Prediction of new gene products and characterization of hypothetical proteins of *Bifidobacterium breve* DS15-17 *In Silico*

Predição de novos produtos gênicos e caracterização de proteínas hipotéticas de *Bifidobacterium breve* DS15-17 *In Silico*

DOI:10.34119/bjhrv5n2-012

Recebimento dos originais: 15/02/2022
Aceitação para publicação: 03/03/2022

Bianca Damasceno
Biomédica
Instituição: Universidade do Estado do Pará (UEPA) - Campus Marabá
Endereço: Avenida Hiléia, s/nº, Agrópolis do Incra, Marabá, Pará, Brasil
E-mail:balvesdamasceno@gmail.com

Ronilson Santos
Graduando em Ciências Biológicas
Instituição: Universidade Federal Rural da Amazônia (UFRA) - Campus Tomé-Açu
Endereço: Rodovia PA 140, 2428, Tomé-Açu, Pará, Brasil
E-mail:ronilson.ssantos01@gmail.com

Mônica Oliveira
Mestre em Computação Aplicada
Instituição: Universidade Federal do Pará (UFPA) - Campus Tucuruí
Endereço: Rodovia BR 422, Vila Permanente, Tucuruí, Pará, Brasil
E-mail:monik.info@gmail.com

Larissa Dias
Doutora em Biotecnologia
Instituição: Instituto Federal do Amapá (IFAP) - Campus Porto Grande
Endereço: Rodovia BR 210, Km 103, Porto Grande, Amapá, Brasil
E-mail:bio.larissadias@gmail.com

Alyne Lima
Doutora em Genética e Biologia Molecular
Instituição: Instituto Federal do Amapá (IFAP) - Campus Porto Grande
Endereço: Rodovia BR 210, Km 103, Porto Grande, Amapá, Brasil
E-mail:alynecslima@gmail.com

Adonney Allan Veras
Doutor em Genética e Biologia Molecular com ênfase em Bioinformática
Instituição: Universidade Federal do Pará (UFPA) - Campus Castanhal
Endereço: Avenida dos Universitários, s/n, Castanhal, Pará, Brasil
E-mail:allanverasce@gmail.com

Pablo de Sá
Doutor em Genética e Biologia Molecular com ênfase em Bioinformática
Instituição: Universidade Federal Rural da Amazônia (UFRA) – Campus Tomé-Açu
Endereço: Rodovia PA 140, 2428, Tomé-Açu, Pará, Brasil
E-mail:pablogomesdesa@gmail.com

Jorianne Alves
Doutora em Genética e Biologia Molecular
Instituição: Universidade do Estado do Pará (UEPA) – Campus Marabá
Endereço: Avenida Hiléia, s/nº, Agrópolis do Incra, Marabá, Pará, Brasil
E-mail:joriannealves@gmail.com

ABSTRACT
The representation of the gene content of an organism is impacted by several factors, ranging from sampling to sequencing and then the genome assembly task. The genome assembly process can generate errors that are related to insufficient coverage in the data set, an inadequate assembly methodology, and finally, errors related to the limitation of the assembly software used. Thus, some genes remain unidentified both in complete and draft genomes, this incomplete gene knowledge impacts on several organisms, mainly of medical and industrial interest, such as *Bifidobacterium breve*, a Gram-positive bacterium, found in the gastrointestinal microbiota of mammals, including humans, and has beneficial probiotic activities. Therefore, the objective of this work is to identify the new gene products not represented in the genome of *Bifidobacterium breve* DS15-17 using the raw reads of this organism. The reads were produced from the sequencing with the Illumina MiSeq platform. PAN2HGENE software was used to identify new gene products. After the analysis, 44 new gene products were identified, 26 with described function and 18 hypothetical proteins. The hypothetical proteins identified were analyzed in the ProtoNet and Superfamily databases.

Keywords: *Bifidobacterium breve,* probiotic, annotation, new products, genome.

RESUMO
A representação do conteúdo gênico de um organismo é impactada por diversos fatores, que vão desde a amostragem até o sequenciamento e depois a tarefa de montagem do genoma. O processo de montagem do genoma pode gerar erros relacionados à cobertura insuficiente no conjunto de dados, uma metodologia de montagem inadequada e, por fim, erros relacionados à limitação do software de montagem utilizado. Assim, alguns genes permanecem não identificados tanto em genomas completos quanto em drafts, esse conhecimento gênico incompleto impacta em diversos organismos, principalmente de interesse médico e industrial, como a *Bifidobacterium breve*, uma bactéria Gram-positiva, encontrada na microbiota gastrointestinal de mamíferos, incluindo humanos, e tem atividades probióticas benéficas. Portanto, o objetivo deste trabalho é identificar os novos produtos gênicos não representados no genoma de *Bifidobacterium breve* DS15-17 utilizando as leituras brutas deste organismo. As leituras foram produzidas a partir do sequenciamento com a plataforma Illumina MiSeq. O software PAN2HGENE foi usado para identificar novos produtos gênicos. Após a análise, foram identificados 44 novos produtos gênicos, 26 com função descrita e 18 proteínas hipotéticas. As proteínas hipotéticas identificadas foram analisadas nas bases de dados ProtoNet e Superfamily.

Palavras-chave: *Bifidobacterium breve,* probiótico, anotação, novo produtos, genoma.1 1

1 INTRODUCTION

Bifidobacterium breve is a Gram-positive bacterium belonging to the intestinal microbiota. This bacterium is part of the group of Lactic Acid Bacteria (LAB) that describes human and animal commensal bacteria with high employability in probiotics (Tojo et al. 2014).

The term probiotic refers to live microorganisms that when ingested in an adequate amount has beneficial properties to the health of the host (Hill et al. 2014). These control intestinal infections, have anticarcinogenic potentials, activate the immune response, enable lactose metabolism in intolerants, reduce the rates of low-density lipoproteins (LDL), among others (Wong et al. 2019).

Studies show that *B. breve* has a beneficial action concerning some diseases, such as allergies (Wong et al. 2019), rheumatoid arthritis (Achi et al. 2019) and rotavirus diarrhea (Rigo-Adrover et al. 2019), acting with a great probiotic potential for intestinal health with positive influences on other physiological systems (Leal et al.2020; Bottacini et al. 2018).

However, the cultivation of bacteria of this genus is considered complex, as they grow only in conditions of microaerophilia and require specific nutrients present in milk. Also, these conditions vary according to the strain analyzed, as in the case of some strains that are more resistant to oxygen and others that are sensitive (Underwood et al. 2017). Thus, omic studies, such as genomics, help in a more detailed understanding of the role of this bacterium in the gastrointestinal system (Bottacini et al. 2018).

Currently, 104 genomes of *B. breve*, 42 complete and 62 drafts, are available in the National Center for Biotechnology Information (NCBI) database. The sequencing of these genomes allowed a better understanding of the importance of the species, mainly for human health (Konstantinidis et al. 2006). Although this whole process aims to represent the genome with greater accuracy, there are still biases that may be inherent in the sequencing, assembly, or annotation processes. And these errors can be propagated in future studies, consequently hindering further analysis (Viana et al. 2020; Veras et al. 2018). Therefore, it is necessary to use new techniques and methodologies to identify genes that have not been represented in the genome and to infer functions to genes without a described function. Thus, the objective of this work is to perform the identification of new gene products of *Bifidobacterium breve* strain DS15-17, by *in silico* approaches, using reads from NGS platform sequencing available in public databases.

2 METHODS

2.1 REFERENCE DATA

In this work, the genome of *Bifidobacterium breve* strain DS15-17 was used, available at NCBI under the accession number NZ_QDIS00000000.1 (under nucleotide database). The size of this genome is 2,260,680 bp and it has 1,996 genes, containing 1,841 proteins, 97 pseudogenes, 53 tRNA and 5 rRNA. This bacterium was sequenced by the Illumina MiSeq platform and the raw reads are available at The European Bioinformatics Institute (EMBL-EBI) under accession number SRR5310870.

2.2 IDENTIFICATION OF NEW GENE PRODUCTS

To identify the new gene products, the computational tool PAN2HGENE (Silva de Oliveira et al. 2021) was used, which analyzes the reads that were not mapped in the reference genome, performing a *de novo* assembly of these reads and then an annotation. The gene products identified in this new assembly are analyzed against the products of the reference genome with the BLAST tool. Thus, identifying new gene products not represented in the original genome.

To confirm the new genetic products identified by PAN2HGENE, a manual validation was performed in the NCBI and Uniprot databases, which fasta sequences were analyzed with the Blast tool.

2.3 CHARACTERIZATION OF NEW HYPOTHETICAL PROTEINS

The characterization of the new hypothetical proteins identified was performed by the ProtoNet software (Rappoport et al. 2011), using default parameters. In addition, GO terms were used to identify and classify proteins in terms of their molecular functions and biological processes. The search for the similarity of the sequences was performed in the SUPERFAMILY databases (Wilson et al. 2009), aiming at the identification and annotation of protein domains.

3 RESULTS AND DISCUSSION

The annotation of the *B. breve* DS15-17 genome deposited in the NCBI contains 1,996 genes. In the analysis by the PAN2HGENE software, 44 new gene products were identified (Table 1), in which 18 products are characterized as hypothetical proteins and 26 have a described function. The complete fasta sequence of each of the 44 new products is available in the Supplementary File S1.

Table 1. List of new gene products identified with PAN2HGENE software.

Sequence ID	Product
Bifidobacterium_breve-DS15-17_2047	Tec protein
Bifidobacterium_breve-DS15-17_2069	hypothetical protein YeeN
Bifidobacterium_breve-DS15-17_2046	Alkanesulfonate ABC transporter substrate-binding protein SsuA
Bifidobacterium_breve-DS15-17_2024	GTP pyrophosphokinase
Bifidobacterium_breve-DS15-17_2027	Phage protein
Bifidobacterium_breve-DS15-17_2048	hypothetical protein
Bifidobacterium_breve-DS15-17_2029	hypothetical protein YeeN
Bifidobacterium_breve-DS15-17_2028	Similar to tetracycline resistance, MFS efflux pump
Bifidobacterium_breve-DS15-17_0528	hypothetical protein
Bifidobacterium_breve-DS15-17_2070	Dipeptidase
Bifidobacterium_breve-DS15-17_2050	Beta-lactamase class C-like and penicillin binding proteins (PBPs) superfamily
Bifidobacterium_breve-DS15-17_2072	NADP-specific glutamate dehydrogenase
Bifidobacterium_breve-DS15-17_2071	Aerobic glycerol-3-phosphate dehydrogenase
Bifidobacterium_breve-DS15-17_2074	hypothetical protein
Bifidobacterium_breve-DS15-17_2030	hypothetical protein
Bifidobacterium_breve-DS15-17_2051	Heterodimeric efflux ABC transporter, multidrug resistance => LmrC subunit of LmrCD
Bifidobacterium_breve-DS15-17_2073	Citrate lyase alpha chain
Bifidobacterium_breve-DS15-17_2054	Uncharacterized MFS-type transporter
Bifidobacterium_breve-DS15-17_2076	hypothetical protein
Bifidobacterium_breve-DS15-17_2031	hypothetical protein
Bifidobacterium_breve-DS15-17_2056	ATP synthase F0 sector subunit c
Bifidobacterium_breve-DS15-17_2078	Membrane associated histidine kinase-like ATPase
Bifidobacterium_breve-DS15-17_2055	ATP synthase F0 sector subunit b
Bifidobacterium_breve-DS15-17_2077	Acetyl-coenzyme A carboxyl transferase beta chain
Bifidobacterium_breve-DS15-17_2033	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase
Bifidobacterium_breve-DS15-17_0034	hypothetical protein

Bifidobacterium_breve-DS15-17_2057	3'->5' exoribonuclease Bsu YhaM
Bifidobacterium_breve-DS15-17_2059	hypothetical protein
Bifidobacterium_breve-DS15-17_0532	hypothetical protein
Bifidobacterium_breve-DS15-17_2017	hypothetical protein
Bifidobacterium_breve-DS15-17_2039	Methionine ABC transporter substrate-binding protein
Bifidobacterium_breve-DS15-17_0853	hypothetical protein
Bifidobacterium_breve-DS15-17_2061	prophage Lp4 protein 12
Bifidobacterium_breve-DS15-17_2063	ABC transporter, substrate-binding protein (cluster 3, basic aa/glutamine/opines)
Bifidobacterium_breve-DS15-17_2041	hypothetical protein
Bifidobacterium_breve-DS15-17_2062	NADH peroxidase Npx
Bifidobacterium_breve-DS15-17_2040	DNA-entry nuclease (Competence-specific nuclease)
Bifidobacterium_breve-DS15-17_2021	Phage portal protein
Bifidobacterium_breve-DS15-17_2043	hypothetical protein
Bifidobacterium_breve-DS15-17_2064	Uncharacterized MFS-type transporter
Bifidobacterium_breve-DS15-17_2023	hypothetical protein
Bifidobacterium_breve-DS15-17_2045	hypothetical protein
Bifidobacterium_breve-DS15-17_2066	hypothetical protein
Bifidobacterium_breve-DS15-17_2022	Zinc ABC transporter, substrate-binding protein ZnuA

3.1 PRODUCTS WITH DESCRIBED FUNCTIONS

3.1.1 Virulence factor and antibiotic resistance

The development of pathogenesis is the result of several virulence factors present in the pathogen plus its ability to adapt using the nutrients available in the host organism (Jha et al. 2019). Among the products identified (Table 1), some are related to the virulence factor due to enzyme coding of ABC transporters and periplasmic binding proteins (PBPs).

In the present work, five ABC systems were identified: ABC system of methionine substrates (Table 1 - Bifidobacterium_breve-DS15-17_2039), ABC system of zinc (Table 1 - Bifidobacterium_breve-DS15-17_2022), ABC system of alkane sulfates (Table 1 - Bifidobacterium_breve-DS15-17_2046), ABC system of glutamine (Table 1 - Bifidobacterium_breve-DS15-17_2063) and heterodimeric multidrug ABC system (Table 1 - Bifidobacterium_breve-DS15-17_2051).

ABC transporters catalyze the uptake of nutrients in microorganisms, contributing to the development of various diseases and antibiotic resistance. They work from energy derived from ATP hydrolysis and may be involved in other processes, such as sporulation, signal transduction, protein secretion and transmembrane transport (Locher 2016). And, PBPs detect essential nutrients and assist in their uptake by carriers, forming ABC systems, which can be divided into three categories according to their functions (Rahmam et al. 2019).

The ABC zinc transporter (ZnuABC) participates in the metal's absorption by the organism, in addition, zinc uptake is essential for colonization and pathology development, such as in *Vibrio cholerae* (Sheng et al. 2015).

The heterodimeric multidrug ABC system favors inherent drug resistance, is present in several species, including *Enterococcus faecalis* and *E. faecium*. The EfrAB pump is present in both, which is expressed in minimal concentrations of gentamicin, streptomycin and chloramphenicol, however, when the expression genes are repressed, the susceptibility to daunorubicin, doxorubicin and ethidium increases, proving the mediation of the pump in the control of multiple drugs in organisms (Hürlimann et al. 2016).

The ABC glutamine transport system is mediated by the GlnBP protein located in the glnPQH operon encoding the ABC system (Rahmam et al. 2019). In *Escherichia coli* it was described that the protein under stress conditions has other binding sites away from the binding bag of immeasurable importance for amino acid uptake (Lv et al. 2017).

Still, on the proteins identified in the present work, we cite acetyl-Coenzyme A carboxyl transferase (ACCase) (Table 1 - Bifidobacterium_breve-DS15-17_2077). This protein converts acetyl-CoA to malonyl-CoA in a stage of fatty acid biosynthesis (Salie and Thelen 2016). The ability to use steroids as a substrate is present in actinobacteria containing mycolic acid, and this ability is an important virulence factor. When analyzing several *M. tuberculosis* mutants, the relationship between the regulatory genes of this metabolic pathway and the expression of ACCase was verified, identifying the enzyme as an intermediate pathway (Crowe et al. 2017).

A class C Beta-lactamases and other Penicillin-Binding Proteins (PBPs) have also been identified (Table 1 - Bifidobacterium_breve-DS15-17_2050). Penicillin-binding proteins (PBPs) are a group of enzymes with affinity and ability to bind to penicillin present in the bacterial membrane, while class C beta-lactamases are part of a family of enzymes well known to target β-lactam antibiotics. Bacterial beta-lactamases deactivate the effect of β-lactam antibiotics such as penicillins, cephalosporins, monobactams and carbapenems by hydrolyzing their beta-lactam rings (Bush and Jacoby 2010). Both are the target of study due to their clinical importance since these are mechanisms of resistance to β-lactam antibiotics. This class of

antibiotics is used in 60% of cases of bacterial infection because it is one of the most effective and accessible agents (Öztürk et al. 2014; Vigouroux et al. 2019).

Another important protein concerning antibiotic resistance identified in the present work is Bifidobacterium_breve-DS15-17_2028 (Table 1), which encodes a protein similar to tetracycline resistance, whose mechanism of action is the efflux pump. Efflux pumps are protagonists in the development of bacterial resistance, which is related to resistance to several antibiotics, among them tetracycline (Cloete et al. 2018).

Tetracyclines have a broad spectrum bacteriostatic action, being used effectively in infections from both Gram-positive and Gram-negative bacteria. This class of antibiotics is widely found in the environment, especially in the soil, which may be related to the spread of antibiotic resistance (Da Rosa et al. 2021; Zhang et al. 2018).

A study, in which the genome of bacteria present in milk was sequenced, identified a gene linked to resistance to tetracycline, which encodes proteins to protect the ribosome and efflux pumps. It was analyzed that the sequenced fragments showed a great similarity with genes of the same function present in lactic acid bacteria (LAB), *Salmonella enterica* and *E. coli*. This allows presuming the occurrence of horizontal transfer, being raw milk cheeses great sources of dissemination of bacterial resistance (Florez et al. 2017).

Because of this ability to transfer horizontally, bacteria marketed as probiotics must undergo resistance tests using culture and microdilution techniques to determine the minimum inhibitory concentration (MIC). Another study demonstrated that *Bifidobacterium animalis* exhibits great resistance to the tetracycline attributed to the *tetW* gene (Morivic et al. 2018).

3.1.2 Biotechnological

B. breve is used in industry due to its probiotic potential, which is why the identification of genes related to industrial processes is of great importance (Bottacini et al. 2018). In this context, the present work identified related proteins, such as Bifidobacterium_breve-DS15-17_2071 (Table 1) which encodes the aerobic enzyme Glycerol-3-phosphate dehydrogenase (*GlpD*) involved in glycerol metabolism. In *L. monocytogenes* this process induces the formation of biofilms of concern in food processing. It was observed that in medium enriched with glycerol under aerobic conditions, *GlpD* is expressed together with glycerol kinase, while in anaerobic conditions its activities are negatively regulated (Tapia et al. 2018).

Another protein identified was Bifidobacterium_breve-DS15-17_2073 (Table 1) which encodes the product Citrate lyase alpha chain. This protein catalyzes the breakdown of citrate into acetate and oxaloacetate, and is composed of alpha, beta and gamma subunits (Schneider

et al. 2002). Citrate is present in abundance in the environment, its fermentation is carried out in eukaryotes, archaea and some pathogenic or non-pathogenic bacteria (Arora et al. 2018). The pathway involves the regulatory system, a carrier, the citrate lyase complex and an oxaloacetate decarboxylase (Martino et al. 2018).

In LABs, citrate lyase is important for its commercial use, since citrate fermentation is indispensable for the production of some foods. Like malolactic fermentation, used in wine production, which is related to citrate and the metabolites involved in its catalysis (Pretorius et al. 2019). As well as the production of high-quality chocolates, where a quick consumption of citric acid presented by the strains producing citrate lyase is necessary (Ouattara et al. 2017).

3.1.3 Survival and adaptation

Bacteria are constantly subjected to stress conditions, mainly pathogenic bacteria and bacteria belonging to the human microbiota (Jakob et al. 2007; Moloney et al. 2013). Despite this, many bacteria can adapt and proliferate through the synthesis of specific proteins induced by stress conditions (Jakob et al. 2007).

The NADH peroxidase (Npx) enzyme (Table 1 - Bifidobacterium_breve-DS15-17_2062) is a flavoprotein present in the pyridine nucleotide, Npx performs the hydrogen peroxide reduction catalysis (Keirsse-Haquin et al. 2018). In group B *Streptococcus*, a type of Npx is involved in the ability of the microorganism to survive within macrophages, which has positive regulation 24 hours after infection, and, probably, the reactive oxygen species released by the defense cells by flavoprotein are feasible, enabling survival bacteria (Korir et al. 2018). *Lactobacillus pentosus* MP-10, when exposed to the action of antimicrobials, also expresses Npx, so the protein acts in times of oxidative stress (Muñoz et al. 2016).

Another protein that may play a role in the survival of the bacterium in stressful environments is Bifidobacterium_breve-DS15-17_2024 (Table 1). This encodes a pyrophosphokinase enzyme, which performs the transfer of gamma-phosphate or beta-phosphate fractions from nucleotide precursors to originate phosphorylated products (Pokhrel et al. 2018). In *M. tuberculosis*, a GTP pyrophosphokinase, called Rel, catalyzes the synthesis of guanosine tetraphosphate (ppGpp), which helps the survival of bacteria in conditions of scarce nutrients (Bag et al. 2014).

Bifidobacterium_breve-DS15-17_2040 (Table 1) encodes a DNA entry nuclease (EndA). This is associated with the membrane and exposed to the surface, and is involved in the competence to aggregate exogenous DNA from the surrounding environment (Midon et al. 2011). In this process EndA is expressed to convert double-stranded DNA into single-stranded

fragments, the nuclease is evenly dispersed throughout the membrane and when the procedure takes place it is centralized within the cell, the result is gene transfer and genetic diversification (Peterson et al. 2013).

In addition to aggregating external DNA, the endonuclease can degrade extracellular chromatin in the form of extracellular neutrophil traps, enabling pneumococcal infection in the host, therefore, it becomes a speculated therapeutic target (Midon et al. 2011; Peterson et al. 2013). A study with *Yersinia Enterocolitica* indicated that the expression of the organism's nuclease is dependent on ions, as well as other Gram-positive pathogens (Möllerherm et al. 2015).

EndA also seems to be involved in the dispersion of biofilms, it was observed that the expansion of the structure is concomitant with the formation of empty spaces and erosion in it, indicating matrix degradation. In *P. aeruginosa*, it was evidenced that the inactivation of EndA impaired the dispersion of biofilms, indicating nuclease as an essential component of biofilm formation (Cherny and Sauer 2019).

3.2 CHARACTERIZATION OF HYPOTHETICAL PROTEINS

In this study, 18 hypothetical protein sequences were also identified, of which 10 demonstrated similarities with proteins existing in the ProtoNet database. Of these 10, four sequences were similar to proteins with available functional annotation. Also, the 18 hypothetical proteins were analyzed by SUPERFAMILY and five of these showed similarities with proteins available in the database. The results of ProtoNet and SUPERFAMILY are represented, respectively, in Tables 2 and 3.

Table 2. Result of the analysis of hypothetical protein sequences by the ProtoNet database.

Sequence ID	Similarity	Cluster Number	Cluster Name
Bifidobacterium_breve-DS15-17_2074	Protein B7P5R7	4093725	Diacylglycerol kinase accessory region
Bifidobacterium_breve-DS15-17_2076	Protein A2TLS7	4134214	Neuropeptide signaling pathway
Bifidobacterium_breve-DS15-17_2017	Protein B4I3A6	3635107	Proteinase inhibitor I7
Bifidobacterium_breve-DS15-17_0853	Protein A9US94	4139975	Proprotein convertase P
Bifidobacterium_breve-DS15-17_2023	Protein A9VR01	4169914	Collagen
Bifidobacterium_breve-DS15-17_2043	Protein A4ATF8	4169915	Collagen
Bifidobacterium_breve-DS15-17_0528	Protein B6NM30	4111733	Metridin-like ShK toxin
Bifidobacterium_breve-DS15-17_0532	Protein B6MTT5	4136768	Frizzled related
Bifidobacterium_breve-DS15-17_2030	Protein B6L842	4025134	Cryptochrome/photolyase, N-terminal domain
Bifidobacterium_breve-DS15-17_2041	Protein B6MXV1	3482469	EGF/extracellular

Table 3. Result of the analysis of hypothetical protein sequences by the SUPERFAMILY database.

Sequence ID	Superfamily	Family	Function	
			Geral	Specific
Bifidobacterium breve-DS15-17_2041	Lambda repressor-like DNA-binding domains	SinR domain-like	Regulation	DNA-binding
Bifidobacterium breve-DS15-17_2043	Major capsid protein gp5	Major capsid protein gp5	Other	Viral proteins
Bifidobacterium breve-DS15-17_2045	TPR-like	Tetratricopeptide repeat (TPR)	General	Protein interaction
Bifidobacterium breve-DS15-17_2069	C-terminal domain of PLC-beta	C-terminal domain of PLC-beta	Other	Unknow function
Bifidobacterium breve-DS15-17_2031	Zinc beta-ribbon	Transcriptional factor domain	Other	Unknow function

The sequence Bifidobacterium_breve-DS15-17_2074 (Table 2) showed high similarity with the protein B7P5R7 (putative diglycerol kinase) located in the cluster 4093725 called diacylglycerol kinase accessory region. The results showed that the B7P5R7 protein has four

molecular functions attributed by GO terms: catalytic activity; ATP link; diacylglycerol kinase activity; and NAD^+ kinase activity.

The NAD^+ kinase activity (NADK) also acts in a catalysis reaction, where the conversion of NAD^+ into $NADP^+$ occurs through phosphorylation. In *S. aureus*, analyzes showed that NADP + is an indispensable molecule in the formation of biofilms and survival of the species in several environmental conditions (Prasad et al. 2017). In *Synechocystis sp*. it has been shown that the molecule is essential for photoheterotrophic growth (Gao and Xu 2012).

Still, due to GO terms, as a result, five biological processes were attributed: Glycerolipid metabolism; Diacylglycerol metabolism; Intracellular signal transduction; Lipid phosphorylation; and protein kinase C-activating G protein-coupled receptor signaling pathway.

Regarding the metabolism of diacylglycerol, diacylglycerol kinase (Dgk), like other protein kinases, in general, catalyzes protein phosphorylation using ATP and GTP molecules (Alberts et al. 2008). Dgk is also present in *Lactobacullus reuteri*, an intestinal probiotic capable of suppressing inflammation resulting from the activation of histamine receptors, it is assumed that Dgk is released into the intestinal lumen by decreasing the amounts of diglycerol causing a reduction in protein kinase C phosphorylation (PKC). In this way, inflammatory responses mediated by histamine from cell signaling mediated by PKC are suppressed, making *L. reuteri* a potential microbial antihistamine (Ganesh et al. 2018).

Intracellular signal transduction is a process capable of identifying environmental signals and adapting cellular behavior and/or metabolism in response. The system also performs surveillance of intracellular conditions and lining structures to trigger responses to stress and neutralize adverse situations (Galperin 2018). In biofilms, several species of bacteria make use of intracellular signal transduction (Miller and Lamont 2019).

The result for the Bifidobacterium_breve-DS15-17_2076 sequence (Table 2) demonstrated similarity with the A2TLS7 protein (Receptor for egg jelly 8) located in the cluster 4134214, neuropeptide signaling pathway. The A2TLS7 protein, described in the cluster, has only the described function of molecular receptor and is an integral component of the membrane.

Bifidobacterium_breve-DS15-17_2017 (Table 2) showed similarity with the protein B4I3A6 (GM18084) belonging to cluster 3635107 called proteinase inhibitor I7. This protein has three annotations of molecular functions attributed by GO terms: Calcium ion binding; DNA binding transcription factor activity; Zinc ion bonding.

The activity of the DNA binding transcription factor occurs when a protein selectively interacts with a specific DNA sequence in the regulatory region of a gene to activate or repress transcription, in order to save cellular energy and adapt to different environments. Because they are essential to cellular functioning, several transcription factors are described in several prokaryotes (Sikder and Kodadek 2005).

And the sequence Bifidobacterium_breve-DS15-17_0853 (Table 2) showed similarity with the protein A9US94 located in cluster 4139975, proprotein convertase P. This protein has a molecular function and a biological process according to the GO term.

The protein has the molecular function of serine protease activity, performing the hydrolysis of internal alpha-peptide bonds in a polypeptide chain. Serine proteases have as their main characteristic active serine residues at their binding site. Proteases, in general, are associated with several pathogens and are investigated as promising therapeutic targets (Agbowuro et al. 2018).

Concerning the biological process, this protein participates in protein processing, which consists of any protein maturation process achieved by cleaving a peptide bond or bonds within a protein. Protein maturation is the process that leads to obtaining the total functional capacity of a protein. Post-translational modifications (PTM) have high dynamics and regulatory potential, which makes them an evolutionary solution to overcome predetermined genetic limits (Lassak et al. 2019).

The other eight sequences analyzed also had similar proteins in the database, but all of them are noted as putative uncharacterized protein. That is, without any information about its functionality. However, all of these hypothetical proteins participate in a cluster that contains information about the other associated proteins, thus being able to infer some function.

The sequences Bifidobacterium_breve-DS15-17_2023 and Bifidobacterium_breve-DS15-17_2043 (Table 2) were similar, respectively, to proteins A9VR01 (repetition of triple collagen helix) and A4ATF8. Both are uncharacterized proteins and belong to cluster 4169914, called Collagen, which has a total of 258 proteins. It is noteworthy that the sequence Bifidobacterium_breve-DS15-17_2043 also showed a result in SUPERFAMILY (Table 3), which was classified as a member of the gp5 capsid main protein superfamily, and the specific function is viral protein.

The results obtained in this analysis are expected since they are sequences related to hypothetical proteins, that is, there are still no studies in the literature proving the functionality of these proteins (Doerks et al. 2004; Hawkins and Kihara 2007). However, the similarity of the hypothetical protein sequences identified in the present work to other sequences deposited

in databases, induces that these proteins have already been identified in other organisms and that they may have similar functions in different species. Seen through homology, proteins with known function can be correlated to proteins without a described function (Da Costa et al. 2018). Thus, further work is encouraged to achieve a better understanding of hypothetical proteins.

4 CONCLUSION

The use of probiotics has gained greater relevance over the years as they are an alternative with fewer side effects than other therapeutic measures, such as antibiotics. To be classified as a probiotic, a microorganism must resist the conditions inherent in the manufacturing and physiological processes of metabolism, in addition to being able to colonize the intestinal microbiota and develop activities beneficial to the host, such as the release of substances against pathogens and metabolites favorable to intestinal health.

The analyzes carried out with the genome of *B. breve* DS15-17 allowed the identification of 44 new gene products, absent in the original annotation deposited in the NCBI database. Among the new products, we highlight the presence of 26 genes with described function, in which genes related to antibiotic resistance such as resistance to tetracycline, class C beta-lactamases and other Penicillin-Binding Proteins (PBPs) were identified. Genes related to virulence factors, biotechnological potential, and survival and adaptation were also identified. In addition to the several new gene products identified, it was also possible to validate that even in deposited genomes, the gene content may not be fully represented, which may negatively impact subsequent analyzes, especially for microorganisms of medical and biotechnological importance.

ACKNOWLEDGMENTS

Thanks to the Brazilian National Council for Scientific and Technological Development (CNPq). This work is part of the research developed by BIOD (Bioinformatics, Omics, and Development research group - www.biod.ufpa.br). JTCA thanks to Pará State University (UEPA), AAOV thanks to Federal University of Pará (UFPA) and PHCGS thanks to Federal Rural University of Amazonia (UFRA).

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

Achi S, Talahalli R, Halami P (2019) Prophylactic effects of probiotic *Bifidobacterium* spp. in the resolution of inflammation in arthritic rats. Springer Nature 103: 6287-6296.

Agbowuro A, Huston W, Gamble A and Tyndall J (2018) Proteases and Protease Inhibitors in Infectious Diseases. Med Res Ver, 38: 1295-133.

Alberts B et al (2008) Molecular biology of the cell. Garland Science Taylor & Francis, New York, NY

Arora G, Chaudhary D, Kidwai S, Sharma D and Singh R (2018) CitE Enzymes Are Essential for *Mycobacterium tuberculosis* to Establish Infection in Macrophages and Guinea Pigs. Front Cell Infect Microbiol, 6: 1-15.

Bag S, Das B, Dasgupta S and Bhadra R (2014) Mutational analysis of the (p)ppGpp synthetase activity of the Rel enzyme of *Mycobacterium tuberculosis*. Arch Microbiol, 196: 575-588.

Bottacini F, Morrissey R, Esteban-Torres M, James K, van Breen J, Dikareva E, Egan M, Lambert J, van Limpt K, Knol J, O'Connell Motherway M, and van Sinderen D (2018) Comparative genomics and genotype-phenotype associations in *Bifidobacterium breve*. Scientific reports, 8(1), 10633. https://doi.org/10.1038/s41598-018-28919-4

Bush K and Jacoby G (2010) Updated Functional Classification of Beta-Lactamases. Antimicrobial Agents and Chemotherapy, 54: 969–976.

Cherny K and Sauer K (2019) *Pseudomonas aeruginosa* Requires the DNA-Specific Endonuclease EndA To Degrade Extracellular Genomic DNA To Disperse from the Biofilm. Journal of bacteriology, 22: 201.

Cloete R, KappI E, Joubert J, ChristoffelsI A and Malan S (2018) Molecular modelling and simulation studies of the *Mycobacterium tuberculosis* multidrug efflux pump protein Rv1258c. Plos One, 13: 1-18.

Crowe A, Casabon I, Brown K, Liu J, Lian J, Rogalski J, Hurst T, Snieckus V, Foster L and Eltisa L (2017) Catabolism of the Last Two Steroid Rings in *Mycobacterium tuberculosis* and Other Bacteria. MBio, 8: 321.

da Costa WLO, Araújo CLA, Dias LM (2018) Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PLoS One, 13: 6.

Doerks T, Von Mering C, Bork P (2004) Functional clues for hypothetical proteins based on genomic context analysis in prokaryotes. J Bioinform Comput Biol, 32: 6321–6326.

Da Rosa G, Merlini L, Bessi W, Netto A, Bezerra K, Da Silva F & Santos R (2021) Resistance profile of *Staphylococcus* sp. isolates from tilapia fillet samples *Oreochromis niloticus* from the northwestern region of Paraná state, Brazil. *Brazilian Journal of Development* 7: 41731-41742.

Florez A, Vazquez L and Mayo B (2017) A Functional Metagenomic Analysis of Tetracycline Resistance in Cheese Bacteria. Front Microbiol, 24: 907.

Galperin M (2018) What bactéria want. Environ Microbiol, v. 20, n. 12, p. 4221-4229.

Ganesh B, Hall A, Ayyaswamy S, Nelson J, Fultz R, Major A, Haag A, Esparza M, Lugo M, Venable S, Whary M, Fox J and Versalovic J (2018) Diacylglycerol kinase synthesized by commensal Lactobacillus reuteri diminishes Protein Kinase C phosphorylation and histamine-mediated signaling in the mammalian intestinal epithelium. Imunol na mucosa, 11: 380-393.

Gao H and Xu X (2012) The Cyanobacterial NAD Kinase Gene sll1415 Is Required for Photoheterotrophic Growth and Cellular Redox Homeostasis in *Synechocystis* sp. Strain PCC 6803. J Bacteriol, 194: 218-24.

Hawkins T and Kihara D (2007) Function prediction of uncharacterized proteins. J Bioinform Comput Biol 5: 1–30.

Hill C, Guarner F, Reid G, et al. (2014) Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat Rev Gastroenterol Hepatol.;11(8):506-514. https://doi.org/10.1038/nrgastro.2014.66

Hürlimann L, Corradi V, Hohl M, Bloemberg G, Tieleman P and Seeger M (2016) The Heterodimeric ABC Transporter EfrCD Mediates Multidrug Efflux in *Enterococcus faecalis*. Antimicrobial Agents and Chemotherapy September, 60: 5400-5411.

Jakob K, Satorhelyi P, Lange C, Wendisch V F, Silakowski B, Scherer S and Neuhaus K (2007) Gene expression analysis of *Corynebacterium glutamicum* subjected to long-term lactic acid adaptation. Journal of bacteriology, 189: 5582–5590.

Jha B, Vyas R, Bhushan J, Sehgal D and Biswal B (2019) Structural Insights Into the Substrate Specificity of SP_0149, the Substrate-Binding Protein of a Methionine ABC Transporter From *Streptococcus Pneumoniae*. Acta Crystallogr F Struct Biol Commun, 1: 520-528.

Keirsse-Haquin J, Picaud T, Bordes L, de Gracia AG, Desbois A (2018) Modulation of the favin–protein interactions in NADH peroxidase and mercuric ion reductase: a resonance Raman study. Eur Biophys J, v. 47, n. 3, p. 205-223. https://doi.org/10.1007/s00249-017-1245-3

Konstantinidis K, Ramette A and Tiedje J (2006) The bacterial species definition in the genomic era. Phil. Trans. R. Soc. B, 36: 1929–1940

Korir ML, Flaherty RA, Rogers LM, Gaddy JA, Aronoff DM, Manning SD (2018) Investigation of the Role That NADH Peroxidase Plays in Oxidative Stress Survival in Group B *Streptococcus*. Frontiers in microbiology, 9, 2786. https://doi.org/10.3389/fmicb.2018.02786

Lassak J, Koller F, Krafczyk R and Volkwein W (2019) Exceptionally versatile – arginine in bacterial post-translational protein modifications. Biol Chem, 400: 1397-1427.

Leal A, Marelli G, Coaioto J, Maia M, Oliveira M, Garcia R, Abrahan T & Oliveira T (2020) Symbotic treatment in women with intestinal constipation: a comparative case study. *Brazilian Journal of Health Review* 3: 431-443.

Locher K (2016) Mechanistic diversity in ATP-binding cassette (ABC) transportes. Nature structural & molecular biology, 23: 487-493.

Lv D, Wang C, Li C, Tan J and Zhang X (2017) An efficient perturbation method to predict the functionally key sites of glutamine binding protein. Elsevier, 67: 62-68.

Martino G, Perez C, Magni C and Blancato V (2018) Implications of the expression of *Enterococcus faecalis* citrate fermentation genes during infection. Plos One, 13:1-18.

Midon M, Schafer P, Pingoud A, Ghosh M, Moon A, Cuneo M, London R and Meiss G (2011) Mutational and biochemical analysis of the DNA-entry nuclease EndA from *Streptococcus pneumoniae*. Nucleic Acids Research, 39: 623–634.

Miller D and Lamont R (2019) Signaling Systems in Oral Bacteria. Adv Exp Med Biol, 1197: 27-43.

Möllerherm H, Neumann A, Schilcher K, Blodkamp S, Zeitouni N, Dersch P, Lüthje P, Naim H, Zinkernagel A, von Köckritz-Blickwede M (2015) *Yersinia enterocolitica*-mediated degradation of neutrophil extracellular traps (NETs). FEMS Microbiology Letters, 362: 1-9.

Moloney R, Desbonnet L, Clarke G, Dinan T and Cryan J (2013) The microbiome: stress, health and disease. Mammalian Genome, 25: 49-74.

Muñoz M, Benomar N, Ennahar S, Horvatovich P, Lerma L, Knapp C, Gálvez A and Abriouel H (2016) Comparative proteomic analysis of a potentially probiotic *Lactobacillus pentosus* MP-10 for the identification of key proteins involved in antibiotic resistance and biocide tolerance. International Journal of Food Microbiology, 222: 8–15.

Ouattara H, Ouattara H, Droux M, Reverchon S, Nasser W, Niamke S (2017) Lactic acid bacteria involved in cocoa beans fermentation from Ivory Coast: Species diversity and citrate lyase production. Int J Food Microbiol, 1: 11-19.

Öztürk E, Ozkirimli E and Özgür E (2015) Classification of Beta-Lactamases and Penicillin Binding Proteins Using LigandCentric Network Models. Plos One, 10:1-23.

Peterson E, Kirrev D, Moon A, Midon M, Janzen W, Pingoud A and Pedersen L (2013) Inhibitors of Streptococcus Pneumoniae Surface Endonuclease EndA Discovered by High-Throughput Screening Using a PicoGreen Fluorescence Assay. J Biomol Screen, 18: 247-57.

Pokhrel A, Poudel A and Purcell E (2018) A Purification and In Vitro Activity Assay for a (p)ppGpp Synthetase from Clostridium difficile. Journal of visualized experiments, 3: 1-8.

Prasad UV, Vasu D, Gowtham RR, et al. (2017) Cloning, Expression and Characterization of NAD Kinase From *Staphylococcus aureus* Involved in the Formation of NADP (H): A Key

Molecule in the Maintaining of Redox Status and Biofilm Formation. Adv Biomed Res, v. 31, n. 6, p. 97. https://doi.org/10.4103/2277-9175.211833

Pretorius N, Engelbrecht L and Toit M (2019) Influence of sugars and pH on the citrate metabolism of different lactic acid bacteria strains in a synthetic wine matrix. Journal of Applied Microbiology, 127: 1490-1500.

Rahmam M, Machuca M and Roujeinikova A (2019) Preliminary X-ray crystallographic studies on the *Helicobacter pylori* ABC transporter glutamine-binding protein GlnH. Drug Discoveries & Therapeutics, 13: 52-58.

Rappoport N, Karsenty S, Stern A, Linial N and Linial M (2011). ProtoNet 6.0: organizing 10 million protein sequences in a compact hierarchical family tree. Nucleic Acids Res.;40 (Database issue): D313-D320. https://doi.org/10.1093/nar/gkr1027

Rigo-Adrover M, Knipping K, Garssen J, Limpt K, Knol J, Franch A, Castell M, Rodríguez-Lagunas M and Pérez-Cano F (2019) Prevention of Rotavirus Diarrhea in Suckling Rats by a Specific Fermented Milk Concentrate with Prebiotic Mixture. Nutrients, 11: 2-15.

Salie MJ and Thelen JJ (2016) Regulation and structure of the heteromeric acetyl-CoA carboxylase. Elsiever, v. 9, p. 1207-1213.

Schneider K, Kästner C, Meyer M, Wessel M, Dimroth P and Bott M (2002) Identification of a Gene Cluster in *Klebsiella pneumoniae* Which Includes citX, a Gene Required for Biosynthesis of the Citrate Lyase Prosthetic Group. Journal of bacteriology, 184: 2439–2446.

Sheng Y, Fan F, Jensen O, Zhong Z, Kan B, Wang H and Zhu J (2015) Dual zinc transporter systems in *Vibrio cholerae* promote competitive advantages over gut microbiome. Infect Immun, 83: 3902-3908.

Sikder D and Kodadek T (2005) Genomic studies of transcription factor–DNA interactions. Current Opinion in Chemical Biology, 9: 38-45.

Silva de Oliveira M, Thyeska Castro Alves J, Henrique Caracciolo Gomes de Sá P & Veras A (2021) PAN2HGENE–tool for comparative analysis and identifying new gene products. *PLOS ONE* 16: e0252414.

Tapia N, Besten H and Abee T (2018) Glycerol metabolism induces *Listeria monocytogenes* biofilm formation at the air-liquid interface. Int J Food Microbiol, 20: 20-27.

Tojo R, Suárez A, Clemente MG, et al. (2014) Intestinal microbiota in health and disease: role of bifidobacteria in gut homeostasis. World J Gastroenterol. 20(41):15163-15176. https://doi.org/10.3748/wjg.v20.i41.15163

Underwood MA, Davis J, Kalanetra KM, Gehlot S, Patole S, Tancredi DJ, Mills DA, Lebrilla CB, Simmer K (2017) Digestion of Human Milk Oligosaccharides by *Bifidobacterium breve* in the Premature Infant. Journal of pediatric gastroenterology and nutrition, 65(4), 449–455. https://doi.org/10.1097/MPG.0000000000001590

Veras A, Merlin B, de Sá P (2018) ImproveAssembly - Tool for identifying new gene products and improving genome assembly. Plos One 13: 2-10.

Viana P, Carvalho E, Nascimento W, Filho R, Oliveira D, Assunção E & Filho S (2020) GENOMIC ANNOTATION ANALYSIS SOFTWARE – GANAS: A TOOL FOR GENOMIC ANNOTATION ANALYSIS. *Brazilian Journal of Development* 6: 89125-89134.

Vigouroux A, Cordier B, Aristov A, Alvarez L, Ozbaykal G, Chaze T, Oldewurtel E, Matondo M, Cava F, Bikard D, Teeffelen S (2019) Class-A penicillin binding proteins do not contribute to cell shape but repair cellwall defects. Elife, 9: 1-26.

Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, Chothia C, Gough J (2009) SUPERFAMILY-sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res, 37: 380-386.

Wong C, Iwabuchi N, Xiao J (2019) Exploring the Science behind *Bifidobacterium breve* M-16V in Infant Health. Nutrients, 11: 2-24. https://doi.org/10.3390/nu11081724

Zhang Y, Boyd S, Teppen B, Tiedje J, Zhang W, Zhu D, Li H (2018) Bioavailability of tetracycline to antibiotic resistant *Escherichia coli* in water-clay systems. Environ Pollut, 243: 1078-1086.