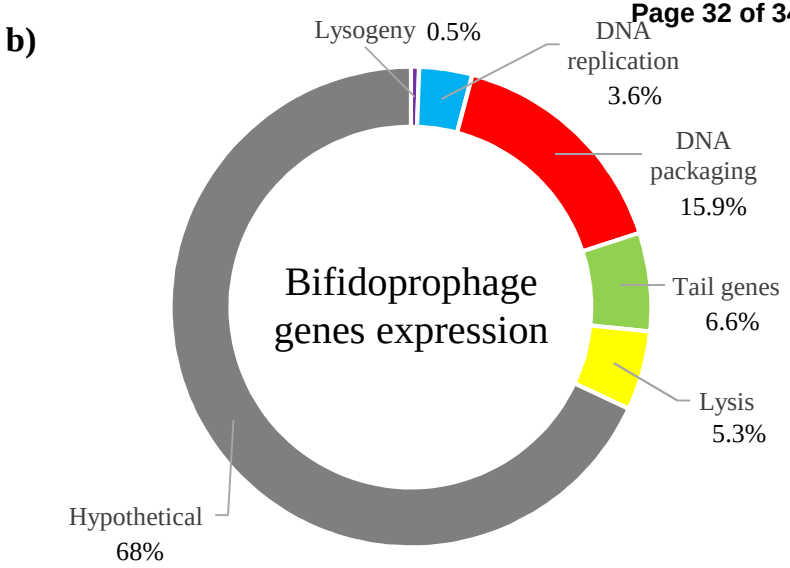
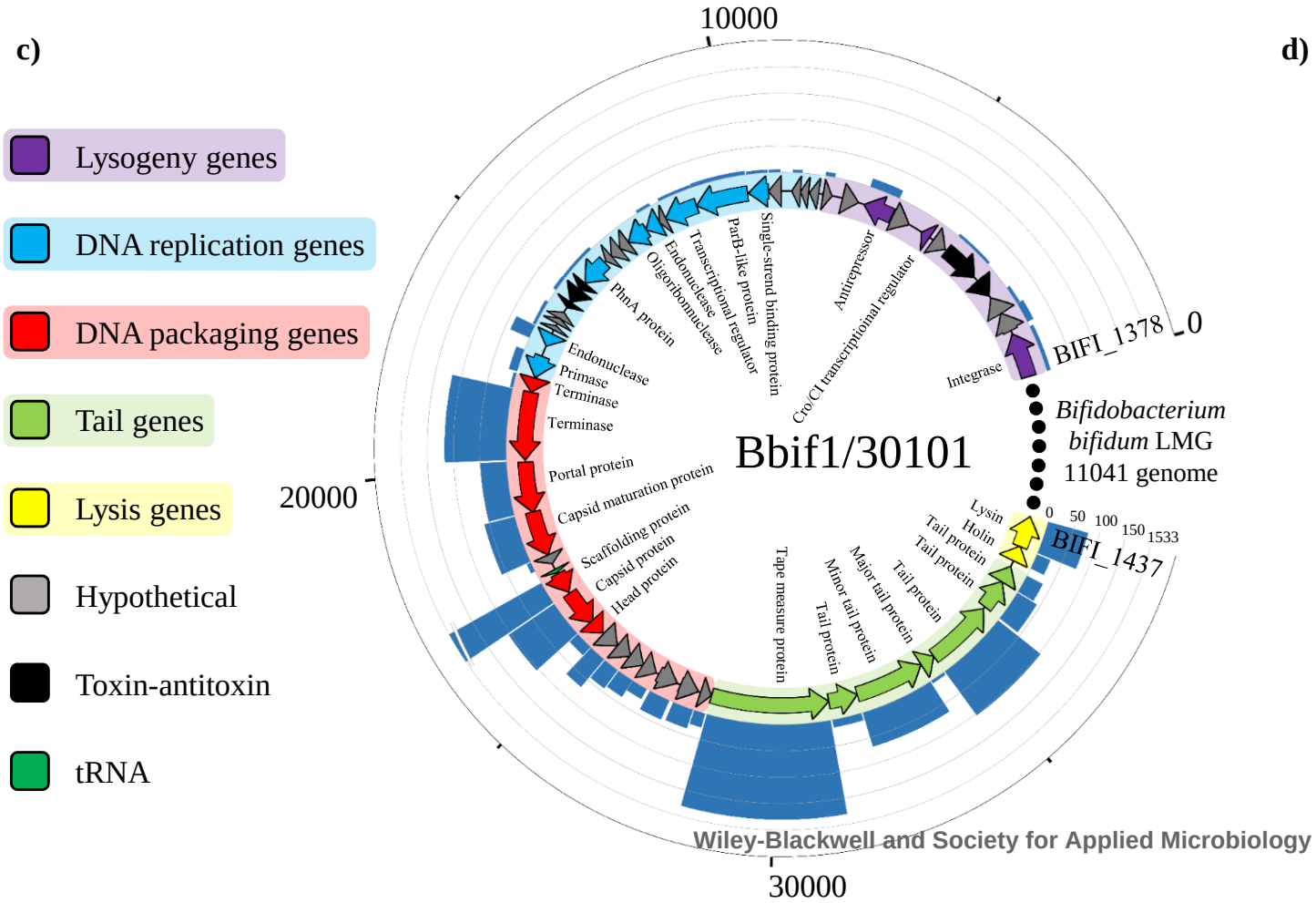

rpkm
100
90
80
70
60
50
40
30
20
10
0
5
13
19
117
210
418
days
B. longum subsp. suis
Bsuis2

Figure 7







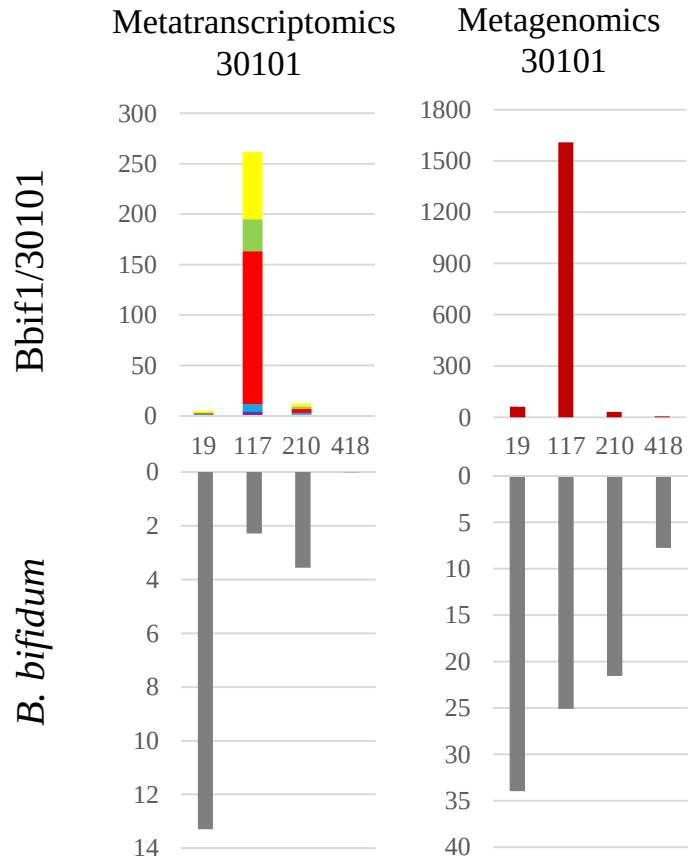


Figure 8

Table 1. List of Bifidoprophage genomes

Number	*Bifidobacterium* Strains	Name	Status	Insertion site	CDS	start	end	Size	GC content	Groups*
1	*B. actinocoloniiforme* DSM 22766	Bact1	complete	-	44	225563	254948	29386	62.1%	5
		Bact2	complete	-	55	914776	953668	38893	62.2%	3
2	*B. animalis* subsp. *animalis* LMG 10508	Bani1	remnant	Lys	9	1364996	1372772	7777	57.3%	-
3	*B. animalis* subsp. *lactis* DSM 10140	Blact1	remnant	-	14	1362362	1370374	8013	59.5%	-
		Bbia1	complete	Met	26	426431	443273	16843	62.1%	1
		Bbia2	complete	-	29	836239	852339	16101	59.7%	3
		Bbia3	complete	Pro	40	1858213	1888649	30431	63.5%	5
4	*B. biavatii* DSM 23969	Bbia4	complete	Gly	37	2365885	2385985	20101	61.8%	4
		Bbia5	complete	-	60	2693965	2731263	37299	61.5%	5
		Bbia6	complete	-	15	3155833	3169736	13904	62.2%	5
		Bbia7	remnant	Ser	5	826233	829712	3480	61.1%	-
5	*B. bifidum* LMG 11041	Bbif1	complete	-	57	1706117	1745146	39030	64.0%	3
6	*B. bohemicum* DSM 22767	Bboh1	remnant	-	6	1152626	1160863	8238	56.2%	-
		Bboh2	remnant	Ala	7	1831230	1836651	5422	49.9%	-
		Bbomb1	complete	Glu	52	1273586	1310010	36425	57.3%	3
7	*B. bombi* DSM 19703	Bbomb2	remnant	-	10	110636	118526	7891	52.7%	-
		Bbomb3	remnant	Lys	18	1741998	1752103	10106	56.0%	-
8	*B. boum* LMG 10736	Bboum1	complete	Pro	51	368586	404862	36277	62.8%	5
9	*B. breve* LMG 13208	Bbre1	complete	-	66	664810	705420	40611	59.5%	4
		Bcall1	complete	-	57	157317	198687	41371	63.5%	5
10	*B. callitrichos* DSM 23973	Bcall2	complete	Arg	38	2043766	2074223	30458	62.6%	5
		Bcall3	remnant	Val	11	1383388	1391190	7803	61.7%	-
11	*B. catenulatum* LMG 11043	Bcat1	complete	Met	23	475181	491377	16197	61.2%	1
		Bcat2	complete	-	56	1724216	1764489	40274	59.3%	5
12	*B. choerinum* LMG 10510	Bcho1	complete	-	74	1067765	1123544	55780	59.0%	2
		Bcho2	remnant	Lys	11	273992	281386	7395	59.6%	-
13	*B. coryneforme* LMG 18911	Bcor1	complete	-	31	1213783	1235097	21315	61.8%	3
		Bcor2	remnant	-	9	1088330	1093162	4833	58.3%	-
14	*B. crudilactis* LMG 23609	Bcru1	complete	-	27	600108	624982	24875	56.1%	2
		Bcun1	complete	-	60	248050	287221	39172	62.7%	5
		Bcun2	complete	Leu	56	422953	462188	39236	64.4%	4
15	*B. cuniculi* LMG 10738	Bcun3	complete	-	44	622569	645806	23238	57.5%	2
		Bcun4	complete	-	25	2092859	2125020	32162	57.7%	2
		Bcun5	complete	-	65	889024	944906	55883	62.1%	2
		Bdent1	complete	-	26	1629039	1655491	26453	62.9%	5
16	B. dentium LMG 11045	Bdent2	complete	Met	24	681010	697366	16357	65.3%	1
		Bdent3	remnant	-	7	707220	714810	7591	61.8%	-
17	*B. gallicum* LMG 11596	Bgallic1	complete	Pro	21	1315403	1327851	12449	58.9%	-
18	*B. indicum* LMG 11587	Bindi1	complete	-	25	1193122	1206638	13517	61.5%	3
19	*B. kashiwanohense* DSM 21854	Bkas1	remnant	-	13	1541485	1555844	14360	56.5%	-
		Binf1	complete	-	56	1288652	1331024	42373	61.2%	3
		Binf2	complete	Ser	66	1660928	1692072	31145	55.7%	3
20	*B. longum* subsp. *infantis* ATCC 15697	Binf3	complete	Thr	28	1694202	1722246	28045	61.8%	2
		Binf4	complete	Leu	61	1956374	1995650	39277	61.1%	4
		Binf5	remnant	-	12	1806020	1813371	7352	60.5%	-
21	*B. longum* subsp. *suis* LMG 21814	Bsuis1	complete	Leu	69	53319	97447	44129	59.2%	4
		Bsuis2	complete	-	59	2054138	2095171	41034	59.7%	3
		Bmagn1	complete	-	32	1557476	1572074	14599	54.5%	-
22	*B. magnum* LMG 11591	Bmagn2	complete	-	44	1751714	1790128	38415	55.1%	2
		Bmagn3	remnant	Gly	9	1140493	1146150	5658	53.9%	-
23	*B. merycicum* LMG 11341	Bmery1	complete	-	54	1455783	1490863	35081	62.8%	5
		Bmon1	complete	His	53	144318	183132	38815	57.6%	3
24	*B. mongoliense* DSM 21395	Bmon2	complete	Val	54	1310262	1347044	36783	58.3%	5
		Bmon3	remnant	Ala	8	1147678	1152336	4659	59.5%	-
		Bmou1	complete	-	54	252163	286222	34060	64.2%	5
		Bmou2	complete	Met	24	2180947	2197325	16379	65.1%	1
		Bmou3	remnant	-	10	969802	976742	6941	56.9%	-
25	*B. moukalabense* DSM 27321	Bmou4	remnant	-	12	1034015	1040521	6507	61.5%	-
		Bmou5	complete	-	59	920776	964425	43650	56.4%	3
		Bmou6	complete	-	58	-	-	40812	60.9%	4
26	*B. pseudolongum* subsp. *globosum* LMG 11569	Bglob1	remnant	-	5	734229	740915	6687	61.3%	-
27	*B. pseudolongum* subsp. *pseudolongum* LMG 11571	Bpseudolon1	remnant	Gly	17	1447010	1454767	7758	60.5%	-
28	*B. psychraerophilum* LMG 21775	Bpsy1	complete	Ala	14	271189	285076	13888	62.3%	2
		Bpsy2	remnant	Arg	14	1476596	1485370	8775	57.4%	-
		Breu1	complete	Met	26	2550146	2566654	16509	60.8%	1
29	*B. reuteri* DSM 23975	Breu2	complete	-	59	1399623	1437870	38248	63.2%	5
		Breu3	remnant	-	8	1179887	1187590	7704	60.7%	-
30	*B. ruminantium* LMG 21811	Brum1	complete	Val	55	459174	497679	38506	58.7%	4
31	*B. saeculare* LMG 14934	Bsaec1	complete	Met	25	1016440	1032693	16254	65.3%	1
		Bsaec2	remnant	Ala	7	1474124	1478875	4752	68.2%	-
		Bsag1	complete	-	67	984189	1022881	38803	61.3%	5
32	*B. saguini* DSM 23967	Bsag2	complete	-	50	2073261	2105479	32219	61.3%	4
		Bsag3	complete	Met	30	2250240	2267938	17699	60.6%	1
		Bsag4	remnant	-	5	1921131	1925600	4470	59.0%	-
		Bsca1	complete	-	57	965412	1005561	40150	63.3%	3
		Bsca2	complete	-	58	1263375	1301227	37853	62.4%	3
33	*B. scardovii* LMG 21589	Bsca3	complete	Gly	59	1333585	2350337	37870	63.5%	5
		Bsca4	remnant	Leu	13	225740	233376	7637	59.6%	-
		Bsca5	remnant	Pro	6	1140946	1147234	6289	63.8%	-
34	*B. stellenboschense* DSM 23968	Bstell1	complete	-	74	741502	780656	39155	66.2%	5
		Bstell2	complete	Ser	28	2213171	2240194	27024	61.7%	2
		Bster1	complete	-	59	188417	228941	40525	62.5%	3
35	*B. stercoris* DSM 24849	Bster2	complete	Leu	68	2045824	2086608	40785	58.8%	4
		Bster3	remnant	-	11	1951629	1958689	7061	59.0%	-
		Bster4	complete	-	42	258482	280479	21998	58.4%	2
36	*B. subtile* LMG 11597	Bsub1	remnant	Gln	8	611714	617744	6031	52.9%	-
		Bsub2	remnant	-	12	2368132	2380675	12544	66.6%	-
37	*B. thermacidophilum* subsp. *thermacidophilum* LMG 21395	Bthermacid1	complete	-	55	857355	891159	33805	63.3%	5
		Bthermacid2	remnant	Ser	5	70308	76708	6401	61.1%	-
38	*B. thermophilum* JCM 1207	Bthermop1	remnant	Ser	13	472391	481766	9376	60.3%	-

* bifidoprophages phylogenetic groups (see also Figure 4)

Table 2. Putative insertion sites

Bifidobacterium Strains	Bifidoprophages	attB	tRNA insertion sites
B. animalis subsp. *animalis* LMG 10508	Bani1	GTGCCCCCCCAGGGATTCGAACCCTGAACCCACGACTTA	Lys
B. animalis subsp. *lactis* DSM 10140	Blact1	GGGGTTCAATCCCCCGCGGCTCCAC	-
B. biavatii DSM 23969	Bbia1	CATGGTTCAAATCCATGCCCCGCTAC	Met
B. bifidum LMG 11041	Bbif1	TATAACCACCGTT	-
B. bombi DSM 19703	Bbomb2	CGTCGGACTCGAACCGACAACCCACGACTT	Lys
	Bbomb3	CCACGGGGATTCGAACCCCGGA	Lys
B. boum LMG 10736	Bboum1	AATCCTGTCAGCCCGACCG	Pro
B. breve LMG 13208	Bbre1	CGGGGTTCGATTCCCCGCGACTCCAC	-
B. callitrichos DSM 23973	Bcall1	AGAGGCGTGAAACGGCGGTATAAAGCCGATTTAT	-
	Bcall3	TCGTTGTCACGATGTTGTCACAGCGCCATCCTACGGGACGGATAAGGCCCGGAATCGTTGGGATTCCGGGCCTT	Val
B. catenulatum LMG 11043	Bcat1	CAAATCCATGCCCCGCTACCAAT	Met
B. choerinum LMG 10510	Bcho1	GATACCACAATTAAACAGCACGGTTGCGTTTCCCTACGCTTGTTGGGATGATTCGTTAAATCGAATCGCGCGACA GTGTTCCCCGCATCACAGCGGGGATGACCCCCTAGATTGTTTGATGCAATCGTGTTCCCCGCATCACAGCGGGGT GGGTTTCAGAGGGCGGAGCCGAGGCGCCGCTCTCAACTAACGGGGCTGGCGGCCTGAGTGGCTGCCGGCCCCGT TTTTTCGTATCCAGACGGTGTGTTGTGCTTGTCTTGGCGGTTATCCGGTCGCGTGAACGGTGTTGATCTGTCCGAT GCGTGTTTTCGTTAAACATTCATTAACTCTTCACGTCGTATGCGAGCGGTGCTAGAACCGGATGATTCTCGTGAA TGTTGCAATCATTGGGTTTTAGCCGTGTTCTTACGTGTCGGACGACCGTCTTGATTGGATAACACCATCGTA	-
	Bcho2	GCTTTTATCGCTGAATCAGTCCCTAT	Lys
B. cuniculi LMG 10738	Bcun1	AGTTCAAGTTCAACGT	-
	Bcun2	GGCGTGCGGGTTCAAGTCCCGCTCCGGACACCGCCAA	Leu
B. dentium LMG 11045	Bdent2	GTTAGCTCAGTCGGTTAGAGCA	Met
	Bdent3	AACCCGTCAACG	-
B. gallicum LMG 11596	Bgallic1	TGTCGTAAAGAATATTAGTGGGGATTAATGCAGCGCGTGACCCGTTCATGACTGAGGAAAGACCACAAGAAAGC CGCCGTTGTTCCGAATCGTTCCGCTAACTGATTTCAAACAATATACCGACT	Pro
B. longum subsp. *infantis* ATCC 15697	Binf1	TCTCAGCTCAGTGCGTCTTTTAACTTGC	-
	Binf4	AAGGTGCCTCCGGTGGGACTCGAACCC	Leu
B. longum subsp. *suis* LMG 21814	Bsuis1	GTGCCTCCGGTGGGACTCGAACCC	Leu
B. magnum LMG 11591	Bmagn3	ATAGAGAGCGGATGACGGGAATCGAA	Gly
B. merycicum LMG 11341	Bmery1	GAGTTCAAGTTCAACGTCTCCAAGTAGCGGTTTCCGCCATCTGGAAACGGCTTGTTTTCAACGGTTTTGAGCTTC CCGACCGTTGGAAACGAGCCGTTTTTGAATGCGCTCGGAATGGCGGTTTCTGCCACGTCTCGGGCGGCTGTTGC CATCGCTTGAGT	-
B. moukalabense DSM 27321	Bmou1	AAAGCCGCCATTTCCGGCGGCTTTCGAATGGTGGAGCTGCGGGGAATCGAACCCC	-
	Bmou2	GTTAGCTCAGTCGGTTAGAGCA	Met
B. pseudolongum subsp. *pseudolongum* LMG 11571	Bpseudolon1	TTTGTTATCAATTTGTTATCACGCGGCC	Gly
B. psychraerophilum LMG 21775	Bpsy2	ATCTCAAGGTCGACGGTTCGAGCCCGTCCGGGGTCAC	Arg
B. reuteri DSM 23975	Breu1	CAAGCCCCTGCGGCCCCACTCATT	Met
	Breu2	CTTGTTTCCGCCGAAGCGGATTGGTGGAGGCGCGGGGAATTGAACCCCGGTC	-
B. saguini DSM 23967	Bsag3	CAAGCCCCTGCGGCCCCACTCATT	Met
B. scardovii LMG 21589	Bsca1	CCGCTATACGGAAAACC	-
	Bsca2	GCGCTATAGTATGAAGCTAT	-
	Bsca3	TCAGTCTTCCAAACTGATTACGCGGGTTCGATTCCCGTC	Gly
	Bsca4	GTGCGAGTGGGGGGAGTTGAACCC	Leu
B. stellenboschense DSM 23968	Bstell1	TTCAATTCCCCGCGACTCCACCACACGAAAGCGCCGCTCCGAAAGGGGCGGCGCTTTTTCGTTG	-
B. stercoris DSM 24849	Bster2	GGGTTCGAGTCCCGCTGGAGGCACTTTTGGAAACCGCCAGA	Leu
B. thermacidophilum subsp. *thermacidophilum* LMG 21395	Bthermacid1	AACGTTGCAATCGCGCCGTTTTTGAATGCACCGGAACAGGCCGTTCC	-