



## RESEARCH

## Open Access

# The genome of *Bifidobacterium pseudocatenulatum* IPLA 36007, a human intestinal strain with isoflavone-activation activity

Ángel Alegría<sup>1†</sup>, Susana Delgado<sup>1†</sup>, Lucía Guadamuro<sup>1</sup>, Ana Belén Flórez<sup>1</sup>, Giovanna E Felis<sup>2</sup>, Sandra Torriani<sup>2</sup> and Baltasar Mayo<sup>1\*</sup>

## Abstract

**Background:** *Bifidobacterium* species, including *Bifidobacterium pseudocatenulatum*, are among the dominant microbial populations of the human gastrointestinal tract. They are also major components of many commercial probiotic products. Resident and transient bifidobacteria are thought to have several beneficial health effects. However, our knowledge of how these bacteria interact and communicate with host cells remains poor. This knowledge is essential for scientific support of their purported health benefits and their rational inclusion in functional foods.

**Results:** This work describes the draft genome sequence of *Bifidobacterium pseudocatenulatum* IPLA 36007, a strain isolated as dominant from the feces of a healthy human. Besides several properties of probiosis, IPLA 36007 exhibited the capability of releasing aglycones from soy isoflavone glycosides. The genome contains 1,851 predicted genes, including 54 genes for tRNAs and five copies of unique 16S, 23S and 5S rRNA genes. As key attributes of the IPLA 36007 genome we can mention the presence of a lysogenic phage, a cluster encoding type IV fimbriae, and a locus encoding a clustered, regularly interspaced, short, palindromic repeat (CRISPR)-Cas system. Four open reading frames (*orfs*) encoding  $\beta$ -glucosidases belonging to the glycosyl hydrolase family 3, which may act on isoflavone glycosides, were encountered. Additionally, one gene was found to code for a glycosyl hydrolase of family 1 that might also have  $\beta$ -glucosidase activity.

**Conclusion:** The availability of the *B. pseudocatenulatum* IPLA 36007 genome should allow the enzyme system involved in the release of soy isoflavone aglycones from isoflavone glycosides, and the molecular mechanisms underlying the strain's probiotic properties, to be more easily understood.

## Introduction

*Bifidobacterium* species are majority bacteria among those inhabiting the gastrointestinal tract (GIT) of animals and humans. They play important roles in maintaining human health via the digestion of foods, production of essential vitamins, and metabolization of endogenous and exogenous compounds, as well as by preventing the colonization and/or overgrowth of pathogens in the GIT [1]. Molecular analyses have shown that members of the *Bifidobacterium*

*catenulatum* group (which includes *B. catenulatum* and *B. pseudocatenulatum*) are abundant in fecal samples from adult humans [2,3]. *B. pseudocatenulatum* strains have a number of probiotic properties, such as the possession of antinutrient-degrading enzymes [4], the ability to bind mutagenic aromatic amines [5], and the capacity to reduce cholesterol levels [6]. However, compared to other bifidobacterial species, the genome of *B. pseudocatenulatum* has been very little explored. The Genomes Online Database (GOLD) (<http://www.genomesonline.org>) only contains the draft sequence of a single strain, *B. pseudocatenulatum* DSM 20438 (Gi02660), plus recently released incomplete sequences of five other strains (D2CA; TSDC19.1-1.1; TSDC19.1-1.2; TSDC19.1-1.3; and TSDC17.2-1.1). Sequence analysis of additional *B. pseudocatenulatum* strains

\* Correspondence: [baltasar.mayo@ipla.csic.es](mailto:baltasar.mayo@ipla.csic.es)

†Equal contributors

<sup>1</sup>Departamento de Microbiología y Bioquímica de Productos Lácteos, Instituto de Productos Lácteos de Asturias (IPLA-CSIC), Carretera de Infesto, s/n, 33300 Villaviciosa, Asturias, Spain

Full list of author information is available at the end of the article

would provide greater insight into the intra-specific variation of this species, and supply information on the genetics that underlay strain-specific capabilities. Recently, *B. pseudocatenulatum* has been used as a cloning host for the expression of natural [7] and synthetic [8] genes. Genomic analyses of strains of this species might allow the confident use of this bacterium in other biotechnological applications.

Bifidobacterial strains have been shown to be involved in the conversion of isoflavone glycosides into aglycones [9,10], a key step in making isoflavones bioavailable and harnessing their estrogenic activity [11]. Indeed, the genomes of sequenced bifidobacteria show an impressive array of genes coding for glycosyl hydrolases, including  $\beta$ -glucosidases, which are thought to be involved in the release of aglycones from dietary polyphenols such as soy isoflavones [12,13]. However, the enzyme(s) involved in the hydrolysis of soy isoflavone glycosides remain(s) mostly unknown. So far, a  $\beta$ -glucosidase from *Bifidobacterium animalis* subsp. *lactis* has been shown to possess aglycone-releasing activity from isoflavones by cloning and expression of its encoding gene in *Bifidobacterium bifidum* [14].

The present work provides a draft genome sequence for *B. pseudocatenulatum* IPLA 36007, an intestinal human strain able to release aglycones from the soy isoflavone glycosides daidzin and genistin. This capability endows it with properties of interest in terms of its use in functional foods.

## Materials and methods

### Isolation and DNA preparation

*B. pseudocatenulatum* IPLA 36007 was isolated among the dominant bacteria from fecal samples of a healthy human, in a study approved by The Ethic Committee of the Asturias Principality, Spain [15]. The strain was grown anaerobically at 37°C in MRS medium (Merck, Darmstadt, Germany) supplemented with 0.25% cysteine (Merck). Genomic DNA was extracted and purified from pure cultures using the GenElute™ Bacterial Genomic DNA kit (Sigma-Aldrich, St. Louis, Miss., USA) following the manufacturer's instructions for extracting DNA from Gram-positive bacteria. The concentration and quality of the DNA was measured using an Epoch microvolume spectrophotometer (BioTek Instruments, Winooski, Vt., USA).

### Aglycone releasing-activity from isoflavone glycosides

Strains were incubated anaerobically in a MRS basal medium without dextrose and supplemented with 2% cellobiose and 100  $\mu$ M daidzin or genistin (Sigma-Aldrich) at 37°C for 24 h. One ml cultures were centrifuged and the cells suspended in the same volume of 0.1 M sodium acetate buffer pH 4.1. Isoflavones and derivatives were then extracted with ethyl acetate (Sigma-Aldrich). The organic phase was evaporated and the dried pellet

suspended in 100  $\mu$ l of methanol. Five 5  $\mu$ l were used for analysis by TLC in silica gel 60 F254 plates (Merck). Isoflavones were separated in a toluene:acetone (2:1) solvent system, revealed by UV light at 365 nm in a transilluminator and visualized with an ImageQuant 350 (GE Healthcare Bio-Sciences, Buckinghamshire, UK).

### Genome sequencing, assembly and annotation

A genomic library of 0.5 kbp was constructed and paired-end sequenced (approximately 155-fold coverage) using a HiSeq 1000 System sequencer (Illumina, Inc., San Diego, CA, USA). Quality-filtered reads were assembled in contigs using Velvet software v.1.2.10. (<https://www.ebi.ac.uk/~zerbino/velvet/>). Gaps within the contigs were closed by direct sequencing of amplicons obtained by PCR with oligonucleotide primers designed to anneal in the flanking regions. The genome was annotated with the RAST annotation system (<http://rast.nmpdr.org/>) and the NCBI Prokaryotic Genome Annotation Pipeline ([http://www.ncbi.nlm.nih.gov/genome/annotation\\_prok/](http://www.ncbi.nlm.nih.gov/genome/annotation_prok/)). The KEGG Pathway (<http://www.genome.jp/kegg/pathway.html>), Uniprot (<http://www.uniprot.org>) and COG (<http://www.ncbi.nlm.nih.gov/COG>) databases were consulted for description of specific genes and proteins. If required, DNA and deduced protein sequences were individually subjected to BLAST analysis (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Multi-blast protein comparisons were performed with the CLC Bioinformatics Database software package (CLC bio, Aarhus, Denmark).

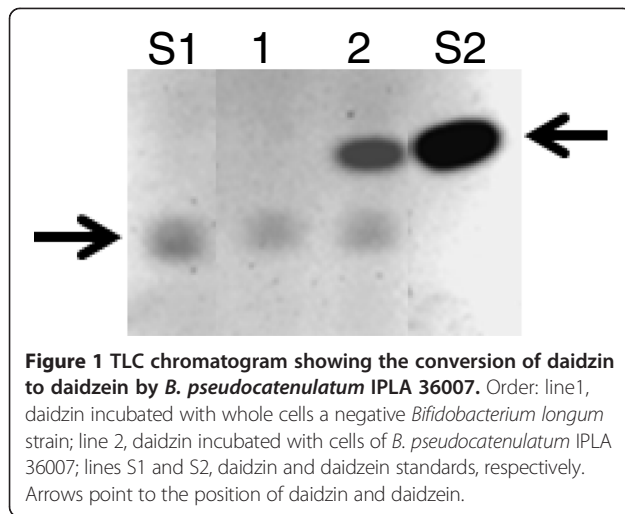
### Nucleotide sequence accession numbers

The results of this Whole Genome Shotgun project have been deposited in the GenBank database under accession number JEOD00000000. The version described in this paper is JEOD01000000.

## Results and discussion

Among a large collection of intestinal bifidobacteria strains from human origin, *B. pseudocatenulatum* IPLA 36007 showed aglycone-releasing activity from isoflavone glycosides. Figure 1 shows the conversion of daidzin into daidzein after incubation with IPLA 36007 cells. Similar activity was detected using genistin as a substrate, which was transformed into genistein (data not shown). As this strain has already shown a bunch of key properties for its use as a probiotic, including among others good survival under conditions simulating those of the GIT, absence of undesirable enzyme activities and atypical antibiotic resistances, and ability to bind human intestinal epithelial cells [16], it was selected for whole genome sequencing in order to get insights on the molecular basis of its relevant phenotypic traits.

The general features of the *B. pseudocatenulatum* IPLA 36007 genome are summarized in Table 1. The draft



genome sequence of IPLA 36007 included 23 contigs varying in size from 203 to 548,016 bp. According to the NCBI Prokaryotic Genome Annotation Pipeline the genome harbors 1,851 genes, of which 1,769 are coding sequences (CDS) and 22 are pseudogenes. Additionally, 60 predicted RNA genes were identified, including five copies of identical 16S, 23S and 5S rRNA genes, a non-coding RNA molecule (ncRNA), and 54 genes coding for tRNAs. The RAST server classified the CDS into 26 classes and 255 subsystems (sets of related functional roles).

This strain was demonstrated plasmid-free (data not shown), and, in agreement, no plasmid-associated genes were found. However, one integrated phage of around 43.5 kbp was recorded. The lysogenic phage region consisted of 76 CDS, and included a gene encoding a retron-type RNA-directed DNA polymerase typical of group II introns [17]. Phage-related sequences have been described in the genome of 22 strains of different bifidobacteria species, but only fragmentary information exists

with regard to their functionality [18]. As in other bacteria, bifidobacterial prophages have been shown to possess classical modular genomic organization in which the DNA lysogeny module and the DNA packaging region are the most highly conserved.

A cluster of nine genes capable of encoding type IV fimbriae or pili was identified. This contained, among others, orthologous genes for *pilA*, *pilB*, *pilC*, *pilM*, and *pilT* (Figure 2). Operons encoding type IV fimbriae have also been found in the genome of other sequenced *B. pseudocatenulatum* strains, in other bifidobacterial species, and in some environmental actinobacteria. Such pili have been well studied in Gram-negative bacteria such as *Neisseria gonorrhoeae*, *Pseudomonas aeruginosa*, *Pseudomonas stutzeri*, *Moraxella bovis* and *Dichelobacter nodosus* [19]. In these pathogenic bacteria, they mediate attachment and adherence to epithelial cells, twitching motility, gliding motility, cell agglutination, and biofilm and fruiting body formation [19]. In addition, they act as receptors for bacteriophages, and are required for extracellular protein secretion and natural transformation [19]. The expression and functionality of the type IV fimbriae genes in *B. pseudocatenulatum* IPLA 36007 has yet to be demonstrated.

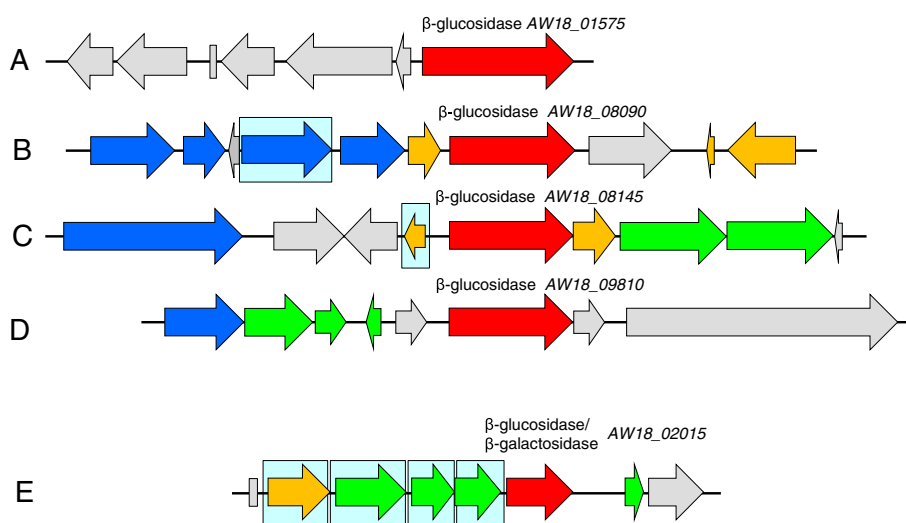
A locus encoding a CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins) system was identified (Figure 3). This locus contained 22 identical repeats of 33 bp (except for one repeat carrying a single C-T transition at position five) and 21 spacers ranging from 32 to 37 bp (Figure 3B). The repeats (5'-GTCGCTCCTCATGGAGAGCGTGGATTGAAA T-3') were preceded by eight CRISPR-associated (*cas*) genes (Figure 3A). Spacers showed no significant homology to DNA sequences on databases. Gene content and gene order analysis showed the CRISPR-Cas system of IPLA 36007 to belong to type I-C [20]. CRISPR-Cas

**Table 1** Key features of the *B. pseudocatenulatum* IPLA 36007 genome

Characteristic	Figures	Observations
Size of the genome	2,328,179 bp <sup>a</sup>	
G + C content	56.4%	
Contigs	23	(from 203 to 548,016 bp long)
Open reading frames (ORFs)	1851	
Coding sequences (CDS)	1769	
Pseudogenes	22	
RNA genes	60	5 rRNA operons, 54 tRNA, 1 ncRNA
Plasmidic genes	0	Plasmid free strain
Integrated phages	1	76 CDS
CRISPR-Cas system	1	22 repeats, 21 spacers, 8 CDS
Glycosyl-hydrolases and glycosyl transferases	>50	
β-glucosidases	5	5 glycosyl hydrolases family_3, 1 glycosyl hydrolase family_1

<sup>a</sup>bp, base pair; ncRNA, non-coding RNA.





**Figure 4 Genetic organization around the putative  $\beta$ -glucosidase-encoding genes.** Five genes encoding glycosyl hydrolases of family 3 (A through D) and a gene encoding a glycosyl hydrolase of family 1 (E) were identified in the *B. pseudocatenulatum* IPLA 36007 genome. Color code: glycosidase genes are in red; in green, genes involved in transport; in blue, genes involved in carbohydrate(s) metabolism; in brown, genes encoding regulator proteins; in gray, genes belonging to a distinct RAST subsystem; boxes represent genes encoding tRNA and 5S rRNA molecules. Pale blue background boxes indicate functionally coupled genes sharing conserved relative positions in the genome of at least four other species.

IPLA 36007 contains a vast array of glycosyl transferases and glycosyl hydrolases CDS (52 genes), including genes that code for xylanases (4), pullulanases (3), amyloamylases (2)  $\alpha$ -amylase (1) and maltodextrin glucosidase (1). It also harbors genes encoding  $\alpha$ - (5) and  $\beta$ - (9) galactosidases,  $\alpha$ - (1) and  $\beta$ - (7) xylosidases,  $\alpha$ - (1) and  $\beta$ - (4) glucosidases,  $\alpha$ -arabinofuranosidases (4),  $\beta$ -mannosidases (2), and  $\alpha$ -rhamnosidases (1). Although belonging to the glycosyl hydrolase family 3, all four  $\beta$ -glucosidases (EC 3.2.1.21) shared very limited amino acid identity (26.2%) (Figure 4, A through D). Multiblast analysis of the  $\beta$ -glucosidases found in the IPLA 36007 genome and those present in bifidobacterial genomes suggested that homologous enzymes are present in bifidobacteria strains belonging to the *Bifidobacterium adolescentis* group (*B. adolescentis*, *B. dentium*, *B. angulatum*, *B. catenulatum*, etc.) (amino acid identity ranging from 80 to 98%). The homology to glycosyl hydrolases from other groups of bifidobacteria was much lower (less than 70% amino acid identity) (Additional file 1: Figure S1). Indeed, the similarity of the deduced  $\beta$ -glucosidases from IPLA 36007 to that of *B. animalis* subsp. *lactis* SH5, which have been proved to act on soy isoflavone glycosides [14], was found to be marginal (23% amino acid identity). An additional gene encoding a putative glycoside hydrolase belonging to family 1 (EC 3.2.1.23) was detected; this might also have  $\beta$ -glucosidase activity (Figure 4, E). Characterization of these genes and their encoded glycosyl hydrolase enzymes would allow those acting on soy isoflavone glycosides to be identified. However, hydrolysis of

isoflavone glycosides some others glycosyl hydrolases of those detected cannot be discarded.

The availability of the genome of the *B. pseudocatenulatum* IPLA 36007 strain should allow the enzymes involved in the release of soy isoflavone aglycones from isoflavone glycosides to be known. This is essential for the rational use of IPLA 36007 as a probiotic in functional foods. The ability of IPLA 36007 to colonize the GIT could be exploited to deliver the aglycone-releasing activity straightway into the human intestine. Comparison of sequences from different sequenced strains would provide greater insights into the genetic variation within this species. It will further allow the core genome and pangenome of *B. pseudocatenulatum* to be identified, while contributing towards defining the gene set required to be competitive in the human GIT.

## Additional file

**Additional file 1: Figure S1.** Multiblast analysis of putative  $\beta$ -glucosidases from *Bifidobacterium pseudocatenulatum* IPLA 36007 to those on the genomes of other bifidobacteria in the databases.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

BM and SD designed the study. AA, ABF and LG performed the experiments. AA, SD, BM, GEF and ST analyzed the data. BM and SD wrote the manuscript. AA, GEF and ST checked and edited the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

This study was supported by a project from the Spanish Ministry of Economy and Competitiveness (Ref. AGL2011-24300). We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). A.B. Flórez and A. Alegría were supported by a contract of the JAE Program from CSIC and a short-term fellowship from FEMS, respectively.

### Author details

<sup>1</sup>Departamento de Microbiología y Bioquímica de Productos Lácteos, Instituto de Productos Lácteos de Asturias (IPLA-CSIC), Carretera de Infesto, s/n, 33300 Villaviciosa, Asturias, Spain. <sup>2</sup>Dipartimento di Biotecnologie, Università degli Studi di Verona, Strada Le Grazie, 15, 37134 Verona, Italy.

Received: 23 June 2014 Accepted: 16 July 2014

Published: 23 July 2014

### References

1. Leahy SC, Higgins DG, Fitzgerald GF, van Sinderen D: Getting better with bifidobacteria. *J Appl Microbiol* 2005, **98**:1303–1315.
2. Matsuki T, Watanabe K, Fujimoto J, Kado Y, Takada T, Matsumoto K, Tanaka R: Quantitative PCR with 16S rRNA-gene-targeted species-specific primers for analysis of human intestinal bifidobacteria. *Appl Environ Microbiol* 2004, **70**:167–173.
3. Junick J, Blaut M: Quantification of human fecal *Bifidobacterium* species by use of quantitative Real-Time PCR analysis targeting the *groEL* gene. *Appl Environ Microbiol* 2012, **78**:2613–2622.
4. Haros M, Carlsson NG, Almgren A, Larsson-Alminger M, Sandberg AS, Andlid T: Phytate degradation by human gut isolated *Bifidobacterium pseudocatenulatum* ATCC27919 and its probiotic potential. *Int J Food Microbiol* 2009, **135**:7–14.
5. Faridnia F, Hussin AS, Saari N, Mustafa S, Yee LY, Manap MY: In vitro binding of mutagenic heterocyclic aromatic amines by *Bifidobacterium pseudocatenulatum* G4. *Benefic Microbes* 2010, **1**:149–154.
6. Al-Sheraji SH, Ismail A, Manap MY, Mustafa S, Yusof RM, Hassan FA: Hypocholesterolaemic effect of yoghurt containing *Bifidobacterium pseudocatenulatum* G4 or *Bifidobacterium longum* BB536. *Food Chem* 2012, **135**:356–361.
7. Alvarez-Martín P, Flórez AB, Margolles A, del Solar G, Mayo B: Improved cloning vectors for bifidobacteria, based on the *Bifidobacterium catenulatum* pBC1 replicon. *Appl Environ Microbiol* 2008, **74**:4656–4665.
8. Losurdo L, Quintieri L, Caputo L, Gallerani R, Mayo B, De Leo F: Cloning and expression of synthetic genes encoding angiotensin-I converting enzyme (ACE)-inhibitory bioactive peptides in *Bifidobacterium pseudocatenulatum*. *FEMS Microbiol Lett* 2013, **340**:24–32.
9. Tsangalis D, Ashton JF, McGill AEJ, Shah NP: Enzymic transformation of isoflavone phytoestrogens in soy milk by  $\beta$ -glucosidase producing bifidobacteria. *J Food Sci* 2002, **67**:3104–3113.
10. Raimondi S, Roncaglia L, De Lucia M, Amaretti A, Leonardi A, Pagnoni UM, Rossi M: Bioconversion of soy isoflavones daidzin and daidzein by *Bifidobacterium* strains. *Appl Microbiol Biotechnol* 2008, **81**:943–950.
11. Crozier A, Jaganath IB, Clifford MN: Dietary phenolics: chemistry, bioavailability and effects on health. *Nat Prod Rep* 2009, **26**:1001–1043.
12. Schell MA, Karmirantzou M, Snel B, Vilanova D, Berger B, Pessi G, Zwahlen MC, Desiere F, Bork P, Delley M, Pridmore RD, Arigoni F: The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* 2002, **99**:14422–14427.
13. Dabek M, McCrae SI, Stevens VJ, Duncan SH, Louis P: Distribution of beta-glucosidase and beta-glucuronidase activity and of beta-glucuronidase gene *gus* in human colonic bacteria. *FEMS Microbiol Ecol* 2008, **66**:487–495.
14. Youn SY, Park MS, Ji GE: Identification of the  $\beta$ -glucosidase gene from *Bifidobacterium animalis* subsp. *lactis* and its expression in *B. bifidum* BGN4. *J Microbiol Biotechnol* 2012, **22**:1714–1723.
15. Delgado S, Suárez A, Mayo B: Bifidobacterial diversity determined by culturing and by 16S rDNA sequence analysis in feces and mucosa from ten healthy Spanish adults. *Dig Dis Sci* 2006, **51**:1878–1885.
16. Delgado S, Sullivan E, Fitzgerald G, Mayo B: In vitro evaluation of the probiotic properties of human intestinal *Bifidobacterium* species and selection of new probiotic candidates. *J Appl Microbiol* 2008, **104**:1119–1127.

17. Tourasse NJ1, Kolstø AB: Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution. *Nucleic Acids Res* 2008, **36**:4529–4548.
18. Ventura M, Turróni F, Foroni E, Duranti S, Giubellini V, Bottacini F, van Sinderen D: Analyses of bifidobacterial prophage-like sequences. *Antonie Van Leeuwenhoek* 2010, **98**:39–50.
19. Han X, Kennan RM, Parker D, Davies JK, Rood JI: Type IV fibrillar biogenesis is required for protease secretion and natural transformation in *Dichelobacter nodosus*. *J Bacteriol* 2007, **189**:5022–5033.
20. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV: Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011, **9**:467–477.
21. Sampson TR, Weiss DS: CRISPR-Cas systems: new players in gene regulation and bacterial physiology. *Front Cell Infect Microbiol* 2014, **4**:37.
22. Pokusaeva K, Fitzgerald GF, van Sinderen D: Carbohydrate metabolism in bifidobacteria. *Genes Nutr* 2011, **6**:285–306.

doi:10.1186/1757-4749-6-31

**Cite this article as:** Alegría et al.: The genome of *Bifidobacterium pseudocatenulatum* IPLA 36007, a human intestinal strain with isoflavone-activation activity. *Gut Pathogens* 2014 **6**:31.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

