SUPPLEMENTARY TEXT

Comparative genomics. A comparative study was undertaken to determine putative orthology between the *B. adolescentis* 22L protein-encoding sequences with those of other completely sequenced genomes belonging to the *Bifidobacterium* genus (Fig. S1). These results revealed 626 putative orthologs that were shared between all these genomes (Fig. S1). Using an *in silico* approach to predict average nucleotide identity (ANI) values between microbial genomes (1), we showed that, as expected, the *B. adolescentis* 22L genome is very similar (ANI value 97.92) to the *B. adolescentis* ATCC15703, and that the genomic structure of 22L is highly syntenic with that of *B. adolescentis* ATCC15703. While comparing the *B. adolescentis* 22L with other members of the *B. adolescentis* phylogenetic group and *B. dentium* Bd1 the obtained ANI value of 80.73 is considerably below the cut-off value for species circumscription (ANI value 94), confirming that 22L belongs to *B. adolescentis* species. A varying range of sequence identity was detected between the ORFs shared by 22L and ATCC15703 genomes; with the large majority displaying an identity of 90-100 %, and just 119 ORFs showing a similarity of less than 90 % (Fig. S1).

However, 98 ORFs were shown to be uniquely present in the genome of *B. adolescentis* 22L. Among these unique 22L DNA regions (Fig. S2), the largest includes genes predicted to specify two capsular polysaccharide biosynthesis loci [BADO_0389-BADO_0407 (Region 1) and BADO_1638-BADO_1647 (Region 11)], four restriction/modification (R/M) systems [(BADO_0658-BADO_0661 (Region 2), BADO_0961-BADO_0963 (Region 3), BADO_1029-BADO_1031 (Region 4) and BADO_1504-BADO_1506 (Region 9)], a gene cluster encoding pilus type IVa [BADO_1197-BADO_1207 (Region 5)], two prophage DNA regions [BADO_1216-BADO_1275 (Region 6) and BADO_1575-BADO_1597 (Region 10)], a Regulator of Chromosome Condensation 1 (RCC1) locus [BADO_1286-BADO_1300 (Region 7)], which has previously been described to also occur in the genome of the bifidobacterial insect-derived strain *B. asteroides* PRL2011 (2) and a

Clustered of Regularly Interspersed Short Palindromic Repeats (CRISPR) locus [BADO_1326-BADO_1332 (Region 8)] (3).

The mobilome of *B. adolescentis* 22L genome. Analysis of G+C content, BLASTP best-match, amino acid usage and codon preference of the *B. adolescentis* 22L genome revealed various chromosomal regions that are presumed to have been acquired by Horizontal Gene Transfer (HGT) (Fig. S2). Notably, a large part of these includes *B. adolescentis* 22L unique genes such as two loci encoding prophage (BADO_1216-BADO_1275 and BADO_1575-BADO_1597), three loci encoding exo-/capsular-polysaccharides (BADO_0389-BADO_0407, BADO_1363-BADO_1372 and BADO_1638-BADO_1647), and four gene clusters each specifying a putative R/M system (BADO_0658-BADO_0661, BADO_0961-BADO_0963, BADO_1029-BADO_1031 and BADO_1501-BADO_1504). Furthermore, the genome of *B. adolescentis* 22L harbors six insertion sequences (IS) (Table 1) (Fig. S2). Putative mobile elements identified in the *B. adolescentis* 22L genome are also represented by one CRISPR locus (BADO_1325-BADO_1333) (3), which were demonstrated to act as an RNA interference defense system against the invasion of foreign genetic material, in particular phages in other bacterial genomes.

References

1. **Richter M, Rossello-Mora R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences of the United States of America **106:**19126-19131.
2. **Bottacini F, Milani C, Turroni F, Sanchez B, Foroni E, Duranti S, Serafini F, Viappiani A, Strati F, Ferrarini A, Delledonne M, Henrissat B, Coutinho P, Fitzgerald GF, Margolles A, van Sinderen D, Ventura M.** 2012. Bifidobacterium asteroides PRL2011 genome analysis reveals clues for colonization of the insect gut. PloS one **7:**e44229.
3. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.
4. **Coutinho PM, Henrissat B.** 1999. Life with no sugars? Journal of molecular microbiology and biotechnology **1:**307-308.

Table S1. Primer information related to genes targeted in RT-qPCR experiments.

Gene Target	Sequence
BADO_0089_fw	3' - GAATCCGGTTCCAACGAATC - 5'
BADO_0089_rev	3' - CGGCAGACTACTGCCATACA - 5'
BADO_0090_fw	3' - GCAGCAGAATGGTATCAGCA - 5'
BADO_0090_rev	3' - TCAGCAGAGGCTTCACAATG - 5'
BADO_0091_fw	3' - CGCTGCCAGCTGGTATCTAT - 5'
BADO_0091_rev	3' - GCGGAGACAGTCTTGCTAGG - 5'
BADO_0093_fw	3' - GGCGATGGTAACAAGGATTC - 5'
BADO_0093_rev	3' - CGGCAGACTACTGCCATACA - 5'
BADO_0094_fw	3' - GCGCATCAAGAAGAACGATG - 5'
BADO_0094_rev	3' - CTGGTCCGATCCGTCTTACTC - 5'
BADO_0095_fw	3' - GATAACGCAGGAAGGCAATG - 5'
BADO_0095_rev	3' - CATCCTTGGTGGTGATTGTG - 5'
BADO_0445_fw	3' - GTGGTCTTGTGACAGCCTTG - 5'
BADO_0445_rev	3' - CATACGGTGTGCAGGTCATC - 5'
BADO_0446_fw	3' - CACCTATATGGCGCAATCCT - 5'
BADO_0446_rev	3' - GCCATTCGCATACCTATCGT - 5'
BADO_0447_fw	3' - CGACGACTACCAGACCAGTG - 5'
BADO_0447_rev	3' - CGGTCTTCGTGAAGGAGAAG - 5'
BADO_1455_fw	3' - CAGAACTATGCGGCTGTCAA - 5'
BADO_1455_rev	3' - CTGTACCACGTGACGGTGTC - 5'
BADO_1456_fw	3' - CGATCTCACTGCATCTGCTC - 5'
BADO_1456_rev	3' - CGGCTTATCGACCTTACGAG - 5'
BADO_1458_fw	3' - GATGAATCGGACGCCAAG - 5'
BADO_1458_rev	3' - CAGTCGCTGAATGTCGGATA - 5'
BADO_1545_fw	3' - CCTCGGCTGATTCCTATGAT - 5'
BADO_1545_rev	3' - CAATGTCAATCGCCTCGTAA - 5'
BADO_1546_fw	3' - GCACGCTCATCTACAGCATC - 5'
BADO_1546_rev	3' - GCTGTTGGTGGCCTTGTAGT - 5'
BADO_1547_fw	3' - GCAGGCTACGACACCTATCC - 5'
BADO_1547_rev	3' - CCAGATCGAACTTCAGCACA - 5'
*pil*A_Fw	3' - CGGACGTTAAGAATGCTTCC - 5'
*pil*A_Rev	3' - TGGTAAGAGTCACGCCATCA - 5'
*pil*B_Fw	3' - GGCGAGAAGATTGTTATGCG - 5'
*pil*B_Rev	3' - GTAGGACCGGTGACCAACAC - 5'
*pil*C_Fw	3' - TCGGGCAACTATGTGATTGA - 5'
*pil*C_Rev	3' - GATTGGGCCTCTTCGTCATA -5'
*pil*D_Fw	3' - GACTGTCCTTCCTTGCCGTA - 5'
*pil*D_Rev	3' - CGGCAGTAGCACTACGTTGA - 5'
*pil*M_Fw	3' - GTGCCGTTGTCTGTAGTGGA - 5'
*pil*M_Rev	3' - ATGGAGGTGATCACGAAACC - 5'
*pil*V_Fw	3' - GAGGTATTGGTGGCGATTGT - 5'
*pil*V_Rev	3' - GGATGGCGCAGAAGTATGAT - 5'
*pil*T_FW	3' - CGGATATCTTTCCGGTCTCA - 5'
*pil*T_Rev	3' - GGCATTGAATTAAGCGAAGC - 5'

Table S2. Genome features of *B. adolescentis* 22L

Genetic features	Value
Size	2,203,222 bp
G+C content	59.29%
Number of identified	1725
Unknown function	411
Assigned function	1314
COG categories:	
RNA processing and modification	2
Energy production and conversion	39
Cell cycle control, cell division, chromosome partitioning	20
Amino acid transport and metabolism	160
Nucleotide transport and metabolism	53
Carbohydrate transport and metabolism	164
Coenzyme transport and metabolism	38
Lipid transport and metabolism	36
Translation, ribosomal structure and biogenesis	129
Transcription	91
Replication, recombination and repair	92
Cell wall/membrane/envelope biogenesis	70
Cell motility	6
Posttranslational modification, protein turnover, chaperones	45
Inorganic ion transport and metabolism	56
Secondary metabolites biosynthesis, transport and catabolism	2
General function prediction only	124
Function unknown	95
Signal transduction mechanisms	38
Intracellular trafficking, secretion, and vesicular transport	9
Defense mechanisms	45
Phage regions	2
IS transposase families:	
- ISL3	2
- IS256	3
- IS1595	1
CRISPR	1
Fimbrial systems	5
Transporters (genes)	
- ABC systems	149
- PTS systems	2
- MFS systems	21

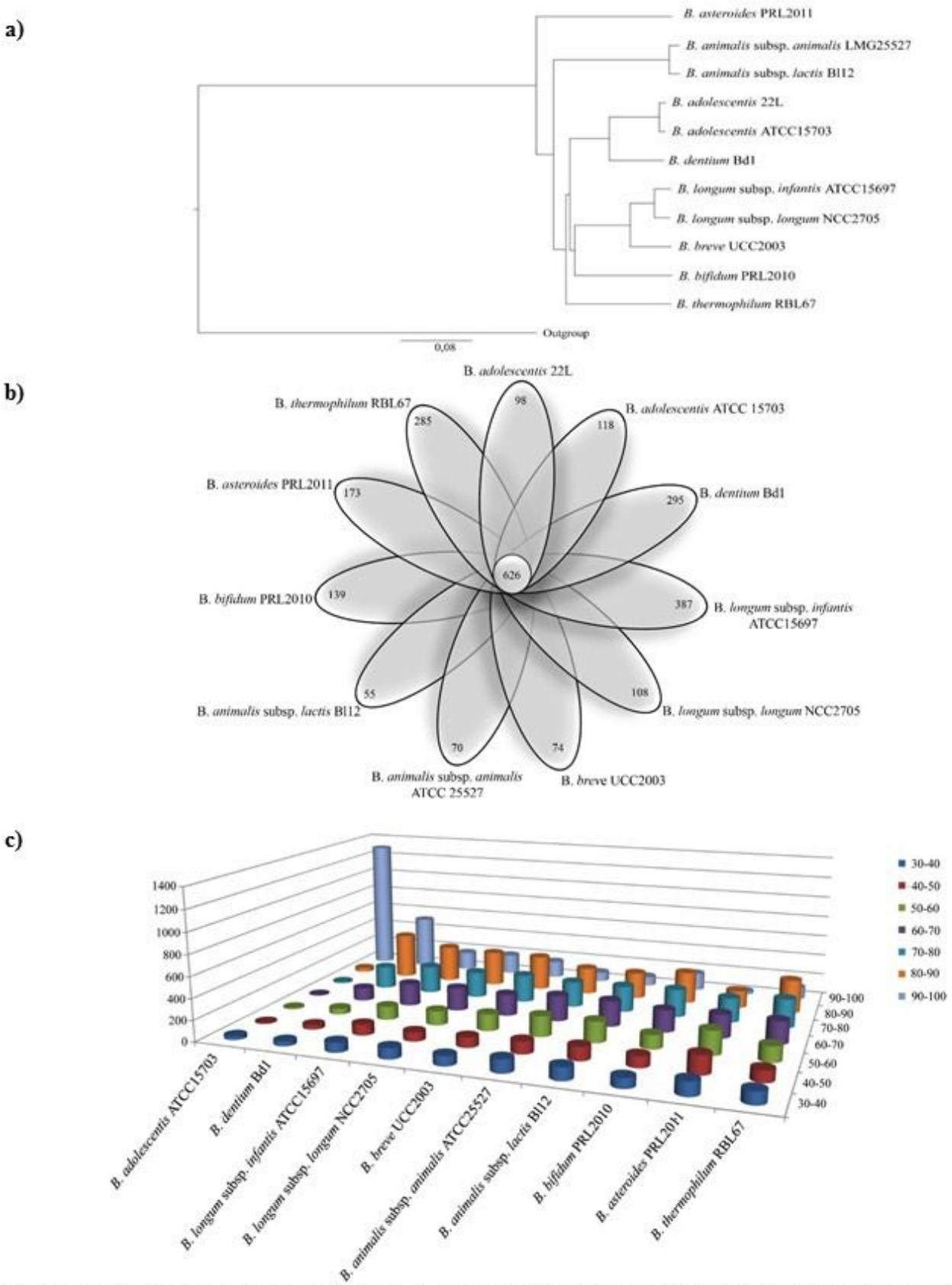

Figure S1. Comparative genomic analyses of *B. adolescentis* 22L with different bifidobacterial genomes. Panel a depicts phylogenetic supertree based on the sequences of *Bifidobacterium* core proteins, using SplitsTree. Panel b represents a Venn diagram of homologs shared between sequenced bifidobacterial genomes. Circle sizes are proportional to members contained in each set. Panel c shows the percentage of amino acid identity of the top-scoring self-matches for protein-coding genes in the analysed bacteria using the predicted proteome of *B. adolescentis* 22L as a reference. For each bacterium, the deduced protein-coding regions for each gene were compared with those derived from the *B. adolescentis* 22L genome. The y axis represents the number of genes detected, whereas the z axis the level of identity.

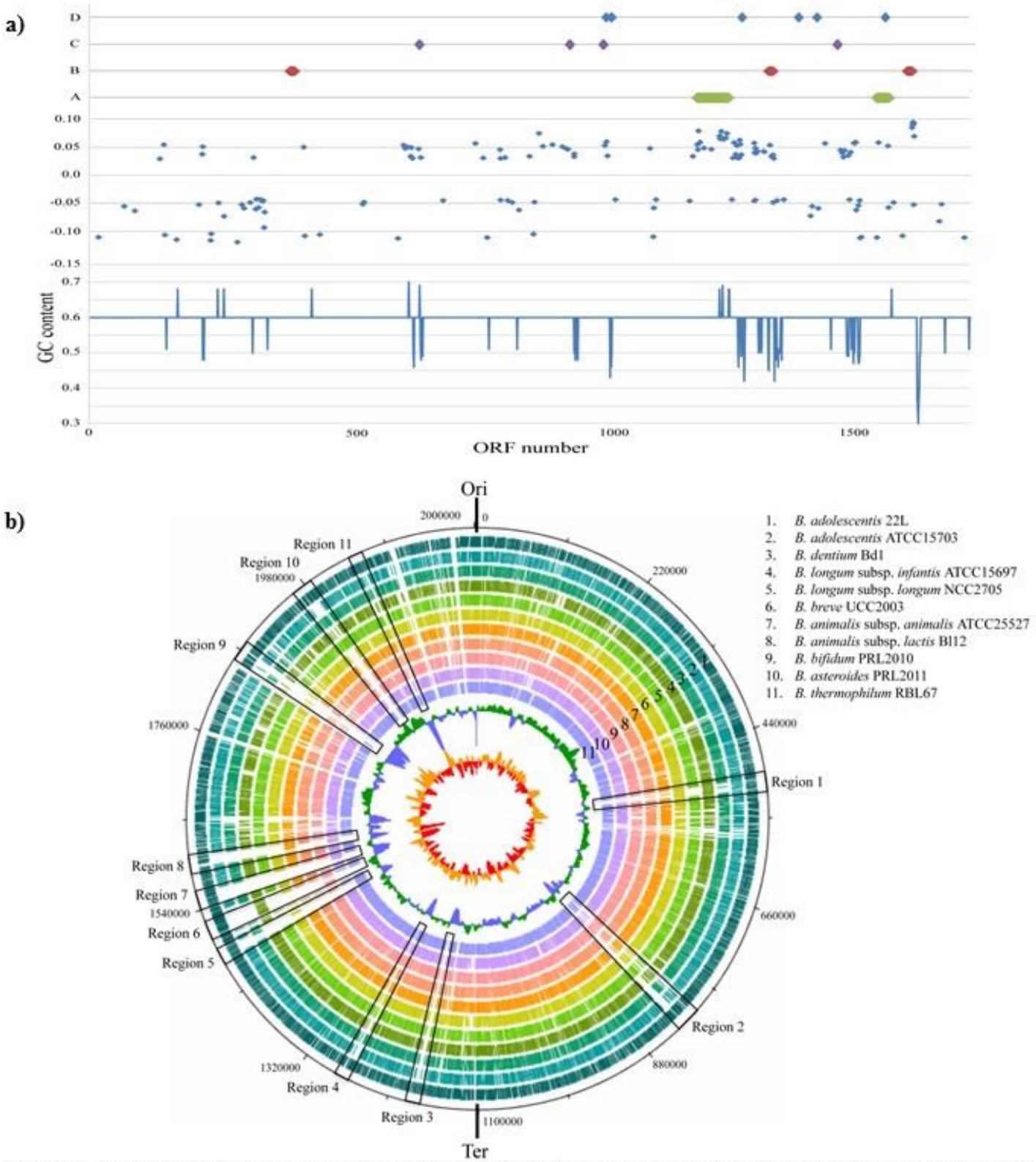

Figure S2. Variable *B. adolescentis* 22L genomic regions identified with respect to the other bifidobacterial genomes. Panel a shows mobile genetic elements of the *B. adolescentis* 22L genome. The first plot from the bottom indicates the deviation of the G+C content of each ORF of the *B. adolescentis* 22L genome from the mean average (59.29%). In the second plot each dot represents an ORF displaying a biased codon usage determined by factorial correspondence analysis of codon usage. The other plots are identified by a one-letter abbreviation: A, prophage; B, capsular polysaccharide biosynthesis; C, R/M system and D, IS element. Panel b represents a circular genome atlas of *B. adolescentis* 22L with mapped orthologs (defined as reciprocal best FastA hits with more than 30% identity over at least 80% of both protein lengths) (circle 1) respect to seven other publicly available *Bifidobacterium* genomes (circles 2-11). The regions identified in 22L genome compared to the other so far available complete bifidobacteria genomes are mapped. Regions one and 11 indicates two capsular polysaccharide biosynthesis locus, Region two, three, four and nine encode for R/M system, Region five represents cluster genes encoding for a pilus type IVa, Region six and 10 include two prophage-like elements, Region seven encompasses a RCC1 locus and Region eight including CRISPR locus.

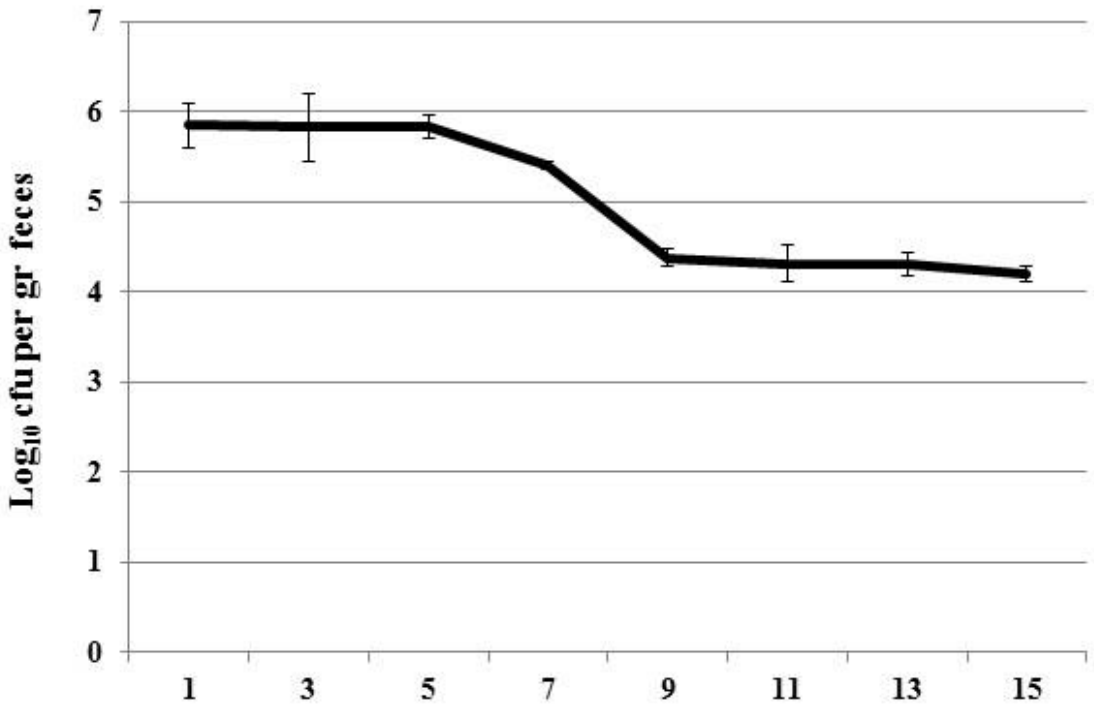

Figure S3. Population sizes of *B. adolescentis* 22L colonizing the intestine of BALB/c mice. Each point represents the average of the log-population size SD for five mice

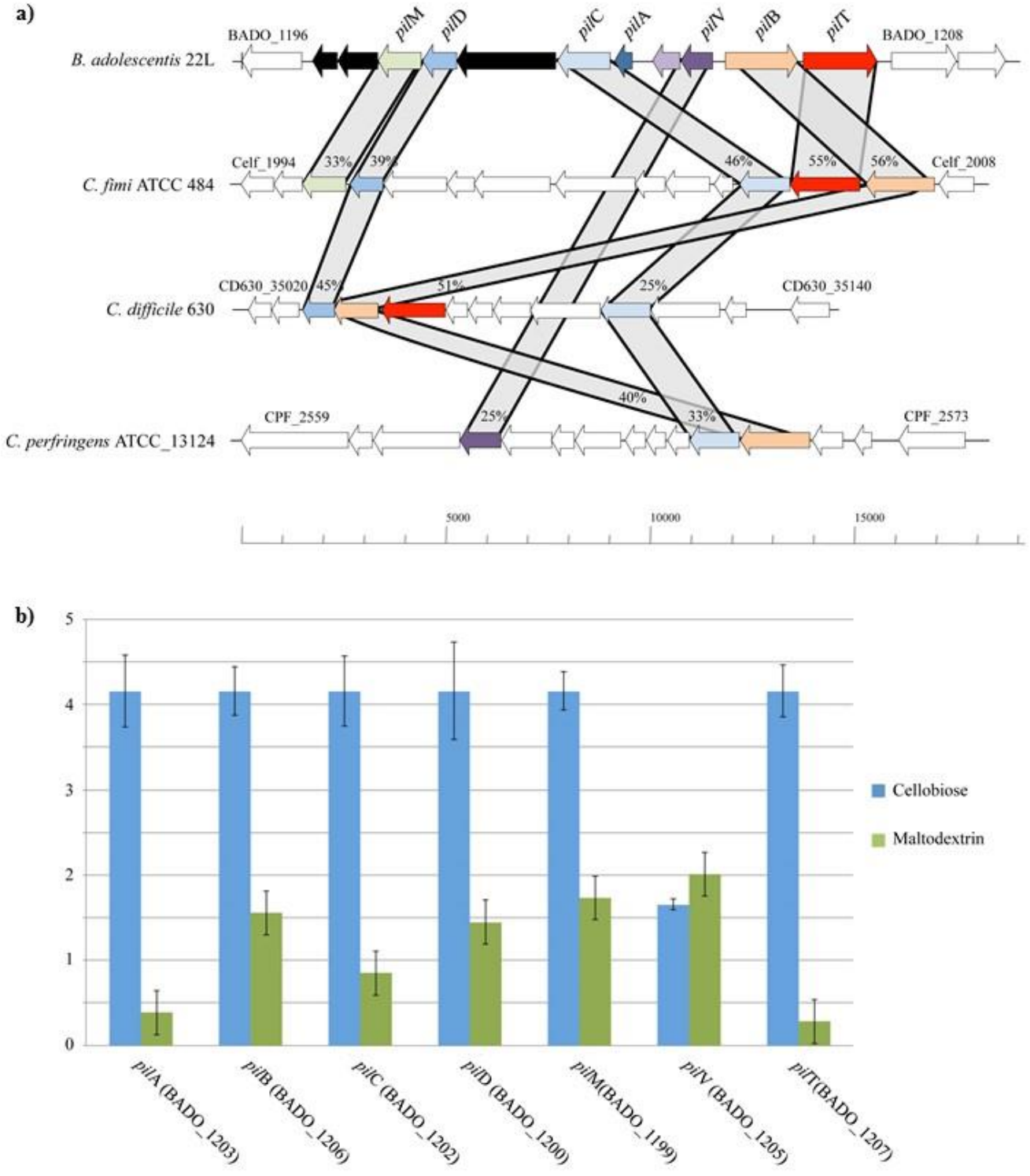

Figure S4. Pilus type IVa locus identified in the chromosome of *B. adolescentis* 22L. Panel a shows a schematic comparative genetic map of the pilus type IVa of *B. adolescentis* 22L and other Firmicutes species. Each arrow indicates an open reading frame (ORF), the size of which is proportional to the length of the arrow. Coloring of the arrows represents the different function of the gene as indicated above each arrow. The amino acid identity of the relevant encoded proteins is indicated in percentages. Panel b represents the qRT-PCR relative transcription levels of pilus type IVa encoding genes from *B. adolescentis* 22L upon cultivation in a medium supplemented with maltodextrin or cellobiose, as unique carbon sources, versus growth in glucose supplemented with glucose as unique carbon source. The histograms indicate the relative amounts of the pilus subunit mRNAs for the specific samples. The data displayed are based on RNA preparations from two independent culture experiments.

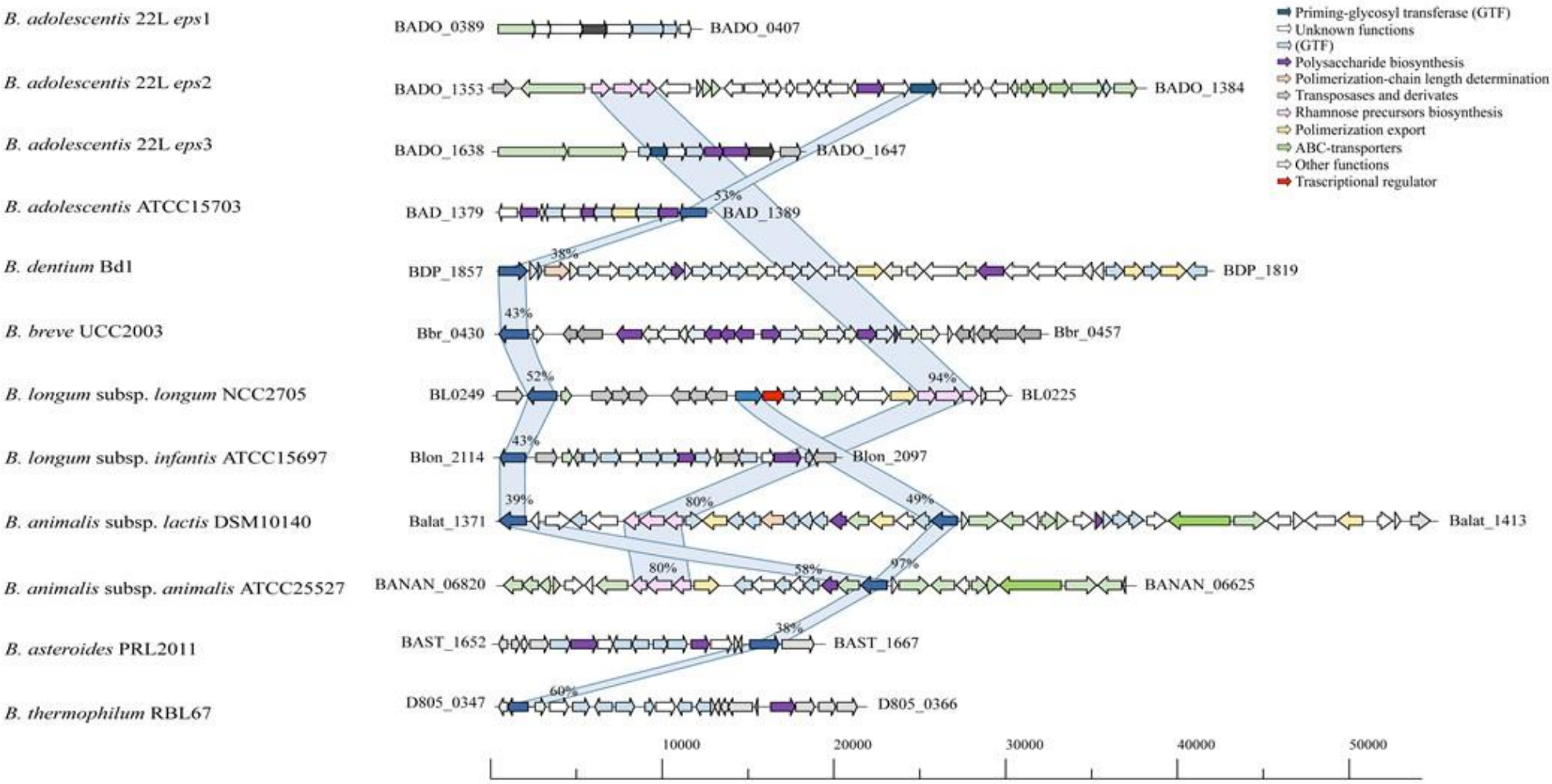

Figure S5. Comparison of the the putative *eps* clusters identified in the genome of *B. adolescentis* 22L with the corresponding loci from different bifidobacteria. The genes were categorized according to their potential functions, which are indicated with the colored arrows in the box. For each strain, the amino acid identity (percentage) of the priming GTF and the rhamnose biosynthesis with respect to those of *B. adolescentis* 22L is indicated.