NNT : 2017SACLE035

Comparative genomics of *Faecalibacterium spp.*

Thèse de doctorat de l'Université Paris-Saclay
préparée à Nom de

n°581 : agriculture, alimentation, biologie, environnement et santé (ABIES)

Thèse présentée et soutenue à Ville Belo Horizonte - Brazil, le 24 Mai 2018, par

Benevides Leandro

Composition du Jury :

CHATEL Jean-Marc PhD, INRA (MICALIS)	Président
FERNANDES Gabriel PhD, Fundação Oswaldo Cruz (Centro de Pesquisas René Rachou)	Rapporteur
JAN Gwenael PhD, INRA (STLO)	Rapporteur
CHATEL Jean-Marc PhD, INRA (MICALIS)	Examinateur
SOARES Siomar PhD, UFTM (Department of Immunology and Parasitology)	Examinateur
AZEVEDO Vasco PhD, UFMG (Department of Biology)	Directeur de thèse
LANGELLA Philippe PhD, INRA (MICALIS)	Co-Directeur de thèse

Thèse de doctorat

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

PhD Thesis

COMPARATIVE GENOMICS OF *Faecalibacterium* spp.

Student: LEANDRO DE JESUS BENEVIDES
Supervisors: VASCO ARISTON DE CARVALHO AZEVEDO
PHILIPPE LANGELLA
Co-Supervisors: SIOMAR DE CASTRO SOARES
JEAN-MARC CHATEL

BELO HORIZONTE
MAY – 2018

LEANDRO DE JESUS BENEVIDES

COMPARATIVE GENOMICS OF *Faecalibacterium* spp.

Thesis presented as partial requirement for the degree of Doctor in Genetics, to the Department of General Biology at the Institute of Biological Sciences, Federal University of Minas Gerais and the degree of Doctor in “Sciences du Vivant” to the University of Paris-Saclay.

Student: LEANDRO DE JESUS BENEVIDES
Supervisors: VASCO ARISTON DE CARVALHO AZEVEDO
PHILIPPE LANGELLA
Co-Supervisors: SIOMAR DE CASTRO SOARES
JEAN-MARC CHATEL

BELO HORIZONTE
MAY – 2018

I dedicate this work to my mom, who always encourage me to pursuit my dreams and always guided me in the academic life.

ACKNOWLEDGEMENTS

I would like to thank everybody who worked with or stood by me since when I first
decided to begin my research way of life.

I would to specially thank to:

- My grandmothers. Matriarchs of two wonderful families.

- My parents. For its unconditional love, for the trust, support and constant encouragement.

- My sister, by the smiles, cries, hugs, conversations, silence ... Thank you for being my life.

- My family for, each one in its own way, always support me.

- My friends, especially Gabriel, Lucas, Marcus and Victor, for the adventures lived and for always being close, even when far away.

- To Priscilla, for all the patience, care, attention and much love she gave me.

- To my supervisors, for always helping with their knowledge and accurate observations.

“Not Everything That Is Faced Can Be Changed, but Nothing Can Be Changed Until It Is Faced.”

JAMES BALDWIN

Table of Contents

Abbreviations

AD	Atopic Dermatitis
ANI	Average Nucleotide Identity
ARDB	Antibiotic Resistance Genes Database
BPA	Bisphenol A
CAPES	*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*
CARD	Comprehensive Antibiotic Resistance Database
CD	Crohn's Disease
CNPq	*Conselho Nacional de Desenvolvimento Científico e Tecnológico*
CoA	Acetyl-coenzyme A
CRC	Colorectal Cancer
DDH	DNA-DNA Hybridization
EOS	Extremely Oxygen Sensitive
GEIs	Genomic Islands
GIT	Gastrointestinal Tract
GOLD	Genomes Online Database
IBD	Inflammatory Bowel Diseases
IBS	Irritable Bowel Syndrome
IFN-γ	Interferon *gamma*
INRA	*Institut National de la Recherche Agronomique*
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCB	Locally Collinear Blocks
LGCM	Laboratory of Molecular and Cellular Genetics
MAM	Microbial Anti-inflammatory Molecule
MI	Metabolic Islands
ML	Maximum- Likelihood
MP	Metabolic Pathways
NGP	Next Generation Probiotic
NGS	Next-Generation Sequencing

PAI	Pathogenicity Islands
PATRIC	Pathosystems Resource Integration Center
PPARγ	Peroxisome Proliferator-Activated Receptor *gamma*
QPS	Qualified Presumption of Safety
RI	Antibiotic Resistance Islands
SCFA	Short Chain Fatty Acid
SI	Symbiotic Islands
UC	Ulcerative colitis
UFMG	Federal University of Minas Gerais
wgMLST	whole genome Multi-Locus Sequence Typing

Abstract

Within the human colon, the genus *Faecalibacterium* is the main member of the *Clostridium leptum* cluster and comprises the second-most common representative genus in fecal samples, after *Clostridium coccoides*. It has been recognized as an important bacterium promoting the intestinal health and today is considered as a potential next generation probiotic. Until recently, it was believed that there was only one species in this genus, but since 2012, some studies have begun to suggest the existence of two phylogroups into the genus. This new proposition of reclassification into this genus increases the importance of new studies, with all strains, to better understand the diversity, the interactions with the host and the safety aspects in its use as probiotic. Briefly, in this work we introduce the comparative genomics analyzes to the genus *Faecalibacterium* performing a deep phylogenetic study and evaluating the safety aspects for its use as a probiotic. The phylogenetic analyzes included not only the classical use of 16S rRNA gene, but also the utilization of 17 complete genomes and techniques like whole genome Multi-Locus Sequence Typing (wgMLST), Average Nucleotide Identity (ANI), gene synteny, and pangenome. Also, this is the first work to combine an analysis of pangenome development with ANI analysis in order to corroborate the assignment of strains to new species. The phylogenetic analyzes confirmed the existence of more than one species into the genus *Faecalibacterium*. Moreover, the safety assessment involved the (1) prediction of horizontally acquired regions (Antibiotic resistance islands, Metabolic islands and phage regions), (2) prediction of metabolic pathways, (3) search of genes related to antibiotic resistance and (4) search of bacteriocins. These analyzes identified genomic islands in all genomes, but none of than are exclusive to one strain or genospecies. Also, were identified 8 genes related to antibiotic resistance mechanisms distributed among the genomes. 126 metabolic pathways were predicted and among than some were highlighted: Bisphenol A degradation, Butanoate metabolism and Streptomycin biosynthesis. In addition, we studied the genomic context of one protein (Microbial Anti-inflammatory Molecule - MAM) first described by our group. This

investigation shows that MAM appears close to genes related to sporulation process and, in some strains, close to an ABC-transporter.

I Presentation

I.1 Collaborators

This work was performed on the Laboratories of Molecular and Cellular Genetics (LGCM), at Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil; the Commensals and Probiotics-Host Interactions Laboratory, at Micalis Institute, *Institut National de la Recherche Agronomique* (INRA), Jouy-en-Josas, France; and the Department of Immunology, Microbiology and Parasitology, at Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, Brazil

The work was supported by: *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq).

II Preface

II.1 State of the Art

With the advent of new sequencing technologies (Next-Generation Sequencing - NGS) in 2005, there was a massive deposit of whole genomes in public databases and the large number of genomes available improved the development of comparative genomics (Metzker, 2010). Combined to the accelerated deposit of prokaryotic genomes, it is noticed a great interest in generating data related to organisms associated to human health. Among them, microorganisms of the intestinal microbiota as *Faecalibacterium prausnitzii*. *F. prausnitzii* is a commensal bacteria and is the main member of the *Clostridium leptum* subgroup, the most representative in the fecal samples (Tap et al., 2009; Walker et al., 2011a). This bacterium has aroused great interest due to its immunomodulatory properties and its role as indicator of, and an actor in, intestinal health (Khan et al., 2012; Sokol et al., 2008b). Besides that, the abundance and presence of *F. prausnitzii* in various mammals than humans, such as pigs (Castillo et al., 2007), mice (Nava and Stappenbeck, 2011), calves (Oikonomou et al., 2013) and chickens (Scupham, 2007), suggests that this bacteria is a functionally important member of the microbiota and possibly have an impact on the physiology and health of the host. Changes in the abundance of *F. prausnitzii* have been widely described in various intestinal and metabolic diseases in humans such as colorectal cancer (CRC), Crohn's disease (CD) and ulcerative colitis (UC) (Miquel et al., 2013; Rajilić-Stojanović et al., 2011; Sokol et al., 2008b). Due its ubiquity and beneficial effect in intestinal health, some studies suggests the use of *F. prausnitzii* as probiotic (Miquel et al., 2013, 2014; Sokol et al., 2008b).

Due the importance of *F. prausnitzii* in human health, is important to keep studying this bacterium to better understand its commensal relationship, to answer some unresolved questions and to evaluate its probiotic potential. Besides, there were a lack of knowledge of genetic diversity and genomic data of this species.

Based on this, this work combined efforts to perform analyzes based on comparative genomic to better characterize the genomes of different strains of *F. prausnitzii*. This approach can help to understand the genetic diversity, lifestyle and the probiotic properties of these species.

II.2 Manuscript Structure and author's contributions

The thesis is divided into Introduction and 2 chapters based on research articles, as follow:

a. The Introduction shows an overview of the genus *Faecalibacterium* and the comparative genomics approaches explored in the research articles presented here;
b. The first chapter presents a research article showing the phylogenetic study of 17 genomes of *Faecalibacterium* spp*.* through different strategies like 16S rRNA, whole-genome Multi-Locus Sequencing Type (wgMLST), Average Nucleotide Identity (ANI), Pangenome and genome synteny;
c. The article in the second chapter describes the safety assessment properties of 18 genomes of *Faecalibacterium* spp.. In this work were performed comparative genomic analyzes like the prediction of genomic islands and phage regions, search of bacteriocins and antibiotic resistance genes and also the prediction of metabolic pathways;
d. Additional files for the research articles are appended after the referred article, in the same section. After the chapters, we present the general conclusions of the work. Finally, after the references, there is an "appendices" section, where one can find the *curriculum vitae*.

III Introduction

III.1 Intestinal Microbiota

III.1.1 Microbial diversity in the GI tract

The human gut microbiota has become the subject of extensive research in recent years and our knowledge of the resident species and their potential functional capacity is rapidly growing. Our gut harbors a complex community of over 100 trillion microbial cells which influence human physiology, metabolism, nutrition and immune function while disruption to the gut microbiota has been linked with gastrointestinal conditions such as inflammatory bowel disease and obesity (Guinane and Cotter, 2013).

It has been estimated that our gut contains in the range of 1000 bacterial species and 100-fold more genes than are found in the human genome. The gastrointestinal tract (GIT) is populated by a complex community of microorganisms, classified as being transient or indigenous and that has co-evolved with the host over thousands of years to form an intricate and mutually beneficial relationship (Thursby and Juge, 2017). The microorganisms considered as transient are not capable to survive in the difficult conditions find in the GIT for more than a few days, so does not colonize. However, those microorganisms are often found in the GIT by the fact that are present in the food consumed daily by humans, as yogurts, cheeses and other fermented food. Otherwise, the indigenous microorganisms are adapted to the conditions found in the GIT and can survive and colonize the host (Artis, 2008; Ley et al., 2006; Velasquez-Manoff, 2015).

The composition of the microbiota in the GIT varies a lot between species and inside each species. The majority of the human intestinal bacteria belong to two phyla, the *Bacteroidetes* and the *Firmicutes* (Mariat et al., 2009). The phylum *Bacteroidetes* consists of three classes, of which the class *Bacteroidetes*, containing the well-known genera *Bacteroides* and *Prevotella*, is probably the most well studied (Underwood et al., 2015). The *Firmicutes* includes the

Faecalibacterium prausnitzii, which represents 3.5% of the microbiota in GIT, being the most abundant specie (Arumugam et al., 2011; Lay et al., 2005) and presents anti-inflammatory properties (Martín et al., 2014a, 2015; Quévrain et al., 2017; Sokol et al., 2008b).

Host physiology and intestinal microbiota are intimately connected. This is evident from the fact that each distinct anatomical region along the GI tract is characterized by its own physicochemical conditions, and that these changing conditions exert a selective pressure on the microbiota. The bacterial amount generally increases going down the gastrointestinal tract: 10^2–10^3 per ml in the highly acidic environment of the stomach, to 10^5 per ml in the upper small intestine and up to 10^{12} per ml in the colon. However, the terminal ileum might contain higher numbers of bacteria than in the colon. The distal colon is the site of the lowest diversity and the caecum with the highest one. While the small intestine is prevalent populated by aerobic species, there is a dominance of anaerobic species in the colon, consistent with the offer of oxygen on those sites (Figure 1)(Mowat and Agace, 2014).

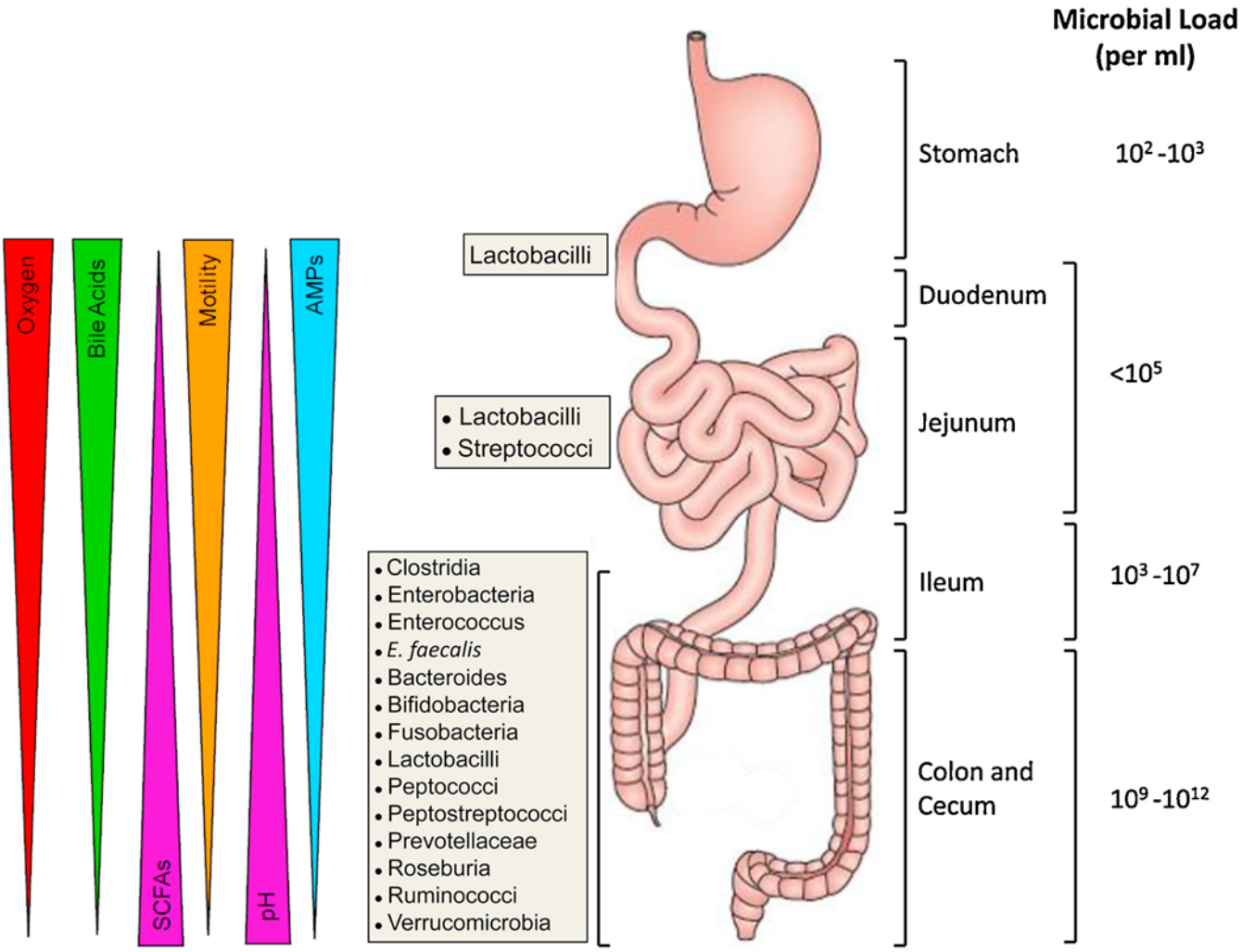

Figure 1 - Distribution of environmental factors along the length of the intestine. Figure from (Reinoso Webb et al., 2016).

The development of the microbiota is generally believed to begin from birth, from various sources, especially from gestational conditions, model of delivery, breasting and from searching environment with mouth. In addition, life events such as illness, antibiotic treatment and changes in diet can causes chaotic shifts in the microbiota (Thursby and Juge, 2017). By around 3 years of age, the composition, diversity and functional capabilities of the infant microbiota resemble those of adult microbiota, with considerable influence of age, diet, health status, stress and other conditions (Rodríguez et al., 2015). The part of inheritance of microbiota remains unclear and studies are contradictory. It is a challenge to quantify which part of the microbiota is issued from the mother,

function of host genes or dependent of the environment (Ley et al., 2005; Turnbaugh et al., 2009; Villanueva-Millán et al., 2015; Zoetendal et al., 2001).

III.1.2 Relationships between host and microbiota

The intestinal microbiota of healthy individuals can confer several health benefits relating to, for example, pathogen protection, nutrition, host metabolism and immune modulation. The extremely divergent arrangement of this gut microbiota between individuals is known and it has been described that the functional gene profiles are comparable, suggesting that the knowledge about of the metabolic activity of microbiota components could be more pertinent than its taxonomical composition (Guinane and Cotter, 2013). The key roles of the microbiota can be generally divided into three groups: (I) metabolic: metabolism of lipids and cholesterol, cleavage of some polysaccharides and dietary fibers from plants with production of compounds as butyrate [anti-inflammatory properties], metabolism of polyphenols [antioxidant and beneficial actions] and synthesis of amino acids and vitamins; (II) protection of the host against colonization by exogenous pathogens and potentially harmful indigenous microorganisms by competition, modulation of the immune system, production of antimicrobial peptides and bacteriocins; and (III) trophic: modulation of the proliferation, differentiation, maturation reduction of apoptosis of colonic epithelial cells (Goldsmith and Sartor, 2014; Guarner et al., 2006; Moreno-Indias et al., 2014; Velasquez-Manoff, 2015; Villanueva-Millán et al., 2015).

Perturbation of the microbiota composition, also known as dysbiosis, has been recognized in various diseases, direct associated or not with the GI tract (Gerritsen et al., 2011). Complex diseases as autism, Inflammatory Bowel Diseases (IBD) or obesity can be associated with this composition of the microbiota. In autism, it has been described a protective effect of *Bacteroides fragilis* in the development of the disease (Hsiao et al., 2013). *Bacteroides*

fragilis and *Faecalibacterium prausnitzii* are associated with healthy intestinal microbiota and are reduced in patients with IBD, which is characterized by diarrhea, abdominal pains and change in weight. Action of *Bacteroides fragilis* is mediated by polysaccharide A, which has protective effect on colitis (Mazmanian et al., 2008). In obesity, the transfer of the microbiota from obese mice to germ free mice led to a higher weight intake (Turnbaugh et al., 2006). It has also been shown that lean patients present a different microbiota composition when compared with obese patients(Ley et al., 2005). In the obese patients was identified a reduction of Bacteroidetes and an increase of Firmicutes and Proteobacteria (Ley et al., 2005).

III.2 Inflammatory Bowel Diseases

Inflammatory bowel disease (IBD) is a group of illnesses typified by a chronic bowel inflammatory disorder. Ulcerative colitis (UC) and Crohn's disease (CD), the primary constituents of IBD, are precipitated by a complex interaction of environmental, genetic, and immunoregulatory factors (Hanauer, 2006). UC is limited to the colon and is characterized by mucosal inflammation in a superficial way. CD typically causes transmural inflammation, affecting all the layers of the intestinal wall and can affect any region of the gastrointestinal tract in a discontinuous way. CD is normally related with the presence of strictures, abscesses and fistulas as complications. Beside these differences, both diseases present similar symptoms like diarrhea, abdominal pain, rectal bleeding and weight loss. These symptoms on relapse can continue for days, weeks or even months (Cho, 2008; Falvey et al., 2015; Stepaniuk et al., 2015; Tontini et al., 2015). The great diversity of symptoms within UC and CD suggests distinct underlying pathogenetic mechanisms. It is hoped that a better understanding of the environmental, genetic, and immunological mechanisms that produce UC and CD will lead to improved therapy for IBD (Hanauer, 2006). Others chronic inflammatory disorders are highly associated with IBD, like

osteoarthritis and psoriasis, and also complications such as colorectal cancer or blindness (Bernstein et al., 2005; Desai, 2014; Mintz et al., 2004).

III.2.1 Epidemiology

IBD is considered as a global public health problem with a variation of the incidence across countries. More than five million people around the world is affected by UC and CD with 3 million in Europe and 1.4million only in the US. Countries with low incidence are found in Asia, South America and southern and eastern Europe. The number of cases around the world is augmented year after year in both pediatric and adult patients (Ananthakrishnan, 2015; Burisch and Munkholm, 2015; Cosnes et al., 2011; Lovasz et al., 2013; Molodecky et al., 2012; Shivananda et al., 1996) (Figure 1). The peak age of onset for IBD is 15 to 30 years old, although it may occur at any age. About 10% of cases occur in individuals ,18 years old. Both UC and CD have a bimodal age distribution, with a second, smaller peak occurring in individuals ages 50 to 70 years (Hanauer, 2006).

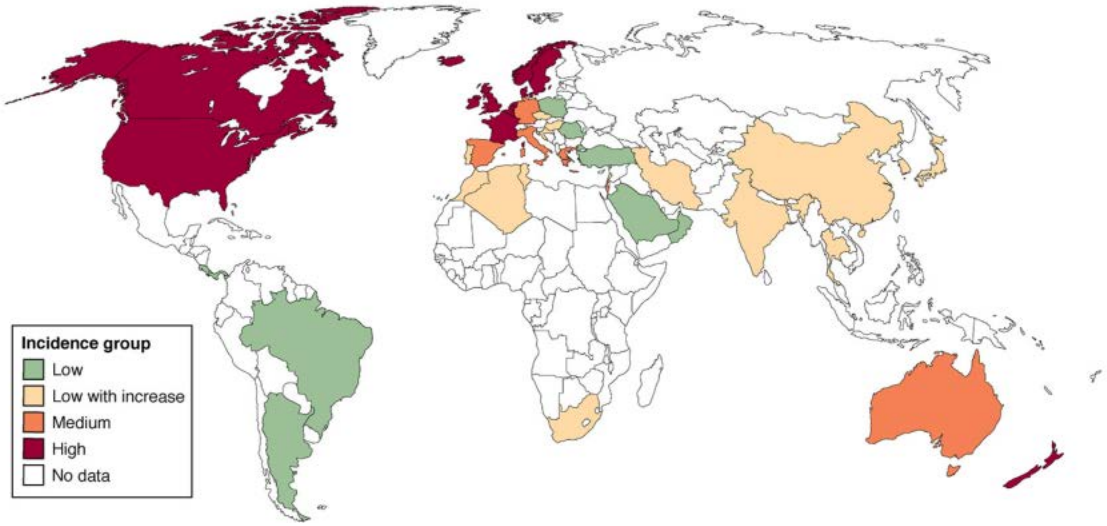

Figure 2 - The global map of inflammatory bowel disease: red refers to annual incidence greater than 10/105, orange to incidence of 5–10/105, green to incidence less than 4/105, yellow to low incidence that is continuously increasing. Absence of color indicates absence of data (Cosnes et al., 2011).

III.2.2 IBD Pathogenesis

It is likely that several factors contribute to the development of mucosal inflammation. Also, variations in influence may account for the clinical diversity seen in UC and CD. Among the distinct aspects in the development of the diseases, some of the most important are: immune response, intestinal microbiota and genetic susceptibility (Zhang and Li, 2014).

III.2.2.1 Immune responses

The investigation of IBD pathogenesis has been dominated for a long time by the studies of mucosal immunity, especially the T cell response. Available evidence suggests that the dysfunctions of innate and adaptive immune pathways contribute to the aberrant intestinal inflammatory response in patients with IBD (Zhang and Li, 2014).

The patients with IBD present alterations in the expression and function of PRRs such as toll-like receptors (TLRs) on the cell surface and NOD-like receptors in the cytoplasm, leading, for example, to the reduced production of antibacterial agents and increase of pathogenic microbial invasion (Abraham and Cho, 2006; Bonen et al., 2003; Noguchi et al., 2009; Watanabe et al., 2004; Wehkamp et al., 2004).

The mucosal immunity, particularly the T cell response, has been studied in the IBD pathogenesis. The Th1 and Th17 responses have been associated with CD, characterized by the production of IL-12, IL-23, IL-27 and IFN-γ, while a

non-conventional Th2 response has been considered in UC with an overexpression of IL-4 and IL-13 (Cobrin and Abreu, 2005; Koren et al., 2013). Prominent levels of IL-17A have been detected in the mucosa of patients with CD and UC, but its activity on IBD seems to be contradictory.

III.2.2.2 Intestinal microbiota and dysbiosis

Many studies have examined the gut microbiota in CD and UC in both inflamed and non-inflamed segments, and found that there is a significantly reduced biodiversity in fecal microbiome in IBD patients compared to that in healthy controls (Zhang and Li, 2014). Among those differences we can highlight the lower quantity of *Faecalibacterium prausnitzii* and Lactobacilli. In patients with UC, it is detected an increase of *Escherichia coli* and a decrease in *Clostridium* spp., while in CD patients it is reported a significant reduction in the Firmicutes and Bacteroidetes and an increase of enterobacteria (Eckburg et al., 2005; Joossens et al., 2011; Martín et al., 2013; Martinez et al., 2008; Sokol et al., 2008b, 2008a; Walker et al., 2011b; Wallace et al., 2014).

Patients with IBD present a reduction in the quantity of bacterial species responsible to produce Short Chain Fatty Acids (SCFA) such as species from Clostridium groups, mainly *F. prausnitzii*. The genes involved in the metabolism of SCFA, like butyrate (that play a key role on the maturation of regulatory T cells) is reduced in patients with IBD (Chang and Lin, 2016; Sokol et al., 2008b; Takahashi et al., 2016; Vangay et al., 2015). Reduction of certain species in the microbiota, as *F. prausnitzii*, could be used as reliable clinical marker in IBD because this reduction is commonly observed in patients during period of active disease or remission. Species from Lactobacillus and Bifidobacteria genus are find to be reduced in IBD with an important impact in the patients because they are important to reduce the expression of pro-inflammatory cytokines in the GIT (Manichanh et al., 2012; Rehman et al., 2016).

III.2.2.3 Genetic susceptibility

In the last decades advances in the DNA sequencing and the analysis of these data have been made, allowing the acknowledge of the genetic contributions to IBD. This is due to the technological advances in DNA analysis and sequencing and the use of huge multinational databases (Duerr, 2007). The Single Nucleotide Polymorphisms (SNPs) are identified through many Genome-Wide Association Studies (GWAS) making possible the identification of genes associated with UC and CD. Studies of gene loci shared by UC and CD may provide new way to find their common pathogenesis (Zhang and Li, 2014). Most part of these genes are associated with cytokine receptor signaling, barrier function or T cell activation and are involved in IBD susceptibility in around 5% of world population (Jostins et al., 2012; Koren et al., 2013; Liu and Stappenbeck, 2016; Ward et al., 2016).

III.2.3 Treatment

Current treatment of IBD may involve the administration of high-dose steroids, immunomodulators (e.g., azathioprine; 6-mercaptopurine), and surgery. Advances in our understanding of the pathogenesis of IBD introduce the possibility for targeted therapy to interrupt the inflammatory cascade (Hanauer, 2006). The corticoids are the immunosuppressive most used, but at long time they can induce several side effects because they are derived from cortisol, a hormone implicated in numerous metabolic functions in the host. Those side effects include mood changes, headache, hyperglycemia, vomiting and weight gain. Besides that, those drugs turn the patient more vulnerable to get infections by the fact that they jeopardize the immune system of the host (Bernstein, 2015; Orlicka et al., 2013; Thia et al., 2008).

Antibiotics can be used in the treatment for some complications in IBD, like fistulas, abscesses and infections by intestinal pathogens. Ciprofloxacin and metronidazole are broad spectrum antibiotics against gram-negative and gram-positive bacteria that are commonly used in the treatment of IBD. Besides their use in the clinical practice, there is controversies about the efficiency to reduce the general symptoms and eradicate dysbiosis because some weeks after the end of the treatment there is a return of the IBD signs (Khan et al., 2011; Rahimi et al., 2007; Sartor, 2010).

In the worst case, a chirurgical intervention is necessary to remove part of the colon and rectum. But even after the surgery it is necessary to continue with the drugs to avoid the return of symptoms. Nevertheless, a high number of patients present a relapse even after surgery and drugs treatment (Sokol et al., 2008b).

Regarding the fact that the recent treatments have no complete effectiveness and present several critical side effects, it is essential that research find new approaches to treat the patients with more safety and strong results (Marteau, 2006; Nielsen and Munck, 2007; Thia et al., 2008).

As formerly pronounced, there is a loss of microbiota diversity in patients with IBD. Considering that, to reverse this microbiota issue, an innovative therapy used has been the fecal material transplant, which consists in the implantation of the microbiota from a healthy patient to an IBD patient. This approach presented success against infection by *Chlostridium difficile*. For IBD, the new technique has shown promising results with decrease of symptoms, disease remission and allowing patients to stop the medication. Nevertheless, studies in large scale must be done to prove the causality of cure after fecal material transplant (Anderson et al., 2012; Ianiro et al., 2014; Youngster et al., 2014).

III.3 *Faecalibacterium* spp.

III.3.1 Taxonomy history

In 1922, the German bacteriologist Otto Carl Willy Prausnitz, also known as Carl Prausnitz-Giles, isolated an organism from pus from a case of pleural empyema which he named *Bacillus mucosus anaerobius*. Hauduroy et al., in 1937, proposed that this organism should be placed in the genus Bacteroides as *Bacteroides praussnitzii* because it was an anaerobic, gram-negative, non-sporeforming rod (Cato et al., 1974). In 1973, the International Subcommittee on Taxonomy of the Anaerobic, Gram-Negative Rods recommended that the anaerobic, gram-negative, non-sporeforming rods whose major metabolic product is butyric acid without isobutyric and isovaleric acids should be placed in the genus *Fusobacterium*. As all strains classified as *B. praussnitzii* in this time, were anaerobic, gram-negative, non-sporeforming rods that produce large amounts of butyric acid and no isobutyric or isovaleric acid, this species have been reclassified as *Fusobacterium prausnitzii* (Cato et al., 1974). In 1996, a 16S rRNA gene analysis of a strain found in humans (ATCC 27766) and other *Fusobacterium* strains showed that *Fusobacterium prausnitzii* have a close relationship with other members of the *C. leptum* cluster (phylum Firmicutes, class Clostridia, family *Ruminococcaceae*; (WANG et al., 1996). After that, Duncan et al. (2002) evaluated the growth requirements and the fermentation products of different strains of *Fusobacterium prausnitzii* and also performed a phylogenetic analysis based on 16S rRNA gene sequences. The results of these analyzes leaded Duncan and collaborators to suggest the creation of the genus *Faecalibacterium* and a proposal to reclassify the *Fusobacterium prausnitzii* species as *Faecalibacterium prausnitzii* (Duncan et al., 2002). In 2012 new studies based on 16S rRNA sequence analyses started to suggest the existence of two phylogroups within this genus (Lopez-Siles et al., 2012).

This new proposal was followed by other studies that correlated changes in the abundance of *Faecalibacterium prausnitzii* with different diseases considering the existence of the phylogroups I and II (Hippe et al., 2016; Lopez-Siles et al., 2016; Song et al., 2016). The contribution of *F. prausnitzii* phylogroups I and II in intestinal disease showed that the phylogroup I was depleted in Crohn's

disease (CD), ulcerative colitis, and colorectal cancer, whereas phylogroup II was specifically reduced in CD (Lopez-Siles et al., 2016). The differences between phylogroups abundance were observed not only in intestinal desease. Hippe et al. (2016), for example, studied *F. prausnitzii* phylotypes faeces of obese and type two diabetes with similar eating behavior compared to a lean control group. They showed that *F. prausnitzii* phylotypes differed in obese with and without developed diabetes type two and these phylotypes may have an influence on developing type two diabetes and might also act as starting points for prevention and therapy of obesity associated disease (Hippe et al., 2016).

III.3.2 Ecology and Diversity

Members of genus *Faecalibacterium* spp*.* are commensal bacteria, ubiquitous in the gastrointestinal tract of humans and other in vertebrate animals such as pigs (Castillo et al., 2007), mice (Nava and Stappenbeck, 2011), calves (Oikonomou et al., 2013) and chickens (Scupham, 2007). Within the human colon, this taxon is the main member of the cluster *Clostridium leptum* and comprises the second-most common representative in fecal samples, after *Clostridium coccoides* (Tap et al., 2009; Walker et al., 2011a).

Faecalibacterium spp*.* are dominant in healthy adults and the population in the intestine is modulated by diverse factors. Some studies (Fallani et al., 2006; Hopkins et al., 2005; Koga et al., 2016) showed that the populations of this bacterium change with age. There is no detection of specific RNA in stools from babies up to the age of 6 months. After that, the value of RNA can be detected, and it increases between ages 6 and 24 months but remains low until 2–3 years old in children (Figure 2).

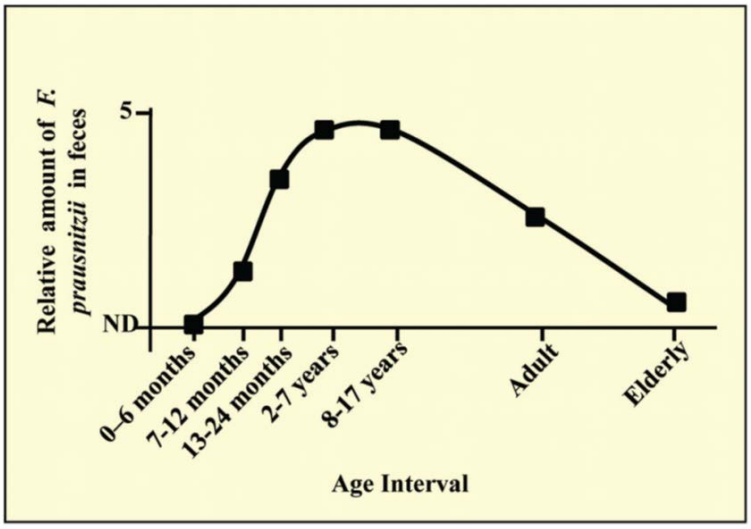

Figure 3 - Kinetics of implantation of *F. prausnitzii*. Changes in human fecal *F. prausnitzii* populations with host age (Miquel et al., 2014)

The low amount of *Faecalibacterium* spp*.* populations in early infancy indicates that the arrival of initial colonizers may facilitate its implantation, consuming of the available oxygen and creating an anaerobic environment favorable for the growth of obligatory anaerobic bacteria, such as *Faecalibacterium* spp*.* So, the implantation of *Faecalibacterium* species, an EOS bacteria, depends on the physicochemical conditions previously created by other commensal bacteria (Miquel et al., 2014).

III.3.3 Metabolism and growth requirements

The genus *Faecalibacterium* include rod shaped, non-spore-forming, non-motile and extremely oxygen sensitive (EOS) bacteria (Duncan et al., 2002; Miquel et al., 2013). Analysis of the *F. prausnitzii* membrane suggests that this bacterium either lacks cell wall lipopolysaccharides (LPS) or displays an unusual LPS composition (Miquel et al., 2014) and there is no consensus if they are Gram negative (Duncan et al., 2002) or Gram positive bacteria (Miquel et al., 2013).

F. prausnitzii isolates can grow well using simple carbohydrates like Glucose, Fructose, Cellobiose, Maltose and Galactose, but some differences exist

between strains in their capability to ferment more complex carbohydrates such as those that are either host or diet derived (Figure 2) (Fructo-oligosacharides, Pectin, Inulin; Duncan et al., 2002; Lopez-Siles et al., 2012). The major end products of glucose fermentation by *F. prausnitzii* strains are formate, tiny amounts of D-lactate and massive quantities of butyrate (Duncan et al., 2002). Most of the isolates tested by Lopez-Siles (2012) can grow on apple pectin and are able to use some pectin derivatives, suggesting that pectin or pectin derivatives could be used as a novel prebiotic approach to stimulate *F. prausnitzii* (Chung et al., 2016). Nine of the eleven strains analyzed by Duncan (2002) and Lopez-Siles (2012) are able to ferment inulin, but only two of them can grow well on this substrate (Table 1) (Lopez-Siles et al., 2017). This evidence corroborates the study of Ramirez-Farias (2009) that observed the stimulation of this species in nutritional interventions with this prebiotic, and suggests that only some members of *F. prausnitzii* population are selectively stimulated by inulin (Chung et al., 2016; Ramirez-Farias et al., 2009). None of the strains evaluated to date can exploit as single energy source any of arabinose, melibiose, raffinose, rhamnose, ribose, xylose, linear and α-1,2-branched dextrans, arabinogalactan, xylan, citrus pectin, or peptides (Louis et al., 2007; Martín et al., 2014b).

Table 1 - Substrates of different origin metabolized by *Faecalibacterium prausnitzii* isolates in vitro (batch pure cultures). Adapted from Lopez-Siles et al., 2017.

Substrate	No. Of utilizers	No. Of strains tested
Simple carbohydrates		
Glucose	11	11
Fructose	4	4
Cellobiose	10	11
Maltose	10	11

Galactose	9	10
Galacturonic acid	7	9
Sucrose	2	4
Melezitose	1	4
Trehalose	1	4
Rhamnose	1	11
Amino acids		
Arginine	4	4
Histidine arylamide	4	4
Glycine arylamide	2	4
Diet-derived		
Fructo-oligosacharides	4	4
Pectin(apple)	10	10
Inulin(chicory)	9	11
Host-derived		
Glucosamine HCl	10	10
N-acetylglucosamine	9	10
Glucuronic acid	6	10

These studies show that *F. prausnitzii* can switch between some substrates derived from the diet or the host. This capability could be investigated further to define novel strategies to use these substrates to reestablish *F. prausnitzii* populations in patients. Some studies already showed that prebiotic treatments can stimulate *Faecalibacterium* populations, but they couldn't differentiate which specie under specific prebiotic intakes grows better (Benus et al., 2010; Lopez-Siles et al., 2017; Ramirez-Farias et al., 2009).

III.3.4 Anti-inflammatory properties

Several anti-inflammatory properties are attributed to these species. Rossi et al. (2016) demonstrated that *F. prausnitzii* A2-165 has a strong capacity to induce IL-10 in human and murine dendritic cells and influence T cell differentiation *in vitro* and *in vivo*. The reduction in IFN-γ^+ cells *in vivo* and the IL-10 induction in the lamina propria cells of damaged intestinal tissue could be a mechanisms contributing to the suppressive effect on inflammation in mouse colitis models (Rossi et al., 2016). Other studies with *F. prausnitzii* cells or their culture supernatant reported the reduction of the severity of induced inflammation in different murine models (Martín et al., 2014b, 2015; Sokol et al., 2008b). These anti-inflammatory effects are associated with secreted metabolites capable of blocking NF-κB activation, IL-8 production (Sokol et al., 2008b) and upregulation of regulatory T cell production (Figure 3) (Qiu et al., 2013). One of these metabolites is a protein, called MAM, of 15 kDa, that have been identified in *F. prausnitzii* A2-165 cultures supernatant, and can inhibit the NF-κB pathway in several intestinal epithelial cells lines (Quévrain et al., 2014). *F. prausnitzii* supernatant has also been shown to attenuate the severity of inflammation through the release of metabolites that improve the intestinal barrier function and that affect paracellular permeability (Carlsson et al., 2013; Martín et al., 2015). Besides that, Wrzosek et al. (2013) (Wrzosek et al., 2013) have shown, using a gnotobiotic model, that *F. prausnitzii* could also influence gut physiology through mucus pathway and the production of the mucus O-glycans, and this

may help to maintain suitable proportions of different cell types of secretory linage in the intestinal epithelium. Another important anti-inflammatory property is the ability to produce butyrate. Butyrate is a short chain fatty acid (SCFA) with immune-modulatory properties, that have beneficial effects in the GIT (Duncan et al., 2002; Macfarlane and Macfarlane, 2011). *F. prausnitzii* has been consistently reported as one of the main butyrate producers found in the intestine (Barcenilla et al., 2000). The butyrate is the main energy source for the colonocytes and it has protective properties against colorectal cancer (CRC) and IBD (Archer et al., 1998; Christl et al., 1996). Butyrate can reduce intestinal mucosa inflammation through inhibiting NF-κB transcription factor activation, upregulating PPARγ and inhibiting interferon *gamma* (IFN-γ) (Klampfer et al., 2003; Lopez-Siles et al., 2017).

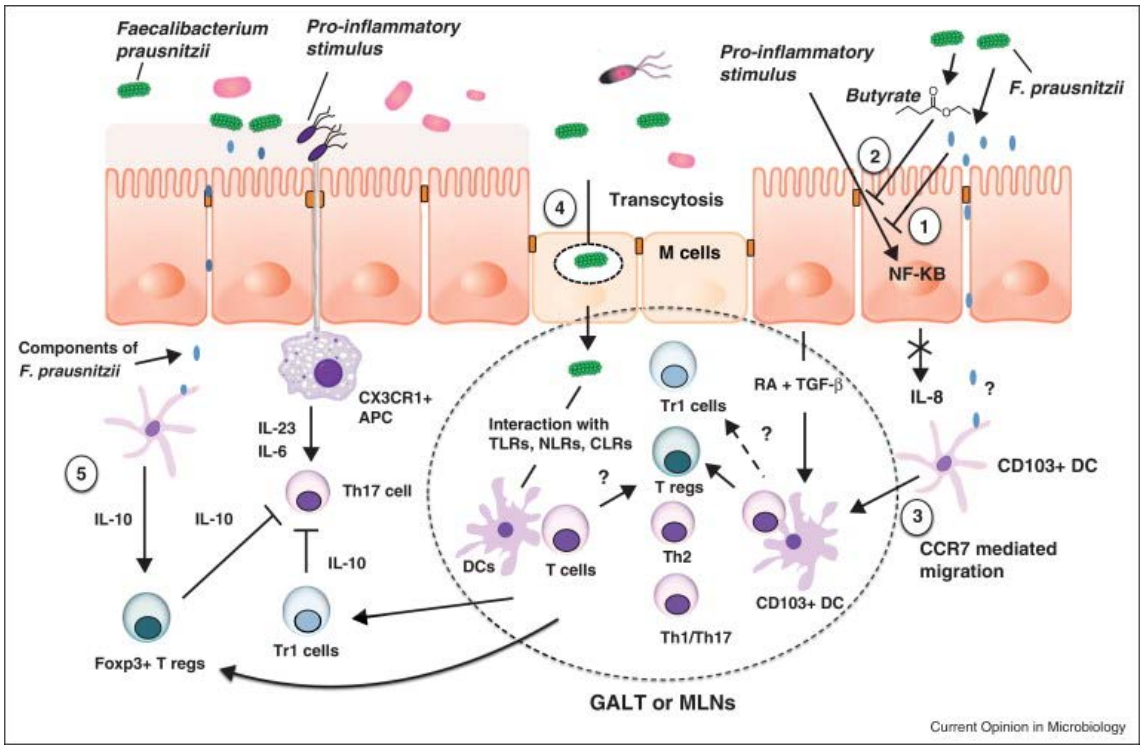

Figure 4 - Proposed anti-inflammatory mechanisms of *F. prausnitzii*. 1. The supernatant of *F. prausnitzii* blocks NF-kB activation induced by a pro-inflammatory stimulus [21__]. 2. Butyrate produced by *F. prausnitzii* inhibits NF-kB activation in mucosal biopsies. 3. *F. prausnitzii* components might interact with CD103+ dendritic cells (DCs) in the lamina propria and stimulate their migration to mesenteric lymph nodes (MLN) and the induction of Tregs. 4. M cell transcytosis of *F. prausnitzii* in organized lymphoid structures may induce Tregs. 5. The capacity of *F. prausnitzii* to induce high amounts of IL-10 in antigen presenting cells may enhance the suppressive activity of Foxp3+ Tregs and block Th17 cells induced by pro-inflammatory stimuli.(Miquel et al., 2013)

Furthermore, over the last few years many studies correlated *Faecalibacterium* spp. depletion with gut diseases, such as IBD, irritable bowel syndrome (IBS), colorectal cancer (CRC), obesity, and celiac disease (Balamurugan et al., 2008; Hansen et al., 2012; Miquel et al., 2013; Rajilić-Stojanović et al., 2011; Sokol et al., 2008b; Swidsinski et al., 2008). Altogether, these findings indicate that *Faecalibacterium* spp. is a functionally important member of the microbiota and has a crucial role in maintaining gut physiology and host health. Because of its important role in GIT homeostasis, *Faecalibacterium* spp.is considered today as a potential next generation probiotic (Martín et al., 2018; Miquel et al., 2015; Sokol et al., 2008b). Despite its relevance in the human gut ecosystem, little is known about the diversity of *Faecalibacterium* spp. and only a few studies have examined isolated strains and used functional approaches (Lopez-Siles et al., 2012, 2017). To better understand the biodiversity and beneficial effect of this species, it is essential to increase our knowledge of several cultured strains and their genomes.

III.3.5 Rules in health and diseases

F. prausnitzii abundance decay has been associated in the mucosa of IBD patients (Chassaing and Darfeuille–Michaud, 2011; Ott et al., 2008; SEKSIK et al., 2006; Sokol et al., 2008a; Tamboli et al., 2004) and other non-intestinal diseases like diabetes related to obesity (Hippe et al., 2016) and atopic dermatitis (Song et al., 2016).

IBD is a group of disorders characterized by a chronic and relapsing inflammation of the gastrointestinal tract (Loftus, 2004). The two most common forms of IBD are Crohn disease (CD) and ulcerative colitis UC. The pathogenesis involves an inappropriate and ongoing activation of the mucosal immune system driven by the presence of the intestinal microbiota in genetically predisposed patient (Sokol et al., 2008b). Fewer types of Firmicutes, mostly

from Ruminococcaceae, were observed in feces of CD patients (Lopez-Siles et al., 2017). Regarding *F. prausnitzii* population, subtype richness is also lower in IBD patients, which frequently tends to only possess one of the two main phylogroups (Figure 4).

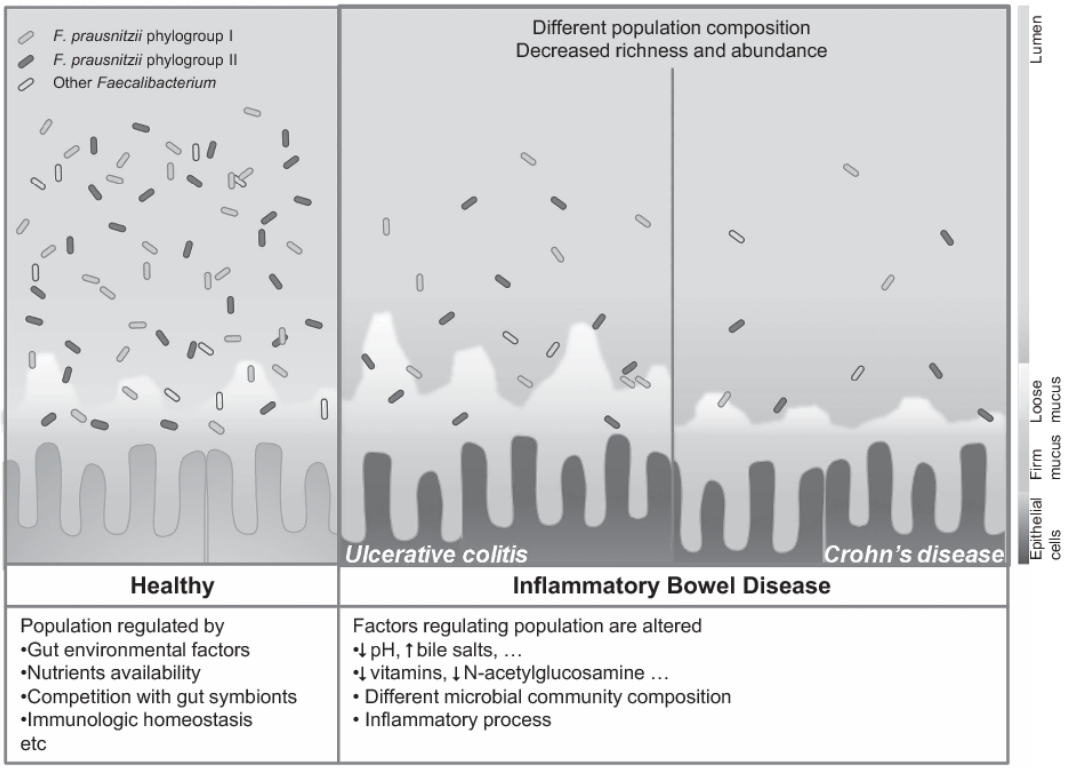

Figure 5 - *F. prausnitzii* populations in healthy gut and in patients with IBD. In IBD patients, alteration of gut environment may affect *F. prausnitzii* population composition and load. These differences can be monitored to discriminate within IBD subtypes (Lopez-Siles et al., 2017).

Factors explaining these differences remain unknown and for that, further studies of isolation and characterization of strains from patients with different intestinal disorders are needed to test the effect of either host or gut physicochemical factors on different *Faecalibacterium* strains (Lopez-Siles et al., 2016). On the one hand an external factor can cause a depletion of *F. prausnitzii*, but also this species reduction can be a contributing factor to disease aggravation. The restoration of normal counts of this species should be explored to determinate its rule on healing and/ or attenuation of disease progression (Lopez-Siles et al., 2017). The effects of IBD treatments on *F. prausnitzii* population levels are still unclear, but some show a positive impact

on *F. prausnitzii* population in the microbiota (Miquel et al., 2013). Several studies have shown its beneficial effects in murine models of IBD and IBS (Carlsson et al., 2013; Martín et al., 2014a; Miquel et al., 2016; Rossi et al., 2016; Sokol et al., 2008b). Some studies showed that different treatments can increase the level of *Faecalibacterium* spp: Rifaximin (Maccaferri et al., 2010), chemotherapy and interferon a-2b (Dörffel et al., 2012), high-dose cortisol therapy or Infliximab (Swidsinski et al., 2008).

III.3.6 Potential use of *F. prausnitzii* as a next generation probiotic

Probiotics are “live microorganisms which when administered in adequate amounts confer a health benefit on the host” (FAO/WHO, 2001). To consider a strain as probiotic, it should be: (i) well-characterized, (ii) achieve safety requirements, and (iii) confer beneficial effects on the host. Nevertheless, a careful selection should be made in each case as probiotic properties are usually strain specific and cannot be extrapolated to another strain even belonging to the same species (Gareau et al., 2010). Most of the probiotics sold and used, both in probiotic research and commercial probiotic development, are from *Lactobacillus* spp. and *Bifidobacterium* spp. genera. The more commonly exploited strains/species among the lactobacilli and bifidobacteria have been accepted as having Generally Regarded as Safe (GRAS) status in the United States (https://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/NoticeInventory/default.htm), or have been granted Qualified Presumption of Safety status by the European Food Safety Authority (EFSA; Ricci et al., 2018). With the development of better culturing methodologies, with the drop in the cost of genome and metagenome sequencing, and more powerful tools to edit and modify bacterial genomes, we are now facing a new era in probiotic research. The knowledge of the composition and function of the human gut microbiome, also accelerated by massively parallel sequencing, has dramatically extended

the range of organisms with potential health benefits and these organisms are referred as Next-Generation Probiotics (NGPs) O'Toole et al., 2017).

Because of its benefic effects in GIT (Figure 4), *F. prausnitzii* is considered today as a potential NGPs (Martín et al., 2018). *F. prausnitzii* is one of the most abundant species to be found in the large intestine and it has been reported to be depleted in individuals with IBD. So, it seems reasonable that if there was a causal link between disease status and the absence of this organism, then by simply feeding it to the individual its health promoting features should be restored and thus it may be considered as an NGP (O'Toole et al., 2017; Sokol et al., 2008b). However, there is a lack of studies demonstrating the efficacy of this organism as a probiotic to be able to reverse the symptoms of inflammatory bowel disease when fed to humans. In animal models, evidence is available and feeding animals with *F. prausnitzii* does lead to or associate with induction of anti-inflammatory cytokines or reduction of pro-inflammatory cytokines in chemically-induced models of colitis/IBD (Martín et al., 2015, 2018; Rossi et al., 2016). There are many studies supporting the beneficial effects of *Faecalibacterium* spp in animal gut, but also there are some lack of knowledge in its biology and phylogeny. For a long time, it was considered the existence of only one specie in this genus, but recent studies showed the possibility of different species and this fact raises the need to carry out new studies to better characterize their beneficial effects, their crosstalk with the host and the possible effectors underlying these effects (Martín et al., 2018). One good strategy to realize that kind of studies is to use comparative genomics approaches with many different *F. prausnitzii* strain genomes (Martín et al., 2017).

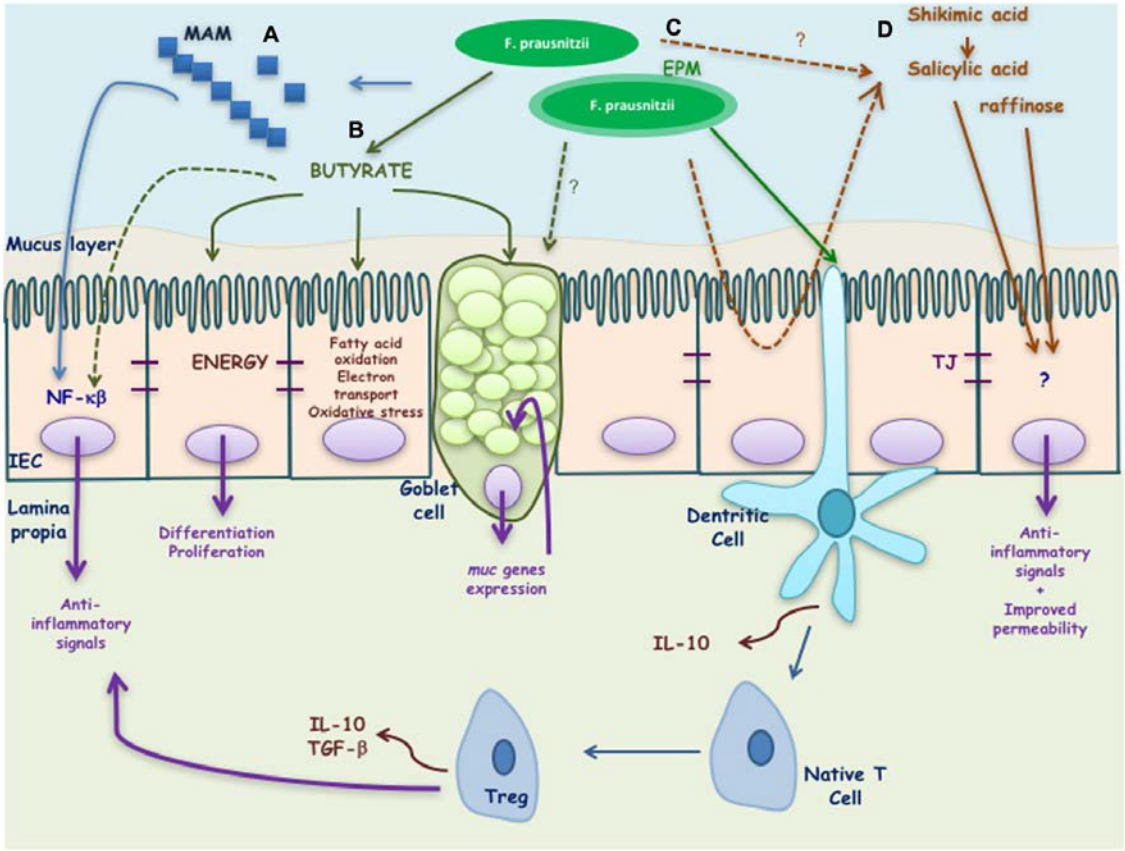

Figure 6 - Putative effectors of Faecalibacterium prausnitzii and its effects on the host. *F. prausnitzii* exerts its benefic effects by means of different effectors: (A) MAM peptides secreted by *F. prausnitzii* block NF-κB activation induced by a pro-inflammatory stimulus. (B) Butyrate produced by *F. prausnitzii* inhibits NF-κB activation and interacts with the intestinal epithelial cells (IEC) driving to the activation of different genes involved on the differentiation, proliferation, and restitution of enterocytes. It is also involved on the regulation of fatty acid oxidation, electrons transport chain, oxidative stress, and apoptosis. In goblet cells it has been described to stimulate muc genes allowing a high production of mucus. (C) EPM produced by *F. prausnitzii* modulates IL-10 cytokine production in antigen presenting cells. Finally, (D) salicylic and shikimic acids are anti-inflammatory molecules able to block inflammation induced by a pro-inflammatory stimulus while raffinose is key in maintaining gut permeability (Martín et al., 2018).

III.4 Comparative genomics

III.4.1 History

Since the development of the Sanger sequencing methodology, several genome projects have been created on the GOLD database (Mukherjee et al., 2017) and, after the advent of next-generation sequencing technologies in

2004, the sequencing cost and dispended time *per* genome sequencing has constantly dropped (Figure 6).

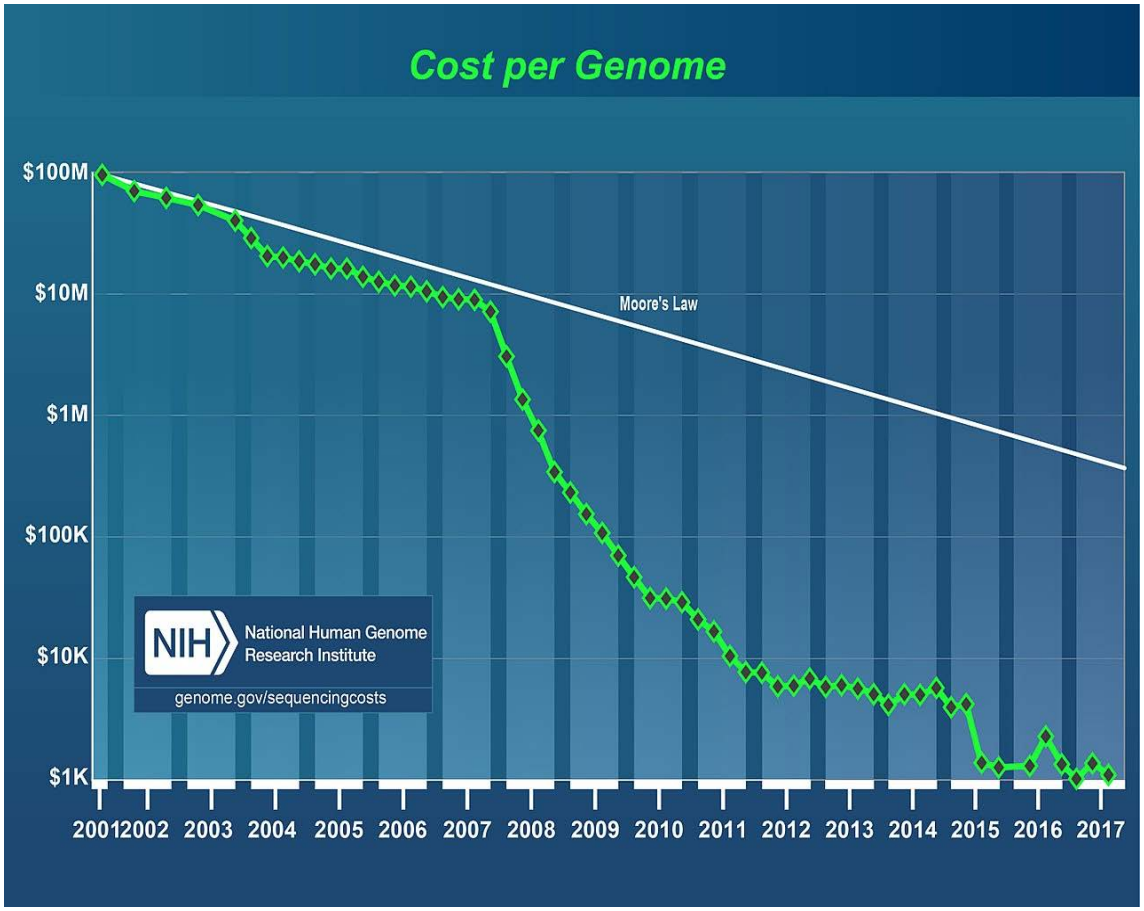

Figure 7 – Cost per genome sequencing (https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/).

As a result of this, the number of genome sequencing projects has increased exponentially. This scenario is also true for prokaryotes, where approximately 85000 genome projects are registered on GOLD database nowadays (Figure 7) (https://gold.jgi.doe.gov/). The microbial genomics may be used for drug resistance research, functional gene study, molecular markers development, metabolic pathway research, determining the origin and evolution of pathogen and also for screening of NGP (Baugher and Klaenhammer, 2011).

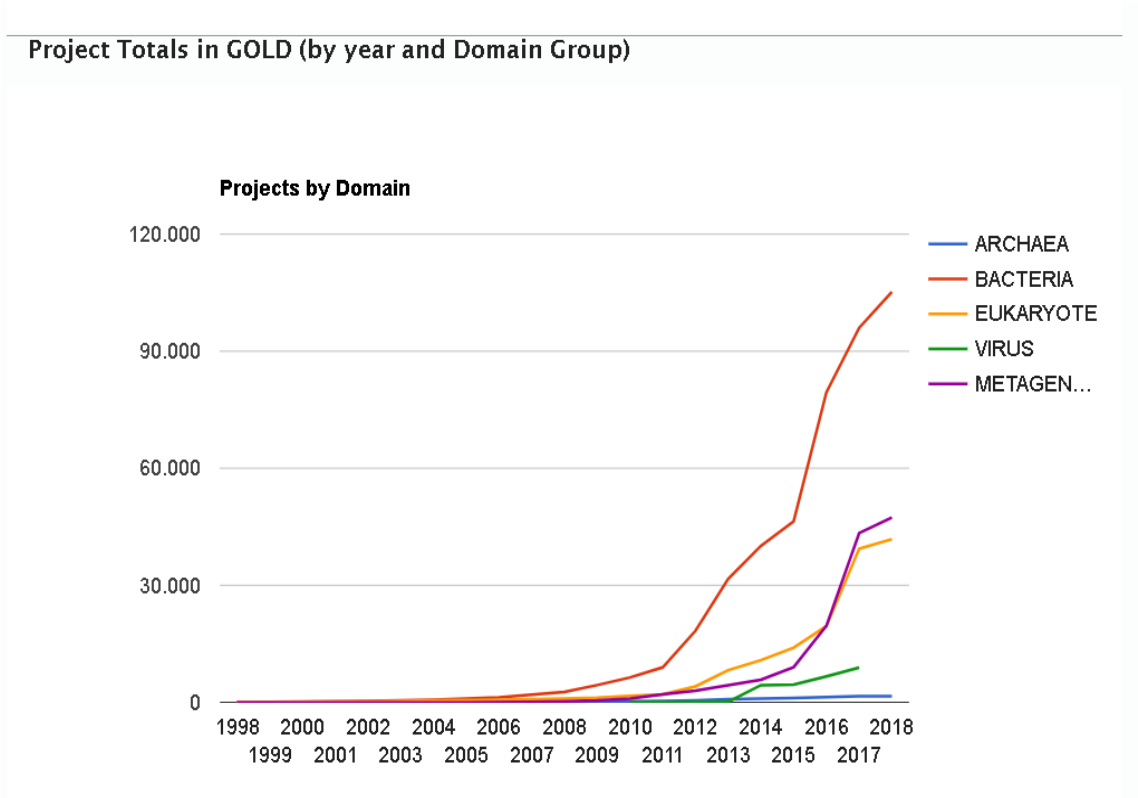

Figure 8 – Project totals in GOLD (by year and domain group). (https://gold.jgi.doe.gov/statistics).

Comparative genomic is the direct comparison of whole genomes from different organisms including the comparison of gene number, gene location, and gene content from these genomes. This approach helps to understand the extent of gene conservation among genomes, providing insights into the mechanism of genome evolution and gene transfer among genomes. It helps to reveal the pattern of acquisition of foreign genes through lateral gene transfer and also helps to determine the core genome of different strains or species, which commonly correspond to the genes that are essential for survival (Xiong, 2006).

III.4.2 Phylogeny

At the beginning of the 18th century, Linnaeus proposed the first hierarchical classification system of the species, based on the grouping of organisms

according to their phenotypic characteristics (Chan et al., 2012). The prokaryotic systematics began as an intuitive science and became increasingly objective using data derived from advances in other areas and is now, once again, undergoing a transitional period. For over a century, microbial systematics have relied on phenotypic information derived from pure culture growth and time-consuming techniques that are difficult to standardize (Goodfellow et al., 2014). The molecular revolution that has transformed all of biology has had as much impact on taxonomy and systematics of bacteria as any other area of biology, and on top of that, rapid advances in the sequencing of whole genomes have provided the platform for a paradigm shift to the systematic community (Emerson et al., 2008). In the early 1970s, approaches based on phenotypic traits began to give way to genotypic approaches, such as DNA-DNA hybridization (DDH), which is still the gold standard in bacterial taxonomy. However, few laboratories now perform DNA-DNA hybridization assays, since they are expensive and technically demanding when compared to a fast and easy sequencing of small signal sequences, such as the 16S ribosomal RNA gene (Chan et al., 2012). Nevertheless, 16S rRNA gene sequences tend to provide insufficient resolution to distinguish between closely related species. These limitations led to the need to develop techniques based on the use of complete genomes in the prokaryotic system (Sangal et al., 2016). We can mention, among these new techniques, the alignment of fragmented genomes for inference of phylogenetic distances, the whole-genome Multi-locus sequencing type (wgMLST) and the Average Nucleotide Identity (ANI). It is critical to understand that each method has a limited resolution covering part, but not all levels of taxonomic information (Figure 7) (Coenye et al., 2005). Recently, large-scale studies have integrated two or more methods based on genome content and chromosome organization (Zhi et al., 2012). These new techniques allow a greater accuracy in the identification of new species (Benevides et al., 2017; Chan et al., 2012; Emerson et al., 2008; Sangal et al., 2016).

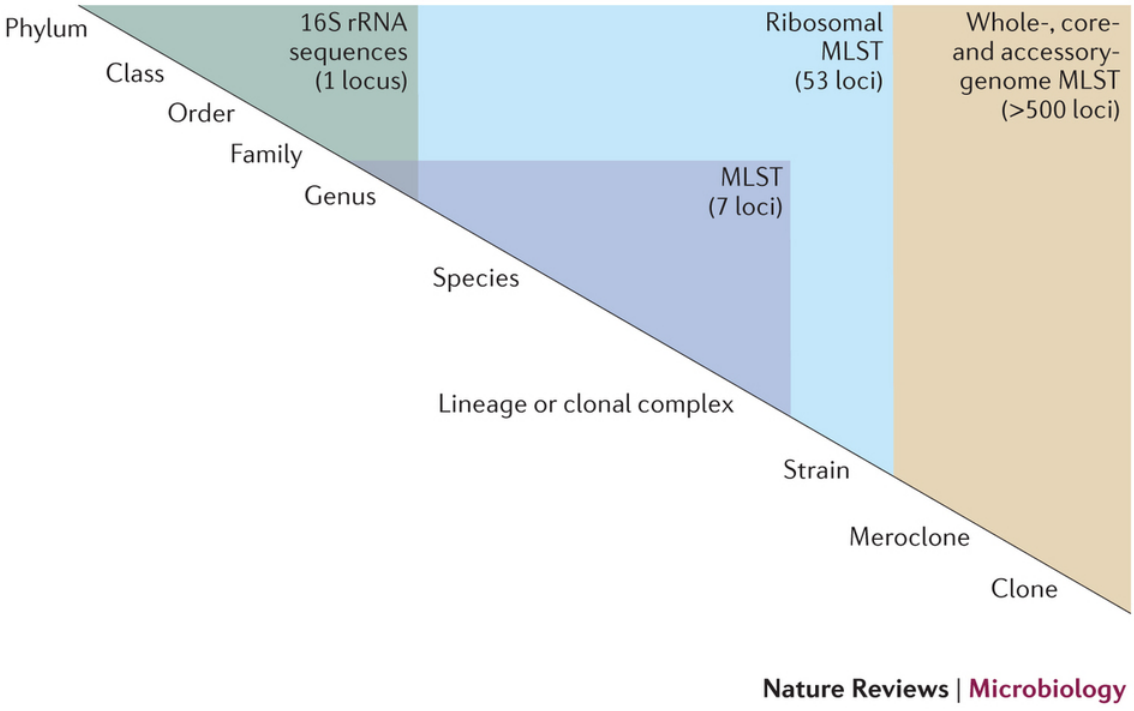

Figure 9 - Relating sequence data to nomenclature schemes. Hierarchical nomenclature schemes are artificial constructs that are developed to facilitate communication and are subject to change as new information becomes available. Nomenclature schemes are dependent on various factors, including sequence relationships, and are ideally genealogically based. The challenge is to map whole-genome sequencing (WGS) data to nomenclature schemes transparently but flexibly at a range of resolutions. The highest discrimination is required for studies of bacterial isolates from one patient or from very closely related transmission chains; these isolates can be thought of as having undergone microevolution. Such studies will require comparisons of whole genomes using whole-genome multi- locus sequence typing (MLST). Progressively lower resolution is required for studies of isolates with more distant common ancestors and, therefore, with more genetic differences. These relationships are best studied using the core genome common to the set of isolates of interest. Genes encoding ribosomal proteins are a particularly useful subset of core genes, and ribosomal MLST accommodates many levels of genealogical relationships, from clonal complexes and lineages to species, genera and beyond. In a database such as the Bacterial Isolate Genome Sequence Database (BIGSdb), multiple gene-by-gene schemes can be implemented alongside other, more conventional sequence- based schemes. Particular genotype summaries of genes or collections of genes can be associated with particular nomenclature schemes, enabling the database to deliver a plain-language report to the user (Maiden et al., 2013a).

The wgMLST is an expansion of the classical MLST technique, however, instead of using pre-defined schemes that use between 7 to 10 housekeeping genes, all genes from all genomes are compared and those that are common to all are used as different *loci* in the evolutionary analyzes (Maiden et al., 2013b). The ANI is a recent technique that emerged as a probable candidate for gold standard in the identification of bacteria, replacing DNA-DNA hybridization. It is now known that an ANI equal to or greater than 95% is equivalent to 70% of

DDH (Emerson et al., 2008; Kim et al., 2014; Konstantinidis et al., 2006). In addition to these techniques that use multiple genes, there are also software, such as the Gegenees (Agren et al., 2012), which perform fragmentation of genomes and compare all against all fragments to create a similarity matrix which can be used in phylogenetic analyzes. Another method that can be used for studies of phylogeny and genomic diversity are Pangenomic analyzes. For this, the subsets of the pangenoma are determined and the development of these subsets for different datasets is calculated. After calculations, the values of *alpha* and *theta* tangent are compared and used for species determination (Benevides et al., 2017; Figueiredo et al., 2015; Tettelin et al., 2008).

III.4.3 Genomic plasticity

Living organisms are defined by the genes they possess and over time, it is possible to observe changes in the content and order of these genetic information. Control of expression of this gene set determines whether an organism can survive changing conditions and can compete for the resources it needs to reproduce. The evolutionary pressure acts on these mechanisms, leading the bacteria to undergo genomic changes. This process of changes in DNA is called genomic plasticity (Bennett, 2004; Soares et al., 2012). This plasticity is a consequence of punctual mutations, genome rearrangements and also of exogenous gene acquisition (Schmidt and Hensel, 2004). This DNA transference between bacterial cells can occur by one of three mechanisms: conjugation, transformation, and transduction (Bennett, 2004).

The transduction mechanism involves DNA transfer mediated by a phage, and this process gains special attention in the case of some bacteria that are used in fermented products for human consumption (Arber, 2014). Analyzes of bacterial genome sequences, generated by metagenomics-based investigations, has demonstrated that bacteriophages are one of the major forces responsible for shaping the diversity and composition of the gut bacterial community (Ventura et al., 2011). Phages are obligate parasites and most of

them have a cycle of multiplication through cell lysis, where hundreds of viral particles are released, ready to infect nearby cells (Garneau and Moineau, 2011). Not only these virulent phages but also temperate ones have been suspected to play a role in dysbiosis of intestinal microbiota (Manrique et al., 2017).

A significant part of the horizontal gene transfer is or has been facilitated by genomic islands (GEIs). These islands can be classified as pathogenicity islands (PAI), metabolic islands (MI), symbiotic islands (SI) and antibiotic resistance islands (RI) and are identified as regions of transfer and incorporation of a significant set of genes into blocks with potential to bring evolutionary changes (Soares et al., 2012). Bioinformatics studies have shown that GEIs tend to carry more 'novel' genes than the rest of the genome, suggesting that they are strongly selected for adaptive and auxiliary functions. They are integrative elements that can be mobilized by conjugation when they have compatible origins of transfer or by integrating in conjugative elements (Guglielmini et al., 2011).

The acquisition and loss of genes by all these mechanisms reflect on the physiological life style and physiological versatility of the microorganism and the results obtained from comparative genomics demonstrate its importance for bacterial evolution (Dobrindt and Hacker, 2001).

III.4.4 Probiotics safety assessment through *in silico* analysis

Probiotic research faces new challenges and comparative genomics could accelerate research into new probiotic strains. In recent years, genome sequencing of gut commensals and symbionts has come to the fore and the development of next generation sequencing and new tools for comparative genomics analysis helps to provide insights into the diversity and evolution of commensal and probiotic bacteria and to reveal the molecular basis for their health-promoting activities (Ventura et al., 2009). These omics approaches

allow the simultaneous analysis of substantial numbers of genes and proteins, permitting to quickly detect and eliminate strains that pose a potential risk, verifying the absence of genes related to virulence or antibiotic resistance transferability and the presence of genes involved in health-promotion (Abriouel et al., 2017; Papadimitriou et al., 2015).

The European Food Safety Authority proposed a pre-market safety assessment of selected groups of microorganisms leading to a Qualified Presumption of Safety (QPS). This proposal suggests that a safety assessment of a defined taxonomic group (e.g. genus or group of related species) could be made based on four pillars (establishing identity, body of knowledge, possible pathogenicity and end use). If the taxonomic group did not raise safety concerns or, if safety concerns existed, but could be defined and excluded, the grouping could be granted QPS status. QPS group would be freed from the need for further safety assessment while microorganisms not considered suitable for QPS would remain subject to a full safety assessment (J.B. Prajapati and Suja Senan, 2013).

IV Goals

IV.1 Main goal

The main goal of this thesis was to characterize *Faecalibacterium* spp*.* through comparative genomics analyses to understand its phylogeny and safety aspects for use as probiotics.

IV.2 Specific goals

The specific goals of this thesis were:

- To sequencing, assembly and annotate ten new genomes of *Faecalibacterium* spp.;
- To perform phylogenetics analyzes of the genus *Faecalibacterium* considering different approaches: 16S rRNA sequences, whole-genome Multi-Locus Sequencing Type (WgMLST), Pangenome and Average Nucleotide Identity (ANI);
- To perform analyzes of genome synteny and genome plasticity including: prediction of genomic islands and phage regions;
- To search genes that encodes bacteriocins and antibiotic resistance;
- To predict the metabolic pathways.

V Research Articles

V.1 Chapter I. New Insights into the Diversity of the Genus *Faecalibacterium*

Leandro Benevides, Sriti Burman, Rebeca Martin, Véronique Robert , Muriel Thomas , Sylvie Miquel, Florian Chain, Harry Sokol, Luis G. Bermudez-Humaran, Mark Morrison, Philippe Langella, Vasco A. Azevedo, Jean-Marc Chatel and Siomar Soares

This article was submitted to Microbial Immunology, a section of the journal Frontiers in Microbiology and published on 22 September 2017.

Here we present a new phylogenetic and comparative study of five sequenced genomes of *Faecalibacterium* spp. available in public databases, combined with twelve new genome sequences isolated from healthy volunteers in Europe and Australia. The phylogenetic relationships among these isolates of *Faecalibacterium* spp. were compared, and pangenomic analysis provided us a more global view of the genomic diversity across these strains. This study was performed with the objective of realize a deep phylogenetic evaluation. Since the first isolation of this bacteria its phylogenetic relation was reviewed sometimes and until now it is not very well established if there are one or more species into the genus *Faecalibacterium*. This phylogenetic study is the first step in the understand about the probiotic properties and will be used in the future researches about the diversity and host-interaction.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 September 2017
doi: 10.3389/fmicb.2017.01790

New Insights into the Diversity of the Genus *Faecalibacterium*

Leandro Benevides[1,2], Sriti Burman[3], Rebeca Martin[2], Véronique Robert[2], Muriel Thomas[2], Sylvie Miquel[2,4], Florian Chain[2], Harry Sokol[2,5,6], Luis G. Bermudez-Humaran[2], Mark Morrison[3], Philippe Langella[2], Vasco A. Azevedo[1], Jean-Marc Chatel[2†] and Siomar Soares[7†]*

[1] Department of General Biology, Federal University of Minas Gerais, Belo Horizonte, Brazil, [2] Commensals and Probiotics-Host Interactions Laboratory, Micalis Institute, Institut National de la Recherche Agronomique, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, [3] Faculty of Medicine, Translational Research Institute, University of Queensland Diamantina Institute, University of Queensland, St. Lucia, QLD, Australia, [4] UMR 6023 Laboratoire Microorganismes: Génome et Environnement, Centre National de la Recherche Scientifique, Université Clermont Auvergne, Clermont-Ferrand, France, [5] Faculté de Médecine Saint-Antoine, Institut National de la Santé et de la Recherche Médicale U1157/UMR7203, AVENIR Team Gut Microbiota and Immunity Equipe de Recherche Labélisée, Université Pierre et Marie Curie, Paris, France, [6] Service de Gastroentérologie, Hôpital Saint-Antoine, Assistance Publique—Hôpitaux de Paris, Paris, France, [7] Department of Immunology, Microbiology and Parasitology, Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, Brazil

OPEN ACCESS

Edited by:
Geanncarlo Lugo-Villarino, UMR5089 Institut de Pharmacologie et de Biologie Structurale (IPBS), France

Reviewed by:
Clara G. De Los Reyes-Gavilan, Consejo Superior de Investigaciones Científicas (CSIC), Spain
Ulisses Padua Pereira, Universidade Estadual de Londrina, Brazil

***Correspondence:**
Siomar Soares
siomars@gmail.com

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Immunology, a section of the journal Frontiers in Microbiology

Received: *03 August 2017*
Accepted: *05 September 2017*
Published: *22 September 2017*

Citation:
Benevides L, Burman S, Martin R, Robert V, Thomas M, Miquel S, Chain F, Sokol H, Bermudez-Humaran LG, Morrison M, Langella P, Azevedo VA, Chatel J-M and Soares S (2017) New Insights into the Diversity of the Genus Faecalibacterium. Front. Microbiol. 8:1790. doi: 10.3389/fmicb.2017.01790

Faecalibacterium prausnitzii is a commensal bacterium, ubiquitous in the gastrointestinal tracts of animals and humans. This species is a functionally important member of the microbiota and studies suggest it has an impact on the physiology and health of the host. *F. prausnitzii* is the only identified species in the genus *Faecalibacterium*, but a recent study clustered strains of this species in two different phylogroups. Here, we propose the existence of distinct species in this genus through the use of comparative genomics. Briefly, we performed analyses of 16S rRNA gene phylogeny, phylogenomics, whole genome Multi-Locus Sequence Typing (wgMLST), Average Nucleotide Identity (ANI), gene synteny, and pangenome to better elucidate the phylogenetic relationships among strains of *Faecalibacterium*. For this, we used 12 newly sequenced, assembled, and curated genomes of *F. prausnitzii*, which were isolated from feces of healthy volunteers from France and Australia, and combined these with published data from 5 strains downloaded from public databases. The phylogenetic analysis of the 16S rRNA sequences, together with the wgMLST profiles and a phylogenomic tree based on comparisons of genome similarity, all supported the clustering of *Faecalibacterium* strains in different genospecies. Additionally, the global analysis of gene synteny among all strains showed a highly fragmented profile, whereas the intra-cluster analyses revealed larger and more conserved collinear blocks. Finally, ANI analysis substantiated the presence of three distinct clusters—A, B, and C—composed of five, four, and four strains, respectively. The pangenome analysis of each cluster corroborated the classification of these clusters into three distinct species, each containing less variability than that found within the global pangenome of all strains. Here, we propose that comparison of pangenome subsets and their associated α values may be used as an alternative approach, together with ANI, in the *in silico* classification of new species. Altogether,

our results provide evidence not only for the reconsideration of the phylogenetic and genomic relatedness among strains currently assigned to *F. prausnitzii,* but also the need for lineage (strain-based) differentiation of this taxon to better define how specific members might be associated with positive or negative host interactions.

Keywords: *Faecalibacterium prausnitzii*, genome sequencing, 16S rRNA gene phylogeny, phylogenomic analysis, Average Nucleotide Identity, gene synteny, pangenome, new species

INTRODUCTION

Members of genus *Faecalibacterium* are commensal bacteria, ubiquitous in the gastrointestinal tracts of animals and humans. Within the human colon, this taxon is the main member of the *Clostridium leptum* cluster, and comprises the second-most common representative in fecal samples, after *Clostridium coccoides* (Tap et al., 2009; Walker et al., 2011). The first characterized isolates were classified as *Fusobacterium prausnitzii*, but its close relationship with other members of the *C. leptum* cluster (phylum Firmicutes, class Clostridia, family Ruminococcaceae) was later established through analysis of the 16S rRNA gene of different strains found in humans (ATCC 27766 and ATCC 27768) (Wang et al., 1996; Duncan et al., 2002). The relative abundance of *F. prausnitzii* in vertebrate animals other than humans, such as pigs (Castillo et al., 2007), mice (Nava and Stappenbeck, 2011), calves (Oikonomou et al., 2013), and chickens (Scupham, 2007), suggests that the species is a functionally important member of the microbiota and likely has an impact on the physiology and health of the host. In that context, changes in the abundance of *F. prausnitzii* have been widely described in various intestinal and metabolic diseases in humans, such as colorectal cancer (CRC), Crohn's disease (CD), and ulcerative colitis (UC) (Sokol et al., 2008; Rajilić-Stojanović et al., 2011; Miquel et al., 2013). Due to its ubiquity and immunomodulatory properties, some studies suggest that *F. prausnitzii* is an indicator of, and an active contributor to, intestinal health and the maintenance of gut homeostasis (Sokol et al., 2008; Miquel et al., 2013, 2014). Despite its relevance in the human gut ecosystem, little is known about the diversity of *F. prausnitzii* (Miquel et al., 2014) and only a few studies have examined isolated strains and used functional approaches (Duncan et al., 2002; Lopez-Siles et al., 2012). To better understand the biodiversity and beneficial effect of this species, it is essential to increase our knowledge of several cultured strains and their genomes.

Recent studies, based on 16S rRNA sequence analyses, have suggested the existence of two phylogroups within this species (Duncan et al., 2002; Lopez-Siles et al., 2012, 2017). Here we present a new phylogenetic and comparative study of five sequenced genomes of *F. prausnitzii* available in public databases, combined with twelve new genome sequences isolated from healthy volunteers in Europe and Australia. The phylogenetic relationships among these isolates of *F. prausnitzii* were compared, and pangenomic analyses provided us a more global view of the genomic diversity across these strains. These data will enable new insights into the contributions of genus *Faecalibacterium* to gut function.

METHODS

Genome Sequencing, Assembly, and Annotation

The genomes used in this study are presented in **Table 1**. The genome data of five different *F. prausnitzii* strains were retrieved from the PATRIC public database. These were combined with genome data from ten newly isolated *F. prausnitzii* strains recovered from stool samples of healthy European volunteers (Martín et al., 2017), as well as two newly isolated *F. prausnitzii* strains recovered from stool samples of healthy Australian subjects (following the guidelines of the University of Queensland Human Research Ethics Committee #2015000775). In Europe, the ten new genomes were sequenced by GATC Biotech Company using the Illumina HiSeq2500 platform; the genomes from the Australian isolates were sequenced using the Illumina NextSeq platform at the Australian Centre for Ecogenomics (www.ecogenomic.org). The genome of the wild-type strain *F. prausnitzii* A2-165 (*F. prausnitzii*_A2-165_PacBio) was sequenced using PacBio single-molecule real-time (SMRT) technology on an RS system (Pacific Bioscience) and assembled by the GATC Biotech Company. The quality of the sequenced reads was checked with FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). All of the genomes (except for wild-type strain A2-165) were assembled using a *de novo* strategy with SPAdes software, v3.8.0. The quality of the assemblies was evaluated using QUAST software (Gurevich et al., 2013) and all genomes were subjected to automated functional annotation using the RAST server (Aziz et al., 2008).

Phylogeny

Phylogenetic analyses were performed on 16S rRNA sequences, whole genome sequences, and the results of whole-genome Multi-Locus Sequence Typing (wgMLST). For the first analysis, the 16S rRNA sequence from the genome of *F. prausnitzii*_A2-165_PacBio was used to perform a BLASTn search in the NCBI database for all sequences belonging to the genus *Faecalibacterium*. The sequence results with more than 82% coverage and 92% identity were collected, and the 16S rRNA gene sequence from *Subdoligranulum variabile* BI-114 was included as an outgroup. Sequences were then aligned using the multiple sequence alignment tool CLUSTALW (Thompson et al., 1994) integrated in MEGA7 software (Kumar et al., 2016). After that, the most appropriate evolutionary model was defined and evolutionary history was inferred using the maximum-likelihood (ML) criterion, based on the Kimura 2-parameter model (Kimura, 1980), with 1,000 bootstrap replicates.

TABLE 1 | Genomic features of *F. prausnitzii* genomes.

Genome name	PATRIC genome ID	Genbank accessions	Isolation country	Sequences	Genome length	GC content	PATRIC CDS
Faecalibacterium prausnitzii M21/2	411485.10	ABED00000000	United Kingdom	25	3,125,761	56.3	2,776
Faecalibacterium prausnitzii SL3/3	657322.3	FP929046	United Kingdom	1	3,214,418	54.81	3,052
Faecalibacterium prausnitzii L2-6	718252.3	FP929045	United Kingdom	1	3,321,367	55.57	3,232
Faecalibacterium cF. prausnitzii KLE1255	748224.3	AECU00000000	USA	139	2,907,000	56.27	2,783
Faecalibacterium prausnitzii AHMP_21	853.123	NOUV00000000	Australia	85	3,019,317	57.36	3,201
Faecalibacterium prausnitzii HMI_19	853.124	NOUW00000000	Australia	63	2,879,169	56.82	2,933
Faecalibacterium prausnitzii CNCM_4540	853.62	NMTQ00000000	France	48	3,043,568	55.7	3,206
Faecalibacterium prausnitzii CNCM_4541	853.63	NMTR00000000	France	78	2,822,838	58.11	2,825
Faecalibacterium prausnitzii CNCM_4542	853.64	NMTS00000000	France	106	2,914,466	55.83	3,071
Faecalibacterium prausnitzii CNCM_4543	853.65	NMTT00000000	France	22	3,080,452	56.2	3,223
Faecalibacterium prausnitzii CNCM_4544	853.66	NMTU00000000	France	71	2,808,526	55.98	2,907
Faecalibacterium prausnitzii CNCM_4546	853.67	NMTV00000000	France	244	3,422,520	54.88	3,611
Faecalibacterium prausnitzii CNCM_4573	853.68	NMTW00000000	France	83	3,275,249	55.9	3,479
Faecalibacterium prausnitzii CNCM_4574	853.69	NMTX00000000	France	38	3,088,985	56.26	3,249
Faecalibacterium prausnitzii CNCM_4575	853.70	NMTY00000000	France	37	3,006,602	57.51	3,077
Faecalibacterium prausnitzii CNCM_4644	853.71	NMTZ00000000	France	36	2,915,240	56.37	3,019
Faecalibacterium prausnitzii A2-165_PacBio	853.73	CP022479	France	1	3,110,044	56.33	3,231

The phylogenomic analysis was performed using Gegenees software (Agren et al., 2012), which calculated the percentage of similarity among the genomes of all strains. Before calculating similarity scores, we used the BLASTn alignment method, with a sequence fragmentation length of 200 bp and a step size of 100 bp. The input files for Gegenees contained the complete genomes in ".fna" format and the resulting similarity matrix was exported in ".nexus" format for phylogenomic analysis using SplitsTree4 software (Huson, 2005). The equal angle method was used to construct the phylogenetic network, which was plotted with NeighborNet.

A wgMLST analysis was performed using the *Build_wgMLSTtree* module in the PGAdb-builder web service tool (Liu et al., 2016). The 17 genome sequences were compared with the resulting PGAdb profile using BLASTn, with filters of 80% coverage and 90% identity.

Average Nucleotide Identity

We also performed an Average Nucleotide Identity (ANI) analysis using the whole-genome sequences. ANI represents a mean of identity/similarity values between homologous genomic regions shared by two genomes. It is generally accepted that ANI values of 95–96% equate to a DNA–DNA hybridization (DDH) value of 70%, and can be used as a threshold for species delineation (Konstantinidis and Tiedje, 2005; Kim et al., 2014).

Gene Synteny Analysis

The progressiveMauve option from the Mauve package (Darling et al., 2004) was used with default parameters to perform orthology comparisons and to evaluate gene synteny among the genomes of *F. prausnitzii*. This genome comparison method also predicts syntenic blocks, which reveal the rearrangement events among the genomes (Darling et al., 2004). This analysis was performed using four different datasets: first, using all 17 genomes, and then using the genome subsets of each of the three groups that resulted from ANI analysis.

Pangenome Calculation

The software program OrthoMCL (Li et al., 2003) was used first to define the cluster of orthologous genes and then, the commonly shared and species-specific genes of all the strains and subgroups. The amino-acid sequences from all coding DNA sequences (CDSs) in each genome were first used in an all-vs.-all BLASTp analysis with an *e*-value of 10^{-6}; the sequences were then clustered using the MCL algorithm. The CDSs observed in all strains were considered to comprise the core genome, while the CDSs harbored by only one strain were considered to be singleton genes.

To calculate pangenome development, we applied Heap's Law, with the formula $n = k^*N^{-\alpha}$, where n is the expected number of genes for a given number of genomes, N is the number of genomes, and the other terms are constants defined to fit the specific curve. According to Heap's law, a value of $\alpha \leq 1$ is representative of an open pangenome; this means that each added genome will contribute some new genes and the pangenome will increase. Instead, an α value >1 represents a closed pangenome, in which the addition of new genomes will not significantly affect the size of the pangenome. The extrapolations of the curves of the core genome and singletons were calculated using the least-squares fit of the exponential regression decay of the mean values, as represented by the formula $n = k^*\exp[-x/t]+tg(\theta)$, where n is the expected subset of genes for x number of genomes, exp is Euler's number, and the other terms are constants defined to fit

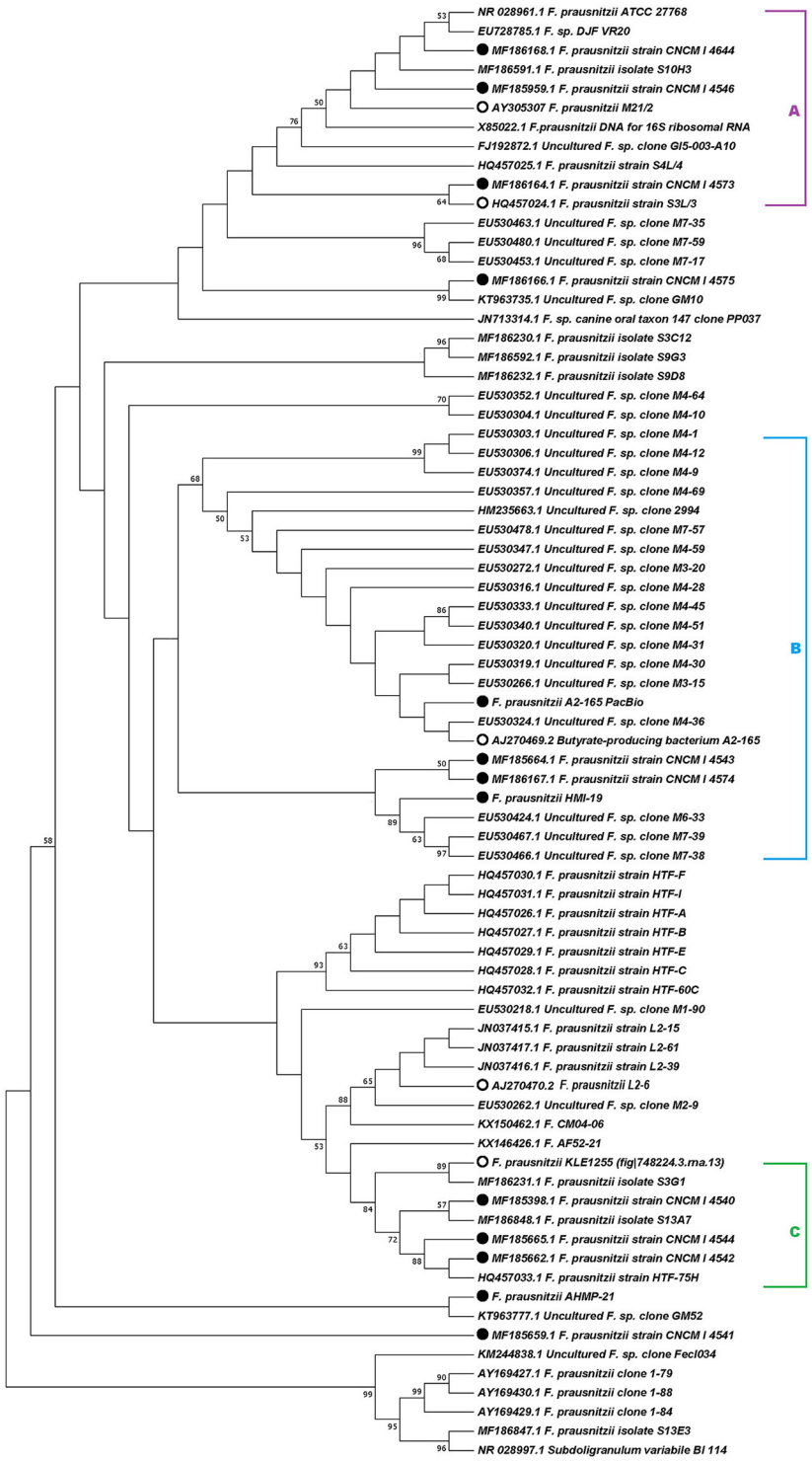

FIGURE 1 | Phylogenetic analysis based on 16S rRNA gene sequences. Evolutionary history was inferred using the maximum-likelihood method based on the Kimura 2-parameter model (Kimura, 1980). The topology of the tree with the highest log likelihood (−3,562.92) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 0.1122)]. The tree is drawn to scale, with branch lengths measured as the number of substitutions per site. The analysis involved 76 nucleotide sequences. All positions containing gaps and missing data were eliminated. The bootstrap analysis was performed with 1,000 replicates. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). Accession numbers of 16S rRNA sequences are given in parentheses. Filled circles indicate the strains newly sequenced for this study and open circles indicate the strains retrieved from PATRIC for genomic analysis.

the specific curve. The formula used to calculate the extrapolated curves of core and singleton genes can be used to predict the final number of CDSs if we consider a high number of genomes. In this formula, the value of tg(θ) represents the convergence value of the size of the core genome or the number of new genes (singletons).

RESULTS

General Features

The number of contigs in the draft genomes varied from 22 to 244. The genome lengths varied by 613,994 bp in size. The GC content varied from 54.81% (*F. prausnitzii* SL3/3) to 58.11% (*F. prausnitzii* CNCM_4541) and the number of predicted CDSs varied from 2,776 to 3,611 (**Table 1**).

Phylogeny

The phylogenetic analysis of 16S rRNA sequences revealed that the genospecies of *Faecalibacterium* can be clustered into different groups. Specifically, the 16S rRNA gene sequences from the new French genomes grouped into clusters A, B, and C (**Figure 1**), as previously proposed by Martín et al. (2017). The 16S rRNA gene sequence from one Australian isolate, *F. prausnitzii* HMI-19, clustered with sequences in group B, whereas the other Australian sequence (*F. prausnitzii*

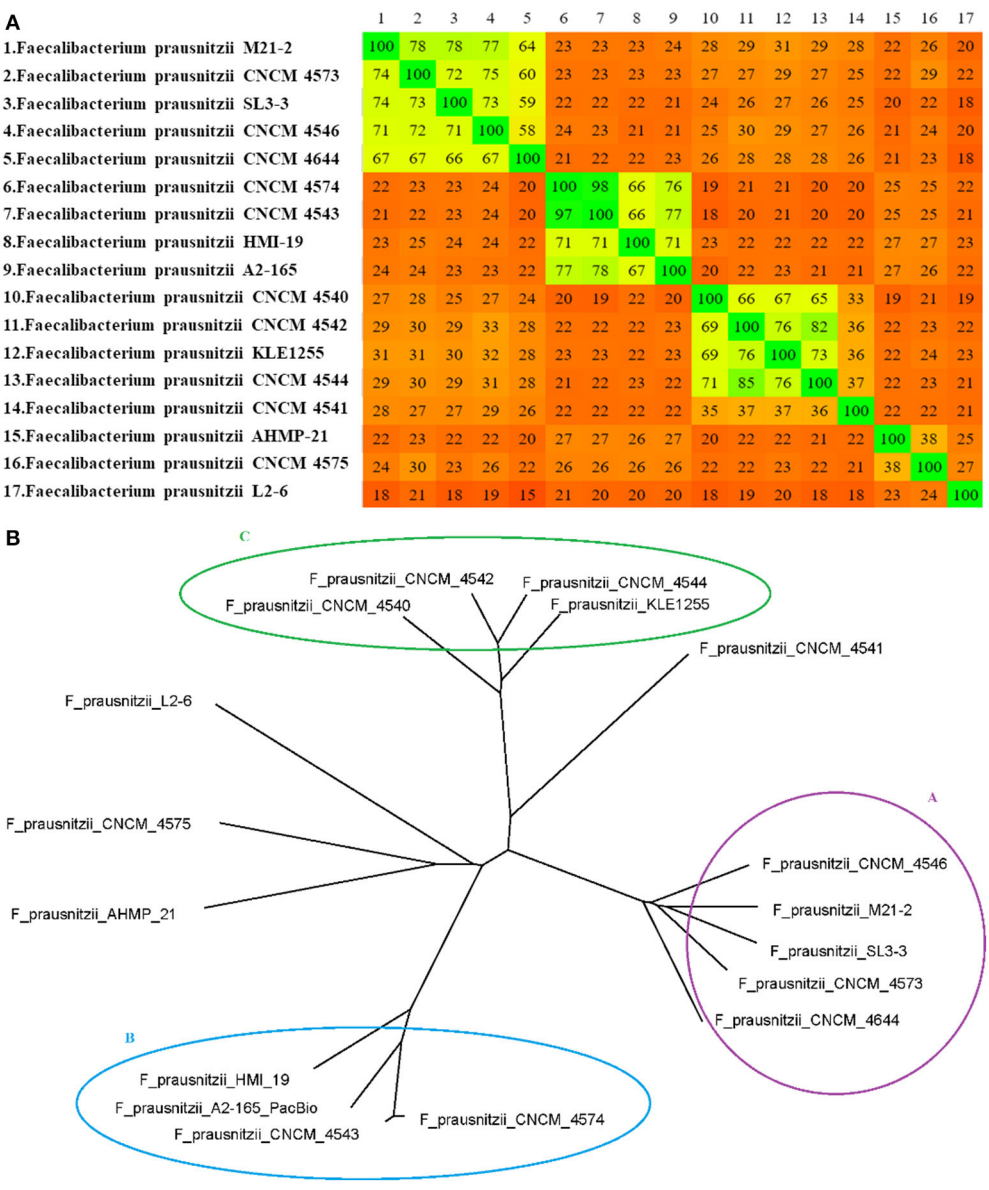

FIGURE 2 | (A) Heatmap and **(B)** distance-matrix-based phylogenetic network of *F. prausnitzii*. The numbers in the heatmap show the percentage of similarity between genomes; the colors vary from red (low similarity) to green (high similarity). The network was constructed using SplitsTree software with NeighborNet and equal angle methods, based on a distance matrix from Gegenees software.

AHMP-21) did not cluster in any of the proposed groups. Likewise, the 16S rRNA sequences from three other strains—*F. prausnitzii*_CNCM_I-4541, *F. prausnitzii*_CNCM_I-4575, and *F. prausnitzii*_L2-6—did not cluster into any of the three groups proposed here.

The distance matrix generated using Gegenees software was plotted as a heatmap (**Figure 2**), in which the similarity among genomes varied from ∼15% (between *F. prausnitzii*_L2-6 and *F. prausnitzii*_CNCM_I_4644) to ∼98% (between the genomes of *F. prausnitzii*_CNCM_I_4543 and *F. prausnitzii*_CNCM_I_4574, which were isolated from the same volunteer). In this analysis, the genomes of *F. prausnitzii*_CNCM_I-4573 and *F. prausnitzii*_SL3/3 clustered together with group A, whereas the genome of *F. prausnitzii*_CNCM_I_4541 was only distantly related to the other strains from group C. As we found with the 16S rRNA analysis, the genome sequences of *F. prausnitzii* AHMP-21, *F. prausnitzii*_CNCM_I-4575, and *F. prausnitzii*_L2-6 did not cluster with any other sequence.

The wgMLST analysis also revealed the presence of three clusters of genomes. Once again, strain *F. prausnitzii*_CNCM_I_4541 was distantly related to other strains in group C, whereas strains *F. prausnitzii* AHMP-21, *F. prausnitzii*_CNCM_I-4575, and *F. prausnitzii*_L2-6 grouped separately from other strains (**Figure 3**).

Average Nucleotide Identity

We performed an Average Nucleotide Identity (ANI) analysis using whole-genome sequence data (**Table 2**). Using an identity cut-off of 94%, this analysis also revealed the presence of the three clusters revealed by the phylogenetic analyses. In addition, the results of the ANI analysis corroborated those of the phylogenomic and wgMLST approaches in finding that the genome sequences of *F. prausnitzii* AHMP-21, *F. prausnitzii*_CNCM_I-4575, *F. prausnitzii*_CNCM_I-4541, and *F. prausnitzii*_L2-6 did not cluster with any other genome sequence. As estimates of ANI are considered to be the gold standard for bacterial species determination, we used the three groups defined here for all further analyses.

Gene Synteny Analysis

Mauve software was used to order the contigs within the genomes and to identify and align regions of local collinearity (called Locally Collinear Blocks, or LCBs), which are regions without local rearrangement of probable homologous sequences that are shared by two or more genomes (Darling et al., 2004). In **Figure 4**, the prediction of LCBs in all strains showed small and numerous regions of homology. When the analysis considered only the genomes within a cluster, the LCBs were larger and less numerous.

Pangenome Calculation

To take a global view of the genome of *Faecalibacterium* and to further explore the genome diversity of this genus, we calculated the size of the pangenome (i.e., the total number of non-redundant CDSs) based on different datasets. When we examined all genomes together, the orthology analysis showed that the pangenome contained a total of 10,366 CDSs (**Figure 5A**), which corresponded to ∼3.33-fold the average total number of genes in each of the 17 strains (3,110.29 CDSs). When we considered only the genomes in group A, we found a total of 5,438 CDSs (**Figure 5B**), ∼1.71-fold the average total number of CDSs in each member strain (3,187.4). The pangenome of group B had 4,311 CDSs (**Figure 5C**), which was ∼1.36-fold the average total number of genes in each member strain (3,159), and group C had 4,686 CDSs in its pangenome (**Figure 5D**), which was ∼1.57-fold the average total number of genes in each member strain (2,991.75). Using the formula $\alpha = 1-\gamma$, we inferred that the α

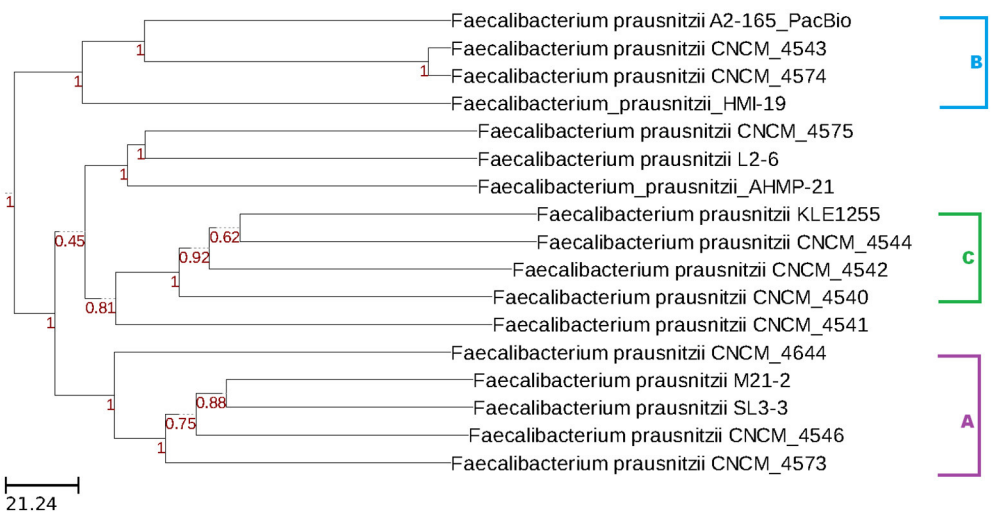

FIGURE 3 | Dendrogram constructed with wgMLST profiles for 17 *F. prausnitzii* genomes. The PGAdb profile from the genomes was used to construct a wgMLST tree using the *Build_wgMLSTtree* module (Liu et al., 2016). Bootstrap values are shown next to the nodes. The dendrogram was constructed with the UPGMA clustering algorithm.

TABLE 2 | Average nucleotide identity.

Strains	CNCM_I_4546	CNCM_I_4573	CNCM_I_4644	M21-2	SL3-3	A2-165_PacBio	CNCM_I_4543	CNCM_I_4574	HMI-19	CNCM_I_4540	CNCM_I_4542	CNCM_I_4544	KLE1255	AHMP-21	CNCM_I_4541	CNCM_I_4575	L2-6
CNCM_I_4546	100	97.37	95.03	97.33	97.36	86.23	86.76	86.85	86.02	86.47	87.6	86.88	87.44	85.48	86.57	86.81	85.72
CNCM_I_4573	97.37	100	95.02	97.13	97.19	86.79	86.2	85.99	86.66	87.11	86.86	86.88	87.21	86.01	86.24	88.42	84.64
CNCM_I_4644	95.03	95.02	100	95.09	94.99	86.28	85.88	85.81	85.6	85.81	86.44	86.34	86.57	85.62	85.91	86.48	84.87
M21-2	97.33	97.13	95.09	100	97.36	86.75	86.03	85.9	85.78	86.48	86.8	86.92	87.35	85.58	86.52	86.79	85.43
SL3-3	97.36	97.19	94.99	97.36	100	86.16	86.47	86.44	86.21	86.03	86.67	86.85	87.05	85.6	86.3	86.33	85.34
A2-165_PacBio	86.23	86.79	86.28	86.75	86.16	100	98.08	97.99	97.12	85.21	86.11	85.47	86.22	86.1	85.28	86.32	86.09
CNCM_I_4543	86.76	86.2	85.88	86.03	86.47	98.08	100	99.9	97.08	84.82	85.58	85.42	85.75	85.85	85.2	85.9	85.71
CNCM_I_4574	86.85	85.99	85.81	85.9	86.44	97.99	99.9	100	97.1	84.77	85.71	85.26	85.63	86	85.02	85.85	85.99
HMI-19	86.02	86.66	85.6	85.78	86.21	97.12	97.08	97.1	100	86.08	85.56	85.59	85.59	85.96	85.47	86.1	85.93
CNCM_I_4540	86.47	87.11	85.81	86.48	86.03	85.27	84.82	84.77	86.08	100	97.62	97.57	97.52	85.03	87.65	85.7	85.61
CNCM_I_4542	87.6	86.86	86.44	86.6	86.67	86.11	85.58	85.71	85.56	97.62	100	98.46	98.1	85.62	88.09	85.97	86.13
CNCM_I_4544	86.88	86.88	86.34	86.92	86.85	85.47	85.42	85.26	85.59	97.57	98.46	100	98.14	85.63	88.05	85.87	86.03
KLE1255	87.44	87.21	86.57	87.35	87.05	86.22	85.75	85.63	85.59	97.52	98.1	98.14	100	85.73	87.94	86.52	86.47
AHMP-21	85.48	86.01	85.61	85.58	85.6	86.1	85.85	86	85.96	85.03	85.62	85.64	85.73	100	85.14	88.31	86.21
CNCM_I_4541	86.57	86.24	85.91	86.52	86.3	85.25	85.2	85.02	85.47	87.65	88.09	88.05	87.94	85.14	100	85.25	85.22
CNCM_I_4575	86.81	88.42	86.48	86.79	86.33	86.32	85.9	85.85	86.1	85.7	85.97	85.87	86.52	88.31	85.25	100	86.87
L2-6	85.72	84.64	84.87	85.43	85.34	86.09	85.71	85.99	85.93	85.61	86.13	86.03	86.47	86.21	85.22	86.87	100

The colors purple, blue, and green corresponds to the clusters A, B, and C respectively.

value of the pangenome of all genomes was 0.56, indicating that the pangenome is probably open and increasing. Similarly, the extrapolation of the pangenome size calculated for groups A, B, and C generated α values of 0.63, 0.77, and 0.66, respectively (**Figure 6**).

Examination of the core genome showed that 1,421 CDSs were shared by all genomes, which corresponded to less than 50% of the average gene content in each genome (3,110.29 CDSs) and represented ∼13.71% of the total pangenome. A separate analysis of the core genome of each group revealed 1,937, 2,036, and 1,940 CDSs, respectively, in groups A, B, and C. The subset of CDSs in all genomes considered to be singletons (i.e. unique to a single genome) contained 4,465 CDSs, while within-group analyses revealed 2,184, 988, and 1,666 singleton CDSs, respectively, within groups A, B, and C (**Figure 5**). By examining the extrapolated curve of the core genome of *Faecalibacterium* ssp., we found that the size of the core genome tended to converge at 1,409 genes, which represented only 13.59% of the pangenome. Within groups A, B, and C, this value increased to 1,910, 2,031, and 1,708 genes, respectively, which represented 35.12, 47.11, and 36.45% of the respective pangenome (**Figure 7**).

DISCUSSION

In bacteria, 16S rRNA sequences have been widely used to study phylogenetic relationships. However, this approach is hampered by the fact that several forces that shape the evolution of bacterial genomes act with different strengths on different parts of the genome and on different bacterial lineages (Janda and Abbott, 2007; Chan et al., 2012). Therefore, to determine the diversity within a bacterial genus or species, it is important to consider not only 16S rRNA sequences, but also the whole genome. Despite this, study of the evolutionary history of genus *Faecalibacterium* has largely been conducted through analyses of 16S rRNA sequences. For example, the first study of 16S rRNA gene sequences of *F. prausnitzii* revealed that this species had been misclassified into genus *Fusobacterium* (Wang et al., 1996; Duncan et al., 2002). After that, Lopez-Siles et al. (2012) used this sequence region to propose the existence of two phylogroups within what is currently considered *F. prausnitzii*. Furthermore, a recent study based on 16S rRNA data showed that there are significant differences among the strains of *F. prausnitzii* in resistance to antibiotics and metabolic activities (Martín et al., 2017). Here we compared the 16S rRNA gene sequences of new *F. prausnitzii* isolates to those previously available, and overall, our results challenge the current concept of a division of isolates into two broad phylogroups. As was initially proposed by Martín et al. (2017), the 16S rRNA gene sequences of our new French *F. prausnitzii* isolates can be clustered into three groups (although one, group B, was indeed supported with a lower bootstrap value than the other two). The Australian isolate *F. prausnitzii*_HMI-19 also clustered into group B, while the other Australian isolate, *F. prausnitzii*_AHMP-21, does not cluster within any of the groups proposed at present. Taken together, our analyses would suggest that there is more phylogenetic complexity in the classification of this species than has been previously shown in other studies (Wang et al., 1996;

FIGURE 4 | Genomic synteny and gene conservation among the genomes of *F. prausnitzii*. The left side of the figure **(A)** shows the LCBs of all genomes studied here. The right side depicts the LCBs of the genomes within each of the three clusters previously obtained from ANI analysis: top right **(B)**—cluster A, middle right **(C)**—cluster B, and bottom right **(D)**—cluster C.

Duncan et al., 2002; Lopez-Siles et al., 2012; Martín et al., 2017). This ambiguity motivated our use of techniques other than 16S analysis in order to better understand the diversity inside genus *Faecalibacterium*.

A whole-genome comparative analysis further validated our findings from the 16S rRNA gene phylogeny. A whole-genome similarity matrix was obtained with Gegenees software and used for a phylogenomic analysis; the resulting phylogenetic tree agreed with the previously performed 16S rRNA analysis in identifying the same three groups of strains: clusters A, B, and C. In this analysis, the genospecies *F. prausnitzii*_CNCM_I_4541 clustered within group C, but the relationship between this strain and other members of group C is very distant, reflecting the low degree of genomic similarity between the former and the latter (∼36% similarity). The same pattern was found for *F. prausnitzii* AHMP-21, *F. prausnitzii*_CNCM_I_4575, and *F. prausnitzii*_L2-6, which were grouped together, but at similarity values ranging from ∼23 to ∼38% (as observed in the heatmap). It is interesting to note that certain strains that were isolated from the same volunteer were quite dissimilar (∼27% similarity between FPR_CNCM_I_4573 and FPR_CNCM_I_4575; ∼37% between FPR_CNCM_I_4541 and FPR_CNCM_I_4542), suggesting that the same individual may harbor different genospecies. The overall abundance of a given genospecies of *Faecalibacterium* within an individual host may be extremely relevant to the study of human diseases, as this overall abundance depends on the disease under study (Hippe et al., 2016; Lopez-Siles et al., 2016). For example, as part of a case-control study of atopic dermatitis (AD) in Korean subjects, Song et al. (2016) reported that 16S rRNA PCR amplicons from stool samples of AD patients were enriched in those similar to strain L2-6 with respect to other strains of *F. prausnitzii*; they also proposed that strain L2-6 can be differentiated from other strains by the existence of a polycistronic region encoding GalNac uptake and metabolism (Song et al., 2016). Our analysis here showed that this strain does indeed demonstrate a very distinct phylogenetic pattern, which increases its potential for use as a reference strain in future AD studies.

To improve the resolution of our phylogenetic analysis, we also applied a strategy based on wgMLST analysis. As opposed to conventional MLST analysis, which uses only a few housekeeping genes, the wgMLST approach takes advantage of a larger number of tracked loci, enabling higher resolution in intraspecies differentiation (Maiden et al., 2013). Our analysis considered only genes that shared more than 80% coverage and 90% identity. Here, the same three groups of genospecies (A, B, and C) were also detected, and again strain *F. prausnitzii*_CNCM_I_4541 was only distantly related to other members of group C. A group containing the isolates *F. prausnitzii*_AHMP-21, *F. prausnitzii*_CNCM_I_4575, and *F. prausnitzii*_L2-6 was also observed. In sum, each of the three phylogenetic analyses we performed suggested the existence of more than one genospecies within the genus *Faecalibacterium*.

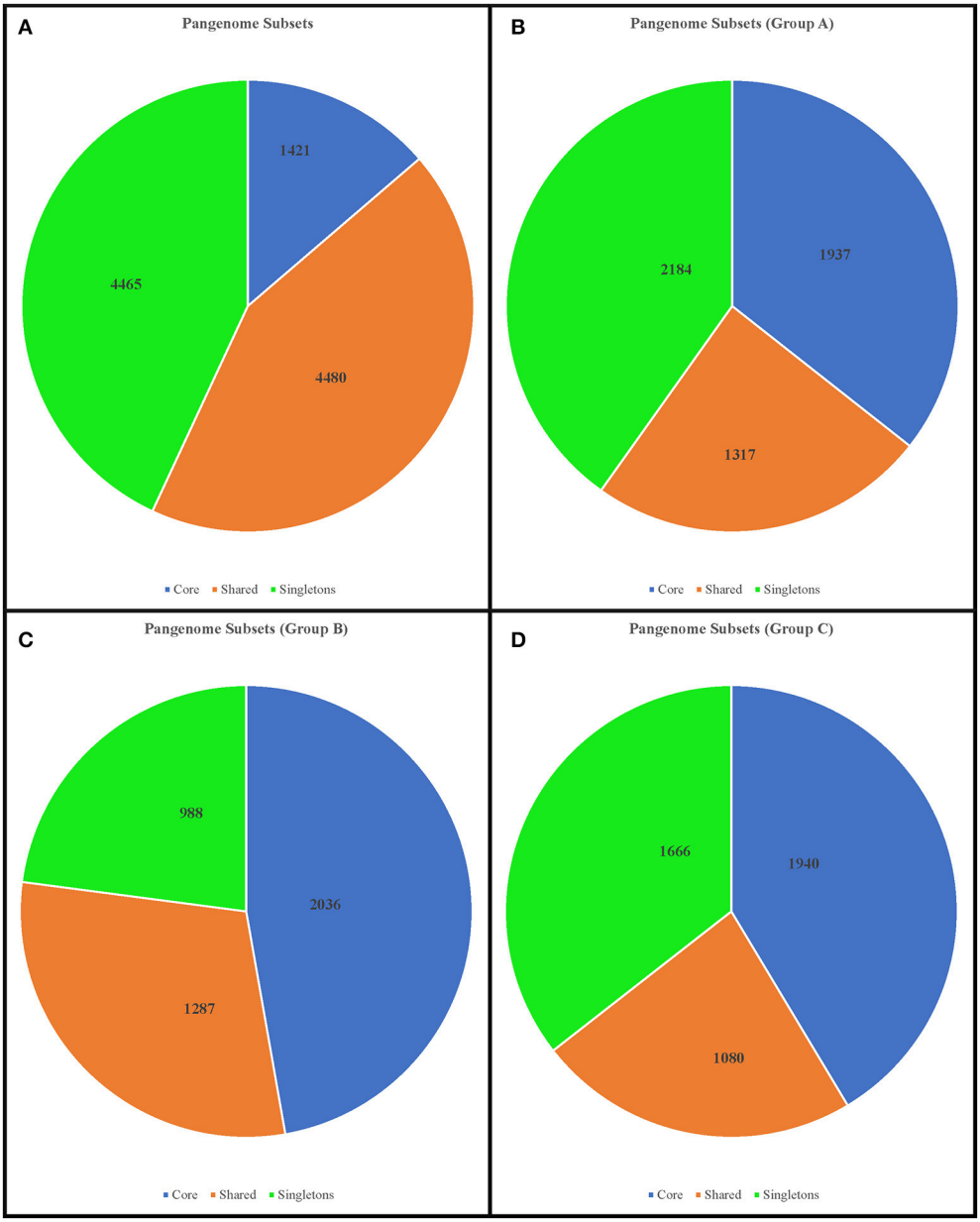

FIGURE 5 | Diagram depicting the subsets of the *Faecalibacterium* pangenome. The numbers represent the coding sequences belonging to each subset. Upper left chart **(A)**: pangenome subsets from an analysis based on all 17 genomes of *Faecalibacterium*. Upper right chart **(B)**: subset based on analysis of 5 genomes from group A. Lower left chart **(C)**: subset based on analysis of 4 genomes from group B. Lower right chart **(D)**: subset based on analysis of 4 genomes from group C.

To further corroborate the existence of these potential new "species," we performed an ANI analysis, which confirmed the new relationships identified in the previous analyses. The ANI analysis supported the classification of *F. prausnitzii*_CNCM_I_4541 as a distinct genospecies separate from group C; likewise, the genomes of *F. prausnitzii*_AHMP-21, *F. prausnitzii*_CNCM_I_4575, and *F. prausnitzii*_L2-6 were found to be quite dissimilar from all other genomes considered.

Using our revised clustering of the *F. prausnitzii* genomes, supported by the ANI results, we then investigated genome diversity via gene synteny analysis and calculations of pangenome. The extent of intra-cluster gene synteny was clearly evident in the Mauve alignments. Furthermore, the number and the lengths of the LCBs in the all-genomes dataset were strikingly different from those in the three intra-cluster datasets, which together indicated a higher degree of genome similarity within than among clusters, particularly with regard to group B. Even within a single genospecies, different genomes had a considerable number of regions with inversions and deletions, which may have arisen from horizontal gene transfer events.

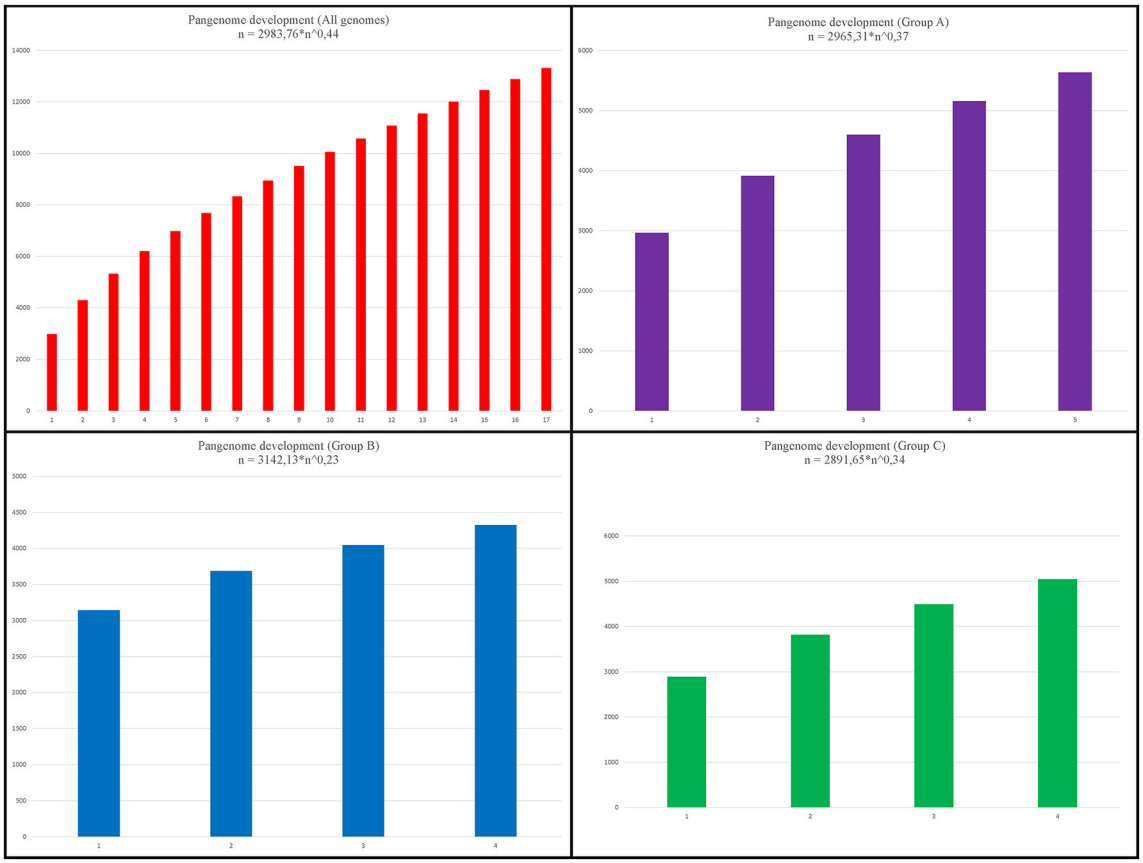

FIGURE 6 | Pangenome development. Upper left chart: pangenome development based on permutations of all 17 genomes of *Faecalibacterium*. Upper right chart: development based on permutations of 5 genomes from group A. Lower left chart: development based on permutations of 4 genomes from group B. Lower right chart: development based on permutations of 4 genomes from group C.

Again, though, the genomes from group B were more similar to each other than were the genomes of the other groups.

The same four datasets were used to perform calculations of pangenome. As might be expected, the number of core-genome CDSs was greater within each cluster than within the dataset containing all 17 genomes, which is consistent with the idea that the genomes within a cluster are from the same species. Extrapolations of pangenome development also corroborated this assumption. The α value generated from an analysis of all genomes indicated that the genus *Faecalibacterium* has an open pangenome ($\alpha = 0.56$), as does each of the groups ($\alpha = 0.63$, $\alpha = 0.77$, and $\alpha = 0.66$, respectively). However, the intra-group α values reveal that these latter pangenomes are increasing in size more slowly than the pangenome of all species (as indicated by higher α values). This means that, if we consider all the genomes to be part of the genus *Faecalibacterium*, each new genome sequenced will increase substantially the number of non-redundant genes in this genus. On the other hand, the genomes within each group tend to be more clonal, and newly sequenced genomes included within these groups will have a less prominent effect on the number of non-redundant genes. We likewise arrived at the same conclusion by analyzing the development of the core genome and singletons: the final core genome tended to be larger within each genospecies than within the all-genome analysis. This phylogenetic approach to pangenome analysis revealed patterns that were totally in accordance with the results of our other analyses.

CONCLUDING REMARKS

Here, we used a variety of methods to analyze 16S rRNA and whole genome data, which together showed that: (i) the current application of phylogroups to differentiate among strains of *F. prausnitzii* should be revised; (ii) this genus contains at least three separate clusters, spanning both phylogroups I and II, which are all derived from a common recent ancestor; and (iii) some strains (e.g., *F. prausnitzii* AHMP-21, *F. prausnitzii*_L2-6, and *F. prausnitzii*_CNCM_I_4575) appear to represent a deeper, more divergent branch of "*Faecalibacterium prausnitzii*." Collectively, our results provide evidence for the reconsideration of the phylogenetic and genomic relatedness among strains currently assigned to *F. prausnitzii*. In addition, they highlight the need for lineage (strain-based) differentiation within this genus to better define how specific members might

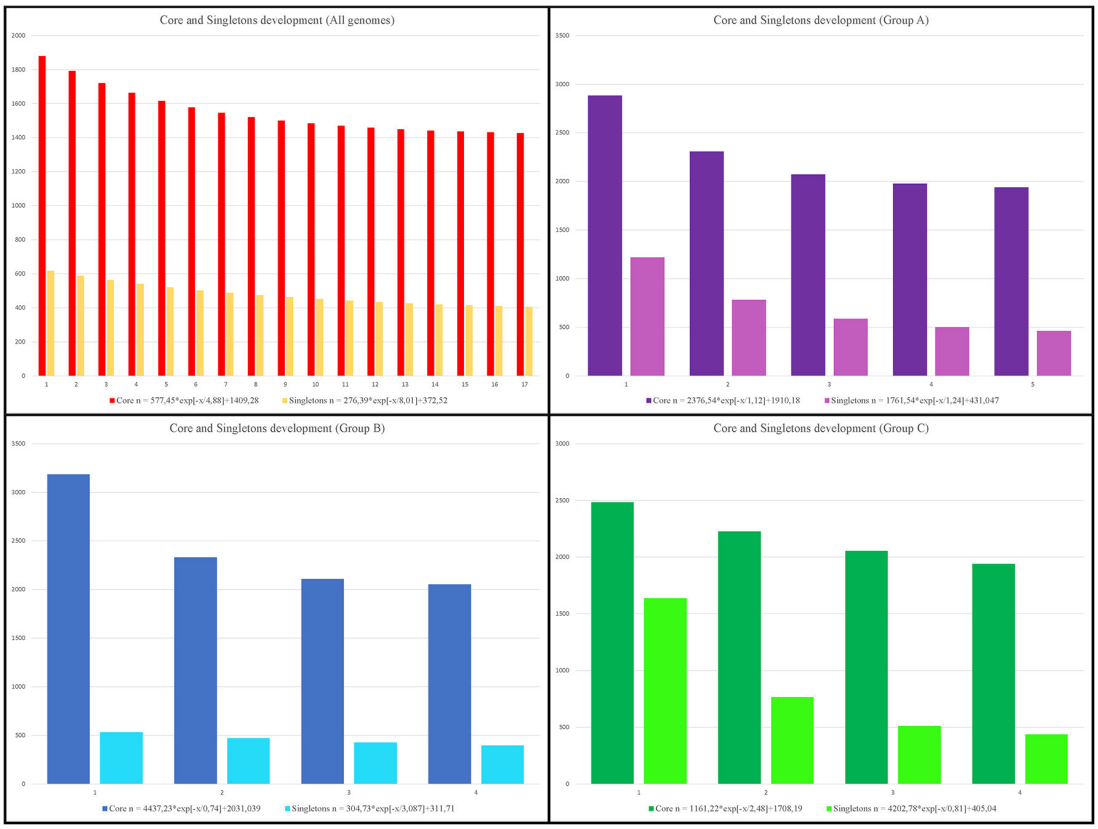

FIGURE 7 | Development of core genome and singletons. Upper left chart: core-genome and singleton development based on permutations of all 17 genomes of *Faecalibacterium*. Upper right chart: development based on permutations of 5 genomes from group A. Lower left chart: development based on permutations of 4 genomes from group B. Lower right chart: development based on permutations of 4 genomes from group C.

be associated with positive or negative host interactions. Such lineage-specific variations might not only explain the variable abundances of *F. prausnitzii* linked with adverse health outcomes (e.g., atopic dermatitis, Crohn's disease, and ulcerative colitis; Swidsinski et al., 2008; Hansen et al., 2012), but also provide new opportunities for the diagnosis and strain-specific treatment of gut inflammation and associated diseases. Also, to the best of our knowledge, this is the first work to combine an analysis of pangenome development with ANI analysis in order to corroborate the assignment of strains to new species. Here, we propose that pangenome subsets and the α value generated by these analyses may be used as an alternative approach, together with ANI, for the *in silico* classification of new species. Although low α values may be found inside a species cluster, due to a high degree of variation among genomes arising from intense horizontal gene transfer events, a high intra-cluster α value may be considered a good indicator of a new, more-clonal species inside the genus.

AUTHOR CONTRIBUTIONS

LB, SB, RM, VR, MT, SM, FC, HS, LGB, MM, PL, VA, JC, and SS designed the experiments, revised the manuscript critically, and participated in the design of the project. LB, SB, RM, VR, SM, FC, and SC performed the experiments and analysis. LB and SC drafted the manuscript.

FUNDING

This work was supported by a scholarship from the International Cooperation Program CAPES/COFECUB (financed by the Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil) and was conducted at the "Institut National de la Recherche Agronomique" (INRA), France. This paper was a part of FPARIS collaborative project selected and supported by the Vitagora Competitive Cluster and funded by the French FUI (Fond Unique Interministériel; FUI: n°F1010012D), FEDER (Fonds Européen de Développement Régional; Bourgogne: 34606), the region of Burgundy, the Conseil Général 21, and the Grand Dijon. This work was also supported by Merck Médication Familiale (Dijon, France) and Biovitis (Saint Etienne de Chomeil, France). RM and SM receive a salary from these same grants. SB is the recipient of an Australian Postgraduate award. The contributions of MM and SB have been partially supported with funds from the University of Queensland, the University of Queensland Diamantina Institute, and the Helmsley Charitable Trust (via the Australasian Gastro Intestinal Foundation).

REFERENCES

Agren, J., Sundström, A., Håfström, T., and Segerman, B. (2012). Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS ONE* 7:e39107. doi: 10.1371/journal.pone.0039107

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Castillo, M., Skene, G., Roca, M., Anguita, M., Badiola, I., Duncan, S. H., et al. (2007). Application of 16S rRNA gene-targetted fluorescence in situ hybridization and restriction fragment length polymorphism to study porcine microbiota along the gastrointestinal tract in response to different sources of dietary fibre. *FEMS Microbiol. Ecol.* 59, 138–146. doi: 10.1111/j.1574-6941.2006.00204.x

Chan, J. Z. M., Halachev, M. R., Loman, N. J., Constantinidou, C., and Pallen, M. J. (2012). Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. *BMC Microbiol.* 12:302. doi: 10.1186/1471-2180-12-302

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Duncan, S. H., Hold, G. L., Harmsen, H. J. M., Stewart, C. S., and Flint, H. J. (2002). Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. *nov., comb. nov. Int. J. Syst. Evol. Microbiol.* 52, 2141–2146. doi: 10.1099/00207713-52-6-2141

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hansen, R., Russell, R. K., Reiff, C., Louis, P., McIntosh, F., Berry, S. H., et al. (2012). Microbiota of *de-novo* pediatric ibd: increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am. J. Gastroenterol.* 107, 1913–1922. doi: 10.1038/ajg.2012.335

Hippe, B., Remely, M., Aumueller, E., Pointner, A., Magnet, U., and Haslberger, A. G. (2016). *Faecalibacterium prausnitzii* phylotypes in type two diabetic, obese, and lean control subjects. *Benef. Microbes* 7, 511–517. doi: 10.3920/BM2015.0075

Huson, D. H. (2005). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Janda, J. M., and Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45, 2761–2764. doi: 10.1128/JCM.01228-07

Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liu, Y. Y., Chiou, C. S., and Chen, C. C. (2016). PGAdb-builder: a web service tool for creating pan-genome allele database for molecular fine typing. *Sci. Rep.* 6:36213. doi: 10.1038/srep36213

Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J., and Martinez-Medina, M. (2017). *Faecalibacterium prausnitzii*: from microbiology to diagnostics and prognostics. *ISME J.* 11, 841–852. doi: 10.1038/ismej.2016.176

Lopez-Siles, M., Khan, T. M., Duncan, S. H., Harmsen, H. J. M., Garcia-Gil, L. J., and Flint, H. J. (2012). Cultured representatives of two major phylogroups of human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl. Environ. Microbiol.* 78, 420–428. doi: 10.1128/AEM.06858-11

Lopez-Siles, M., Martinez-Medina, M., Surís-Valls, R., Aldeguer, X., Sabat-Mir, M., Duncan, S. H., et al. (2016). Changes in the abundance of *Faecalibacterium prausnitzii* phylogroups I and II in the intestinal mucosa of inflammatory bowel disease and patients with colorectal cancer. *Inflamm. Bowel Dis.* 22, 28–41. doi: 10.1097/MIB.0000000000000590

Maiden, M. C. J., van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi: 10.1038/nrmicro3093

Martín, R., Miquel, S., Benevides, L., Bridonneau, C., Robert, V., Hudault, S., et al. (2017). Functional characterization of novel *Faecalibacterium prausnitzii* strains isolated from healthy volunteers: a step forward in the use of *F. prausnitzii* as a next-generation probiotic. *Front. Microbiol.* 8:1226. doi: 10.3389/fmicb.2017.01226

Miquel, S., Martín, R., Bridonneau, C., Robert, V., Sokol, H., Bermúdez-Humarán, L. G., et al. (2014). Ecology and metabolism of the beneficial intestinal commensal bacterium *Faecalibacterium prausnitzii*. *Gut Microb.* 5, 146–151. doi: 10.4161/gmic.27651

Miquel, S., Martín, R., Rossi, O., Bermúdez-Humarán, L. G., Chatel, J. M., Sokol, H., et al. (2013). *Faecalibacterium prausnitzii* and human intestinal health. *Curr. Opin. Microbiol.* 16, 255–261. doi: 10.1016/j.mib.2013.06.003

Nava, G. M., and Stappenbeck, T. S. (2011). Diversity of the autochthonous colonic microbiota. *Gut Microbes* 2, 99–104. doi: 10.4161/gmic.2.2.15416

Oikonomou, G., Teixeira, A. G. V., Foditsch, C., Bicalho, M. L., Machado, V. S., and Bicalho, R. C. (2013). Fecal microbial diversity in pre-weaned dairy calves as described by pyrosequencing of metagenomic 16S rDNA. Associations of *Faecalibacterium* species with health and growth. *PLoS ONE* 8:e63157. doi: 10.1371/journal.pone.0063157

Rajilić-Stojanović, M., Biagi, E., Heilig, H. G. H. J., Kajander, K., Kekkonen, R. A., Tims, S., et al. (2011). Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* 141, 1792–1801. doi: 10.1053/j.gastro.2011.07.043

Scupham, A. J. (2007). Succession in the intestinal microbiota of preadolescent turkeys. *FEMS Microbiol. Ecol.* 60, 136–147. doi: 10.1111/j.1574-6941.2006.00245.x

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J. J., et al. (2008). *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16731–16736. doi: 10.1073/pnas.0804812105

Song, H., Yoo, Y., Hwang, J., Na, Y.-C., and Kim, H. S. (2016). *Faecalibacterium prausnitzii* subspecies–level dysbiosis in the human gut microbiome underlying atopic dermatitis. *J. Allergy Clin. Immunol.* 137, 852–860. doi: 10.1016/j.jaci.2015.08.021

Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M., and Doerffel, Y. (2008). Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. *Inflamm. Bowel Dis.* 14, 147–161. doi: 10.1002/ibd.20330

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-P., et al. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* 11, 2574–2584. doi: 10.1111/j.1462-2920.2009.01982.x

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., et al. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* 5, 220–230. doi: 10.1038/ismej0.2010.118

Wang, R. F., Cao, W. W., and Cerniglia, C. E. (1996). Phylogenetic analysis of *Fusobacterium prausnitzii* based upon the 16S rRNA gene sequence and PCR confirmation. *Int. J. Syst. Bacteriol.* 46, 341–343. doi: 10.1099/00207713-46-1-341

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

V.2 Chapter II. *In silico* evaluation of safety aspects in the use of *Faecalibacterium* strains by genome plasticity and antibiotic resistance approaches

Leandro Benevides, Henrique C. P. Figueiredo, Philippe Langella, Vasco A. Azevedo, Jean-Marc Chatel and Siomar Soares

This article is in preparation for submission in the journal Nucleic Acids Research.

After observing the existence of various species in the genus *Faecalibacterium*, we started to evaluate the safety aspects in the use of *Faecalibacterium* strains as probiotic. To that, we used bioinformatics tools that allowed us to screen candidates for further *in vitro* and *in vivo* analyzes. Among the comparative genomics analyzes to evaluate the safety aspects in the use of *Faecalibacterium* strains as probiotic, we did the (1) prediction of horizontally acquired regions (Antibiotic resistance islands, Metabolic islands and phage regions), (2) prediction of metabolic pathways, (3) search of genes related to antibiotic resistance and (4) search of bacteriocins.

In silico evaluation of safety aspects in the use of *Faecalibacterium* strains by genome plasticity and antibiotic resistance approaches

Leandro Benevides[1,2,5], Letícia C. Oliveira[1], Henrique C. P. Figueiredo[3,4], Philippe Langella[2], Vasco A. Azevedo[1], Jean-Marc Chatel[2] and Siomar Soares[5]

[1]Department of General Biology, Federal University of Minas Gerais, Belo Horizonte, Brazil

[2]Commensals and Probiotics-Host Interactions Laboratory, Micalis Institute, Institut National de la Recherche Agronomique, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

[3]AQUACEN, National Reference Laboratory for Aquatic Animal Diseases, Ministry of Fisheries and Aquaculture, Federal University of Minas Gerais, Belo Horizonte, Brazil

[4]Veterinary School, Department of Preventive Veterinary Medicine, Federal University of Minas Gerais, Belo Horizonte, Brazil

[5]Department of Immunology, Microbiology and Parasitology, Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, Brazil

ABSTRACT

Several studies suggest the use of *Faecalibacterium spp.* as a probiotic. This genus is a major member of the Firmicutes phylum and one of the most abundant bacteria in the healthy human microbiota. Nowadays, it's describe the existence of at least 7 species into this genus and, as the probiotic properties are usually strain-specific ones, it is necessary to evaluate the probiotic potential of these bacteria considering the distinct species and strains. As they are very difficult to isolate and grow, the *in silico* evaluation of its probiotic potential is a good strategy to determinate the best candidate strains to be used as commercial probiotic and to understand its role in human homeostasis. In this work, we performed comparative genomic analyzes to provide insights into the genetic basis of the probiotic potential of *Faecalibacterium* strains. Briefly, we performed analyzes involving (1) prediction of horizontally acquired regions (Antibiotic resistance islands, Metabolic islands and phage regions), (2) prediction of metabolic pathways, (3) search for genes related to antibiotic resistance and (4) search for bacteriocins. Our results show the existence of Antibiotic resistance islands and Metabolic islands in all genomes, whereas phage regions were only not found in strain CNCM_I_4541. 126 metabolic pathways were predicted including important ones like: Bisphenol A degradation, Butanoate metabolism and Streptomycin biosynthesis. Among the genomes, we found 8 genes related to antibiotic resistance mechanisms, but the presence/absence of these genes inside RI seems not to interfere in the mechanisms of antibiotic resistance demonstrated in *in vitro* studies. No bacteriocins were predicted using the software BAGEL4, but our group have previously identified *in vitro* one protein (MAM protein) with some characteristics similar to bacteriocins that were investigated here. Altogether, this work presents a screening of the main safety aspects in the use of *Faecalibacterium* strains as probiotics.

INTRODUCTION

The consumption of probiotic foods worldwide is constantly increasing leaded by the awareness of its beneficial effects in promoting gut and general human health. The recent advances in omics tools and sequencing techniques have supported this increased consumption and improved our understanding of probiotic functionality and also, the specific interactions between probiotics and their human hosts. In this context, a bacterium that has attracted attention is *Faecalibacterium prausnitzii*. *F. prausnitzii*, a dominant human intestinal bacterial species, which is considered as a sensor and an actor of the human intestinal health (Benevides et al., 2017; Martín et al., 2017; Sokol et al., 2008). Within the human colon, the genus *Faecalibacterium* is the main member of the *Clostridium leptum* cluster, and comprises the second-most common representative in fecal samples, after *Clostridium coccoides* (Tap et al., 2009; Walker et al., 2011). Until recently, it was believed that there was only one species in this genus, but since 2012, some studies have begun to suggest a higher complexity (Hippe et al., 2016; Lopez-Siles et al., 2016; Martín et al., 2017; Song et al., 2016). In 2017, a deeper phylogenic analysis considering 17 genomes of this genus has suggested that there could be seven different species, spanning the two classical phylogroups proposed before to *Faecalibacterium* (Benevides et al., 2017). These findings bring new possibilities to the studies of *Faecalibacterium* probiotic potential.

As the probiotic properties are usually strain-specific ones (Pineiro and Stanton, 2007), individual studies are required to assess the probiotic properties of the different *Faecalibacterium* species isolated. Recently, our group analyzed, *in vitro*, the probiotic properties of different strains from *Faecalibacterium* genus (Martín et al., 2017). In this study, we considered diverse metabolic and safety tests and showed that the strains have different profiles endorsing the existence of many species in this genus. Important anti-inflammatory properties have being reported to the reference strain *F. prausnitzii* A2-165, more specifically for its anti-inflammatory effects both *in vitro* and *in vivo* in different murine models of colitis (Martín et al., 2014, 2015; Sokol et al., 2008). These properties are partially attributed to the ability to produce butyrate (Barcenilla et al., 2000;

Duncan et al., 2002) and the expression of a protein know as Microbial Anti-inflammatory Molecule (MAM; Breyner et al., 2017; Quévrain et al., 2014). Butyrate is a short chain fatty acid (SCFA) well- known for its pleiotropic and beneficial effects in the GIT (Duncan et al., 2002; Macfarlane and Macfarlane, 2011) as well as its immune-modulatory properties *in vitro* (Böcker et al., 2003). Butyrate can reduce intestinal mucosa inflammation through the inhibition of NF-κB transcription factor activation, upregulating PPARγ and inhibiting interferon gamma (Lopez-Siles et al., 2017). The MAM protein is exclusive of *Faecalibacterium* and is also capable of inhibiting the NF-κB pathway in several intestinal epithelial cell lines (Quévrain et al., 2014). *In vivo* studies confirmed the anti-inflammatory properties of MAM in a DNBS-induced colitis and in DSS model. In DSS model, MAM was able to inhibit Th1 and Th17 immune response while in DNBS model MAM reduced Th1, Th2, and Th17 immune response and increased TGFβ production (Breyner et al., 2017). Due to the previously reported probiotic properties and the existence of diverse species, it is important to better understand the biodiversity and beneficial effects of these species. For this is important to characterize *in silico*, *in vitro* and *in vivo* the strains that have a potential probiotic potential (Papadimitriou et al., 2015). The European Food Safety Authority proposed a pre-market safety assessment of selected groups of microorganisms leading to a Qualified Presumption of Safety (QPS). This proposal suggests that a safety assessment of a defined taxonomic group (e.g. genus or group of related species) could be made based on four pillars (establishing identity, body of knowledge, possible pathogenicity and end use). If the taxonomic group did not raise safety concerns or, if safety concerns existed, but could be defined and excluded, the grouping could be granted QPS status (J.B. Prajapati and Suja Senan, 2013).

As *Faecalibacterium spp* are Extremely Oxygen Sensitive (EOS) and very difficult to isolate and grow, the *in silico* evaluation of its probiotic potential is a good strategy to determinate the best candidate strains to be used as commercial probiotic and to understand why this bacterium is a major player of human homeostasis.

The bacterial whole-genome sequencing has recently become a low cost-effective and a suitable approach for a rapidly and accurate screening of

potential probiotic candidates (Didelot et al., 2012; Senan et al., 2015). This approach allows us to detect and discard candidate strains that have potential risk factors, like the presence of antibiotic resistance or virulence genes. It also facilitates the analysis and description of functional mechanisms, avoiding the difficulties of isolating and growing the microorganisms (Papadimitriou et al., 2015). Therefore, using an *in silico* approach and the available genomes on public database, we performed bioinformatic analysis to provide insights into the genetic basis of the probiotic potential of *Faecalibacterium* strains. Here we used the 17 genomes used by Benevides et al. (2017) plus the genome *Faecalibacterium prausnitzii* Indica (GenBank: CP023819.1), that is a genome isolated from the gut of healthy adult Indian. This genome was sequenced using Illumina HiSeq and Nanopore technologies and was assembled using SPAdes resulting in one complete chromosome.

Based on the four pillars proposed by the European Food Safety Authority to safety assessment of selected groups of microorganisms, our group have previously performed a deep phylogenetic investigation on the available genomes of *Faecalibacterium* spp (Benevides et al., 2017). Here, we performed analyzes involving (1) Prediction of horizontally acquired regions (Antibiotic resistance islands, Metabolic islands and phage regions), (2) Prediction of metabolic pathways, (3) search of genes related to antibiotic resistance and (4) search of bacteriocins.

METHODS

Prediction of horizontally acquired regions

Genomic islands prediction

Genomic islands (GEIs) were predicted with the Genomic Island Prediction software (GIPSy) (https://www.bioinformatics.org/groups/?group_id=1180/), choosing the option for the prediction of Antibiotic resistance islands (RIs) and Metabolic islands (MIs). GIPSy predicts putative genomic islands by searching

for regions larger than 6 kb that show genomic signature deviations (i.e., deviations in G+C content or codon usage), transposase genes, antibiotic and metabolic-related factors (in our study), and flanking tRNAs. Additionally, it also checks for the absence of the target region from closely related species (Soares et al., 2015). The putative GEIs were predicted using the *Subdoligranulum variabile* DSM 15176 genome [GenBank: NZ_ACBY00000000.2] as the closely related reference organism of a distinct genus. After, we consolidated and manually curated the results.

Phage regions prediction

Prophage sequences were predicted using the PHASTER web server (http://phaster.ca/) with default parameters (Arndt et al., 2016). Briefly, Phaster uses three classification methods, based in the percentage of predicted proteins that share homology to proteins from phage organisms and a database of phage-related keywords to calculate the score of the predictions. Finally, the phages are classified as intact, questionable and incomplete when the scores are >90, in the range between 70 and 90, or lower than 90, respectively (Arndt et al., 2016).

Bacteriocin prediction

The bacteriocin prediction for all genomes was performed in the web-based software tool BAGEL4 (http://bagel4.molgenrug.nl/). This software uses three different databases containing modified or unmodified bacteriocins and the DNA sequences used as input are analyzed in using two different approaches, one based on finding genes commonly found in the context of bacteriocin, and the other based on finding the gene itself (van Heel et al., 2013).

We have also investigated the synteny of one gene, exclusive to *Faecalibacterium spp*, that was identified by our group in other studies and encodes a protein that have characteristics similar to bacteriocins: the Microbial

Anti-Inflammatory Molecule (MAM) protein (Breyner et al., 2017; Quévrain et al., 2014). To perform this analysis, we used the software MAUVE (Darling et al., 2004).

Metabolic pathways prediction

The Comparative Pathway Tool software from Pathosystems Resource Integration Center (PATRIC) (Wattam et al., 2017) was used to reconstruct the *Faecalibacterium* species metabolic pathways using their genome sequences in "fasta" format. Briefly, it uses the annotation of the genomes to extract their EC Number, compare with Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/) and construct the KEGG maps and the Heatmaps. This tool allows us to identify a set of pathways based on taxonomy, EC number, pathway ID and pathway name. The Comparative Pathway Tool exhibits different information, like Genome Count, Unique Gene Count, Unique EC Count, EC Conservation % and Gene Conservation.

Circular comparison map of genomic sequences

To create circular genome comparisons, we used the software BRIG (Alikhan et al., 2011) and all genome sequences in ".fasta" format. This software performs BLAST analyses of all genomes against the reference strain to create the circular comparison map. Here, the analyses were performed using different genomes as reference. Also, the coordinates of the genomic islands and phage regions were added to the figure to visualize genome plasticity events.

Search of genes related to antibiotic resistance

Two databases were used to search for genes related to antibiotic resistance: The Comprehensive Antibiotic Resistance Database (CARD) (Jia et al., 2017)

and Antibiotic Resistance Genes Database (ARDB) (Liu and Pop, 2009). We also verified the G+C Content, Codon Usage and the presence of these genes in Resistance Islands using the software GIPSy (Soares et al., 2015).

RESULTS

Prediction of horizontally acquired genes

Genomic islands prediction

A summary of most analyses performed here is listed in Table 1. The genome of *Subdoligranulum variabile* DSM 15176 was used as reference to predict the GIs to all *Faecalibacterium* genomes. In supplementary files, one can find other characteristics of the predicted GIs, like: island length, location in the genome, number of genes and the prediction score as “Weak”, “Normal” or “Strong”. To visualize the predicted GIs and the BLASTn comparison of the genes among these genomes, the software BRIG was used. Four distinct genomes were used as reference to show the predicted islands by BRIG software. Three genomes representing each of the 3 genospecies proposed by Benevides et al. (2017) (Figures 1-3) and the genome from strain L2-6 (Figure 4). In the figure 1, the genomes of *Faecalibacterium* are represented. In this example, the genome *F. prausnitzii* A2-165_PacBio was used as reference and the GI and phage regions predicted for this genome are represented in the outermost rings. After the BLASTn comparison, some gapped regions could be observed that indicate the absence or low similarity among the genomes. The same kind of result is shown in figures 2-4, using the genomes of the strains CNCM_I_4540 and CNCM_I_4644 as references from the groups C and A on figures 2 and 3, respectively, and in the figure 4, the strain L2-6.

Table 1 – Number of the predicted Genomic islands, Phage regions, Metabolic Pathways and antibiotic resistance genes.

Strains	Genospecies	MI	RI	Phage	Metabolic Pathways	Resistance genes
4541	-	18	14	0	123	5
4575	-	21	20	1	115	3
L2-6	-	24	26	4	113	3
AHMP21	-	17	15	3	115	3
4546	A	18	10	7	123	6
4573	A	13	11	1	125	4
4644	A	12	12	2	117	4
M21-2	A	15	12	1	123	4
SL3-3	A	17	14	2	116	3
4543	B	25	15	4	117	4
4574	B	23	14	8	117	4
A2-165_PacBio	B	24	19	3	116	4
HMI19	B	5	7	1	114	4
Indica	B	8	10	2	114	-
4540	C	25	10	3	119	4
4542	C	19	12	5	119	5
4544	C	15	7	3	118	4
KLE1255	C	17	8	4	117	3

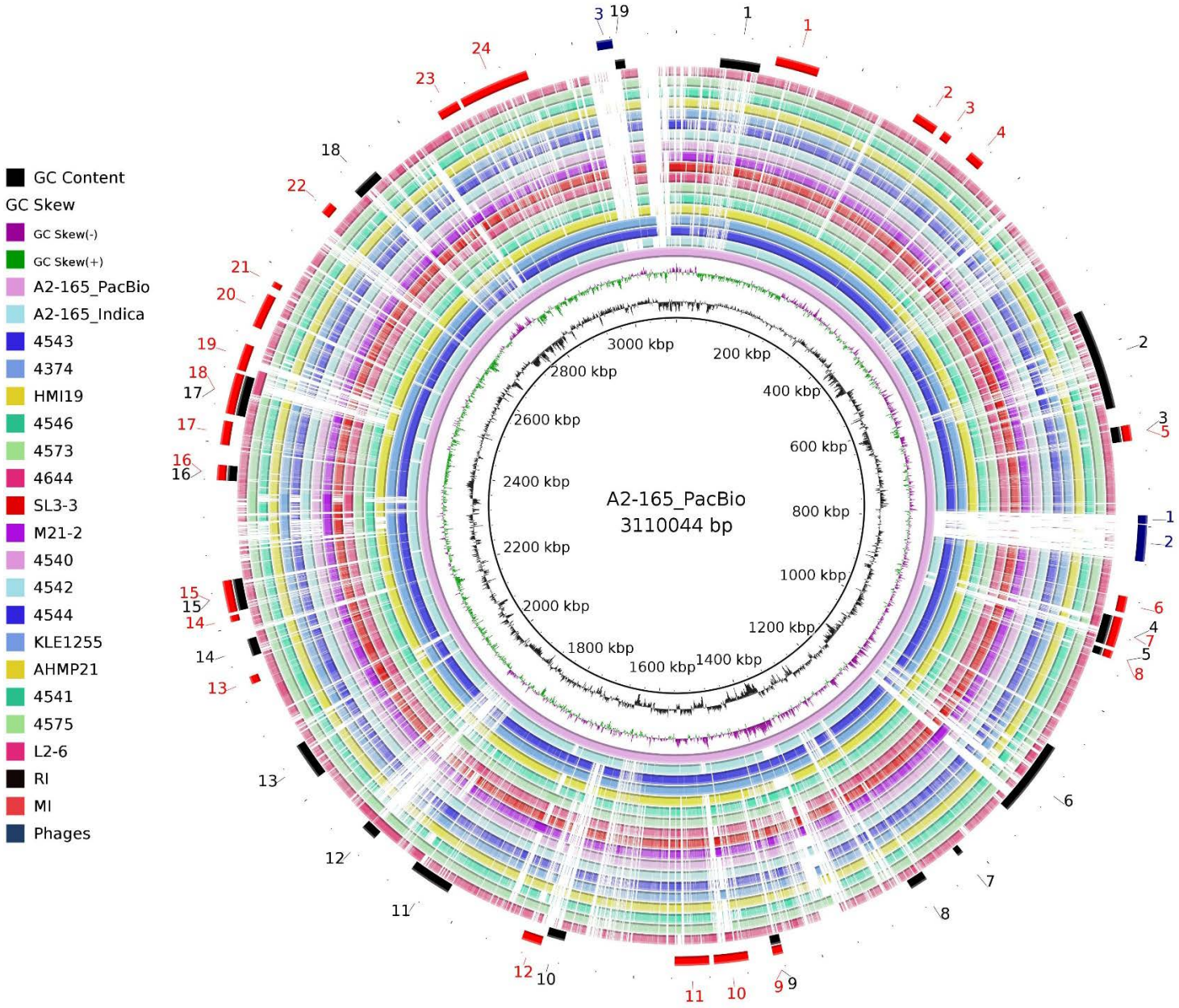

Figure 1 - Visualization of genome comparison using BRIG software. A BLASTn analysis was performed for each strain using the F. prausnitzii A2-165_PacBio as reference genome. The innermost circles represent the genome of Faecalibacterium strains. The outermost circles represent the genomic islands predicted to F. prausnitzii A2-165_PacBio and the probable phage regions. MI: Metabolic Island (red); RI: Antibiotic Resistance Island (black); Phages (navy).

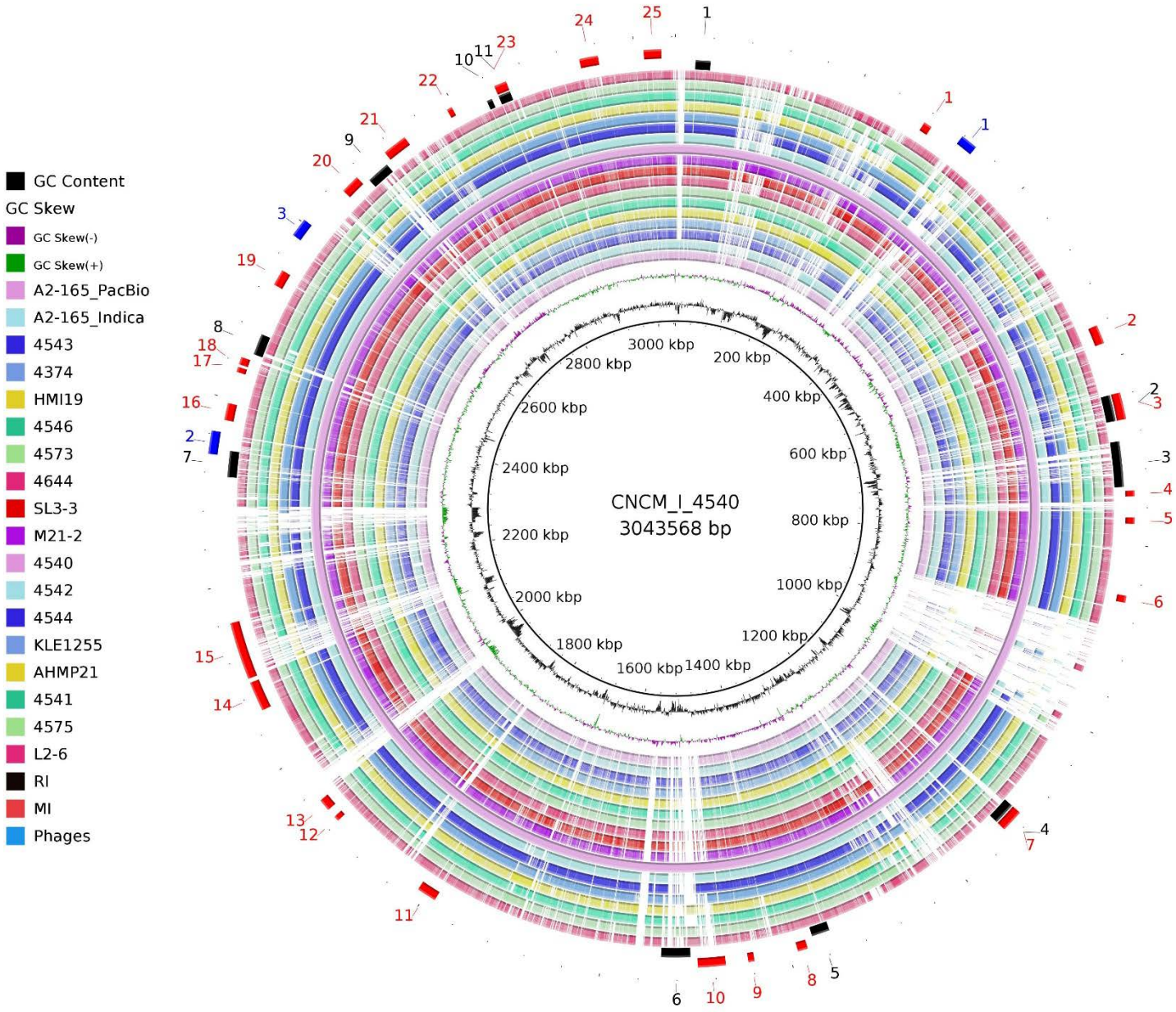

Figure 2 - Visualization of genome comparison using BRIG software. A BLASTn analysis was performed for each strain using the F. prausnitzii CNCM_I_4540 as reference genome. The innermost circles represent the genome of Faecalibacterium strains. The outermost circles represent the genomic islands predicted to F. prausnitzii CNCM_I_4540 and the probable phage regions. MI: Metabolic Island (red); RI: Antibiotic Resistance Island (black); Phages (navy).

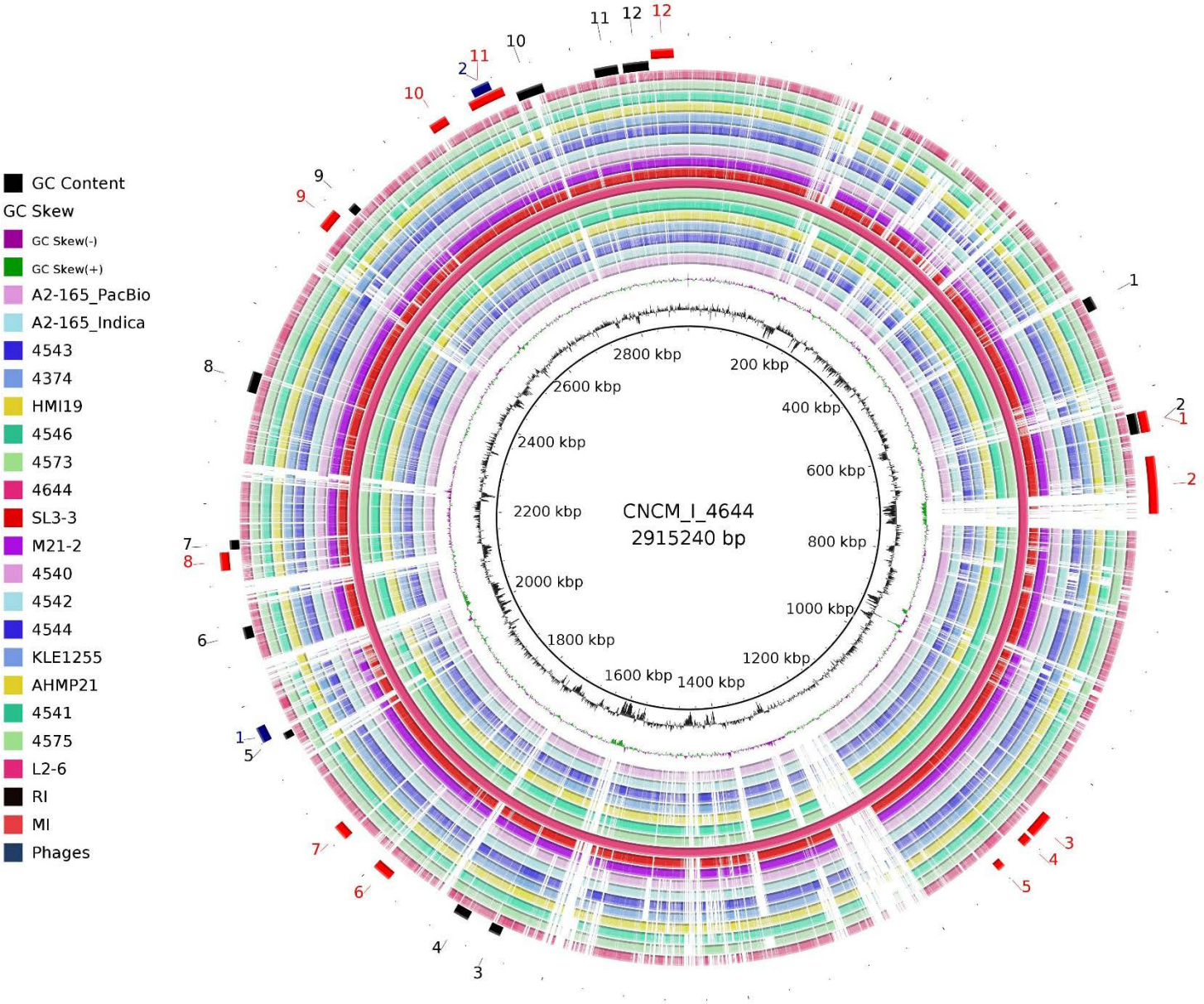

Figure 3 - Visualization of genome comparison using BRIG software. A BLASTn analysis was performed for each strain using the F. prausnitzii CNCM_I_4644 as reference genome. The innermost circles represent the genome of Faecalibacterium strains. The outermost circles represent the genomic islands predicted to F. prausnitzii CNCM_I_4644 and the probable phage regions. MI: Metabolic Island (red); RI: Antibiotic Resistance Island (black); Phages (navy).

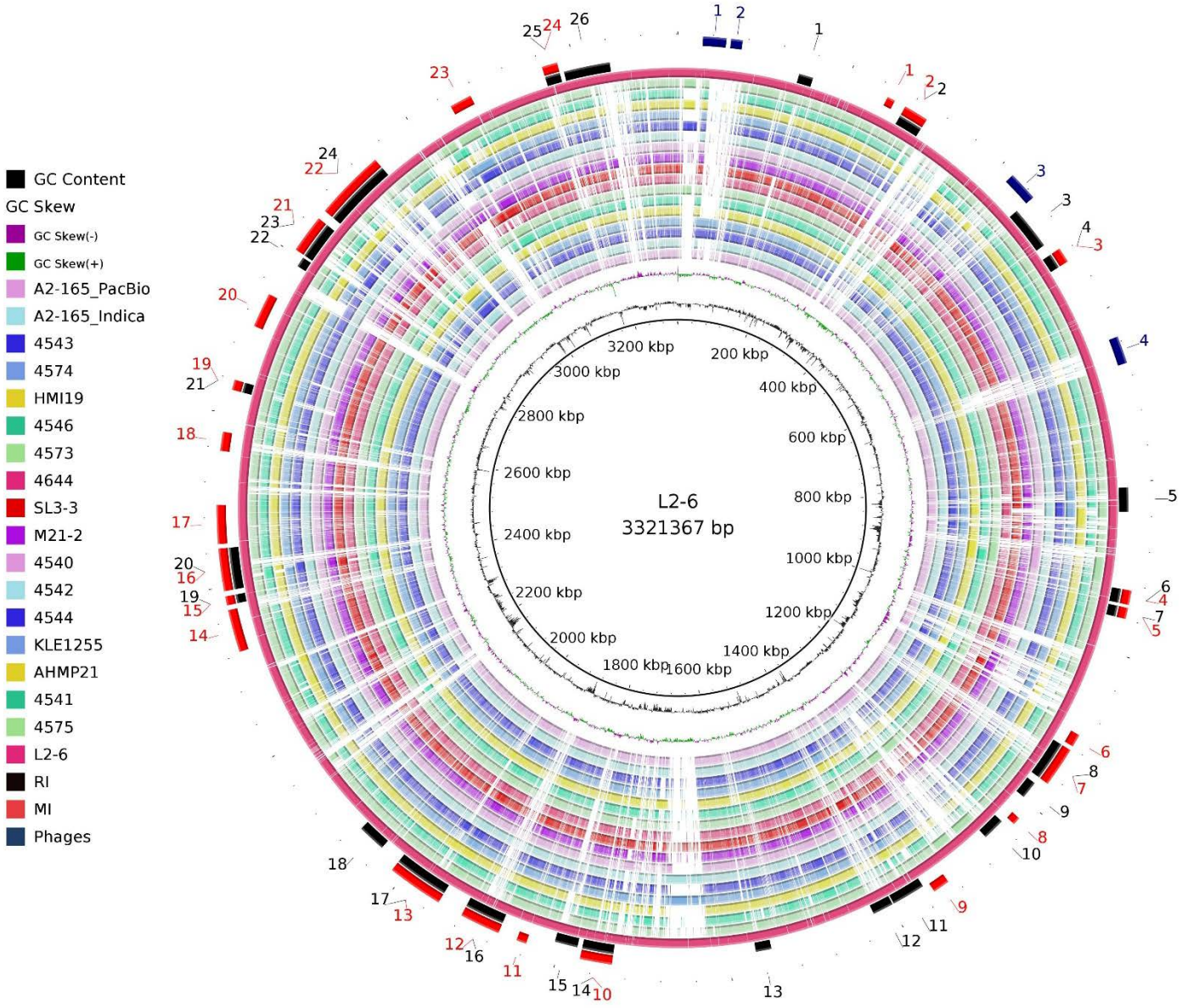

Figure 4 - Visualization of genome comparison using BRIG software. A BLASTn analysis was performed for each strain using the F. prausnitzii L2-6 as reference genome. The innermost circles represent the genome of Faecalibacterium strains. The outermost circles represent the genomic islands predicted to F. prausnitzii L2-6 and the probable phage regions. MI: Metabolic Island (red); RI: Antibiotic Resistance Island (black); Phages (navy).

Phage regions prediction

A summary with the number of predicted phage regions is listed in Table 2. The figures 1-4 show the predicted prophage regions for the genomes *F. prausnitzii* A2-165_PacBio, *F. prausnitzii* CNCM_I_4540, *F. prausnitzii* CNCM_I_4644 and *F. prausnitzii* L2-6, respectively.

Table 2 - Number of predicted phage regions.

Strains	Genospecies	Intact (score > 90)	Questionable (score 70-90)	Incomplete (score < 70)
CNCM_I_4541	-	0	0	0
CNCM_I_4575	-	1	0	0
AHMP-21	-	2	0	1
M21-2	-	1	0	0
CNCM_I_4546	A	0	0	7
CNCM_I_4573	A	1	0	0
CNCM_I_4644	A	0	1	1
L2-6	A	2	0	2
SL3-3	A	1	0	1
CNCM_I_4543	B	2	0	2
CNCM_I_4574	B	2	1	5
A2-165_PacBio	B	1	1	2
Indica	B	2	0	0
HMI-19	B	1	0	0
CNCM_I_4540	C	0	0	3
CNCM_I_4542	C	1	1	3
CNCM_I_4544	C	1	0	2
KLE1255	C	1	2	1

Metabolic pathways prediction

The software PATRIC predicted 126 metabolic pathways (MP) among the genomes. Of these, 108 MP are found in all genomes (supplementary Table 1). A summary of the MP not shared by all genomes can be found in Table 3.

Table 3 - Predicted metabolic pathways that are not common to all genomes.

Pathway Name	Unique Genome Count	Unique EC Count	Genomes with pathway	EC predicted
Glycosphingolipid biosynthesis - globo series	17	3	Genospecies A: 4546, 4573, 4644, M21-2 and SL3-3; Genospecies B: , A2-165_PacBio, Indica, 4543, 4574 and HMI-19; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541, 4575, AHMP21	2.4.1.-; 3.2.1.22; 3.2.1.52
Retinol metabolism	16	3	Genospecies A: 4546, 4573, 4644 and M21-2; Genospecies B: A2-165_PacBio, Indica, 4543, 4574 and HMI-19; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541, 4575, AHMP21	2.4.1-; 2.4.99
Glycosphingolipid biosynthesis - lacto and neolacto series	16	2	Genospecies A: 4546, 4573, 4644 and M21-2; Genospecies B: , A2-165_PacBio, Indica, 4543, 4574 and HMI-19; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541, 4575, AHMP21	1.1.1.-; 1.1.1.1; 3.1.1.-
T cell receptor signaling pathway	13	2	Genospecies A: 4546, 4573, 4644 and M21-2; Genospecies B: A2-165_PacBio, 4543 and 4574; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541, 4575	2.7.10.2; 3.1.3.16
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	13	1	Genospecies A: 4546, 4573, 4644 and M21-2; Genospecies B: Indica and HMI-19; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541, 4575, AHMP21	2.4.1.-
O-Glycan biosynthesis	13	1	Genospecies A: 4546, 4573, 4644 and M21-2;	2.4.1.-

			Genospecies B: Indica and HMI-19; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541, 4575, AHMP21	
Steroid biosynthesis	12	1	Genospecies A: 4546, 4573, 4644 and M21-2; Genospecies B: A2-165_PacBio, 4543 and 4574; Genospecies C: 4540, 4542 and 4544; 4541, 4575	5.-.-.-
Tetrachloroethene degradation	12	3	Genospecies A: 4546, 4573, M21-2 and SL3-3; Genospecies B: 4543 and 4574; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541, L2-6	1.1.1.-; 1.1.1.1; 1.2.1.3
Bisphenol A degradation	12	2	Genospecies A: 4546, 4573, 4644, M21-2 and SL3-3; Genospecies B: A2-165_PacBio, Indica, 4543, 4574 and HMI-19; 4541, L2-6	1.1.1.-; 3.1.1.-
Drug metabolism - cytochrome P450	9	1	Genospecies A: 4546, 4573, M21-2 and SL3-3; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541	1.1.1.1
Metabolism of xenobiotics by cytochrome P450	9	1	Genospecies A: 4546, 4573, M21-2 and SL3-3; Genospecies C: 4540, 4542, 4544 and KLE1255; 4541	1.1.1.1
Toluene and xylene degradation	6	1	Genospecies A: 4546, 4573, M21-2 and SL3-3; 4541, L2-6	1.1.1.-
C21-Steroid hormone metabolism	6	1	Genospecies A: 4546, 4573, M21-2 and SL3-3; 4541, L2-6	1.1.1.-
Linoleic acid metabolism	6	1	Genospecies A: 4546, 4573, M21-2 and SL3-3; 4541, L2-6	1.1.1.-
Primary bile acid biosynthesis	6	2	Genospecies A: 4573; Genospecies B: 4543, 4574 and A2-165_PacBio; Genospecies C: 4542; AHMP21	1.1.1.35; 3.5.1.24
N-Glycan biosynthesis	5	1	Genospecies A: 4546, 4644, 4573 and M21-2; 4541	2.4.1.83
Secondary bile acid biosynthesis	5	1	Genospecies A: 4573; Genospecies B: 4543, 4574 and A2-165_PAcBio; AHMP21	3.5.1.24
Biosynthesis of unsaturated fatty acids	1	1	4540	3.1.2.2

Antibiotic resistance genes

Among the genomes, we found 8 genes related to antibiotic resistance mechanisms considering 2 databases: CARD (Jia et al., 2017) and ARDB (Liu and Pop, 2009). Additionally, we checked the G+C Content and Codon Usage of these genes and their presence in resistance islands (Table 4) and tried to correlate with the antibiogram profile proposed by Martín et al. (2017) (Martín et al., 2017), which used 12 different antibiotics. Among the genomes, 6 genes of Tetracycline resistance, ribosomal protection type => Tet(W) were found. Of these one has G+C Content deviation, five have a Codon usage deviation and four are present in resistance islands (RI). Five genes of Aminoglycoside N(3)-acetyltransferase were found among the genomes: Two with G+C Content deviation, three with Codon Usage deviation and one is present in a RI. Seventeen genes of SSU ribosomal protein S12p (S23e) were found: Two with G+C Content deviation, two with Codon Usage deviation and one of these is in a RI. Seven genes of 23S rRNA (adenine(2085)-N(6))-dimethyltransferase were found: one with G+C Content deviation, two with Codon Usage deviation and two are present in a RI. Twenty three genes that encodes an ABC transporter, ATP-binding protein were found: five with G+C Content deviation, twelve with Codon Usage deviation and fifteen of these are present in a RI. Five genes that encodes a Vancomycin response regulator were found: none of them have G+C Content or Codon Usage deviation and three are present in a RI. Three genes that encodes a Chloramphenicol acetyltransferase were found: none of them have G+C Content deviation, two have Codon Usage deviation and one is present in a RI. One gene that encodes a putative aminoglycoside 6-adenylyltansferase was found. This gene has a G+C Content and Codon Usage deviation and it is present in a RI (Table 4).

Table 4 - Genes putatively coding for antibiotic resistance-related proteins.

Query ID	G+C Content	Codon Usage	Presence in RI
Tetracycline resistance, ribosomal protection type => Tet(W)			
F_prausnitzii_M21_2|853_72_peg_191	NORMAL	DEVIATION	YES

F_prausnitzii_CNCM_4541|853.63.peg.260	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4644|853.71.peg.1614e	LOW	DEVIATION	YES
F_prausnitzii_L2_6|718252.3.peg.1682	NORMAL	DEVIATION	YES
F_prausnitzii_CNCM_4542|853.64.peg.2996	NORMAL	DEVIATION	NO
F_prausnitzii_CNCM_4546|853_67_peg_632	NORMAL	DEVIATION	YES
Aminoglycoside N(3)-acetyltransferase			
F_prausnitzii_M21_2|853_72_peg_99	NORMAL	NORMAL	NO
F_prausnitzii_SL3_3|657322.3.peg.185	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4573|853.68.peg.503	LOW	DEVIATION	NO
F_prausnitzii_CNCM_4644|853.71.peg.515	LOW	DEVIATION	YES
F_prausnitzii_CNCM_4546|853.67.peg.285	NORMAL	DEVIATION	NO
SSU ribosomal protein S12p (S23e)			
F_prausnitzii_A2_165_PacBio|853_73_peg_2960	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4542|853_64_peg_2405	LOW	DEVIATION	NO
F_prausnitzii_CNCM_4575|853_70_peg_189	NORMAL	DEVIATION	NO
F_prausnitzii_HMI_19|FPR_HMI_19_0234	NORMAL	NORMAL	NO
F_prausnitzii_SL3_3|657322_3_peg_2842	NORMAL	NORMAL	NO
F_prausnitzii_KLE1255|853_59_peg_2630	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4546|853_67_peg_3064	HIGH	NORMAL	NO
F_prausnitzii_CNCM_4573|853_68_peg_3199	NORMAL	NORMAL	NO
F_prausnitzii_AHMP_21|FPR_AHMP_21_0323c	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4644|853_71_peg_2644	NORMAL	NORMAL	NO
F_prausnitzii_L2_6|718252_3_peg_50	NORMAL	NORMAL	YES
F_prausnitzii_CNCM_4544|853_66_peg_1628	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4543|853_65_peg_3029	NORMAL	NORMAL	NO
F_prausnitzii_M21_2|853_72_peg_2883	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4540|853_62_peg_2807	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4541|853_63_peg_2686	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4574|853_69_peg_2988	NORMAL	NORMAL	NO
23S rRNA (adenine(2085)-N(6))-dimethyltransferase (EC 2.1.1.184)			
F_prausnitzii_HMI_19|FPR_HMI_19_2029c	LOW	DEVIATION	YES
F_prausnitzii_CNCM_4542|853_64_peg_1310	NORMAL	NORMAL	YES
F_prausnitzii_CNCM_4575|853_70_peg_887	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4546|853_67_peg_1486	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4573|853_68_peg_599	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4540|853_62_peg_1645	NORMAL	DEVIATION	NO
F_prausnitzii_CNCM_4574|853_69_peg_2956	NORMAL	NORMAL	NO
ABC transporter, ATP-binding protein			
F_prausnitzii_CNCM_4541|853_63_peg_170	NORMAL	NORMAL	YES

F_prausnitzii_CNCM_4574|853_69_peg_1928	NORMAL	NORMAL	YES
F_prausnitzii_M21_2|853_72_peg_2431	LOW	DEVIATION	YES
F_prausnitzii_CNCM_4543|853_65_peg_35	LOW	DEVIATION	YES
F_prausnitzii_CNCM_4544|853_66_peg_252	NORMAL	DEVIATION	YES
F_prausnitzii_L2_6|718252_3_peg_948	NORMAL	NORMAL	YES
F_prausnitzii_CNCM_4644|853_71_peg_3000	NORMAL	NORMAL	YES
F_prausnitzii_AHMP_21|FPR_AHMP_21_2879	NORMAL	DEVIATION	YES
F_prausnitzii_CNCM_4573|853_68_peg_378	NORMAL	DEVIATION	NO
F_prausnitzii_CNCM_4546|853_67_peg_477	NORMAL	NORMAL	YES
F_prausnitzii_KLE1255|853_59_peg_138	NORMAL	DEVIATION	YES
F_prausnitzii_SL3_3|657322_3_peg_652	NORMAL	NORMAL	YES
F_prausnitzii_HMI_19|FPR_HMI_19_1164	LOW	DEVIATION	YES
F_prausnitzii_CNCM_4575|853_70_peg_1754	NORMAL	DEVIATION	YES
F_prausnitzii_A2_165_PacBio|853_73_peg_680	NORMAL	DEVIATION	YES
F_prausnitzii_CNCM_4542|853_64_peg_81	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4543|853_65_peg_2023	NORMAL	NORMAL	YES
F_prausnitzii_CNCM_4542|853_64_peg_2143	NORMAL	DEVIATION	NO
F_prausnitzii_HMI_19|FPR_HMI_19_1969	LOW	DEVIATION	NO
F_prausnitzii_KLE1255|853_59_peg_1734	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4544|853_66_peg_2672	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4540|853_62_peg_2391	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4574|853_69_peg_975	LOW	DEVIATION	NO
Vancomycin response regulator VanR			
F_prausnitzii_CNCM_4543|853_65_peg_2392	NORMAL	NORMAL	YES
F_prausnitzii_CNCM_4541|853_63_peg_1200	NORMAL	NORMAL	NO
F_prausnitzii_CNCM_4540|853_62_peg_1989	NORMAL	NORMAL	YES
F_prausnitzii_CNCM_4544|853_66_peg_103	NORMAL	NORMAL	YES
F_prausnitzii_A2_165_PacBio|853_73_peg_933	NORMAL	NORMAL	NO
Chloramphenicol acetyltransferase (EC 2.3.1.28)			
F_prausnitzii_CNCM_4546|853_67_peg_2624	NORMAL	DEVIATION	NO
F_prausnitzii_AHMP_21|FPR_AHMP_21_1629c	NORMAL	DEVIATION	YES
F_prausnitzii_CNCM_4541|853_63_peg_2635	NORMAL	NORMAL	NO
putative aminoglycoside 6-adenylyltansferase			
F_prausnitzii_A2_165_PacBio|853_73_peg_676	LOW	DEVIATION	YES

Bacteriocin prediction

No bacteriocin was found in any genome using the BAGEL4 software. In view of this, we analyzed only the Microbial Anti-inflammatory Molecule (MAM), which present bacteriocin-like activity. The Genomic environment of the anti-inflammatory protein MAM is shown in Figure 5. MAM is a protein produced by *F.prausnitzii* which has been described as anti-inflammatory *in vitro* and *in vivo* (Breyner et al., 2017; Quévrain et al., 2014). In all strains, MAM gene is associated with genes encoding for proteins related to sporulation, SpoIIGA, SigE, SigG and Stage III sporulation protein D (numbers 3, 4, 5 and 6 respectively in the Figure 5). Moreover, in strains from group B the protein MAM is linked to the gene encoding the ABC-type bacteriocin transporter protein (number 2 in the Figure 5).

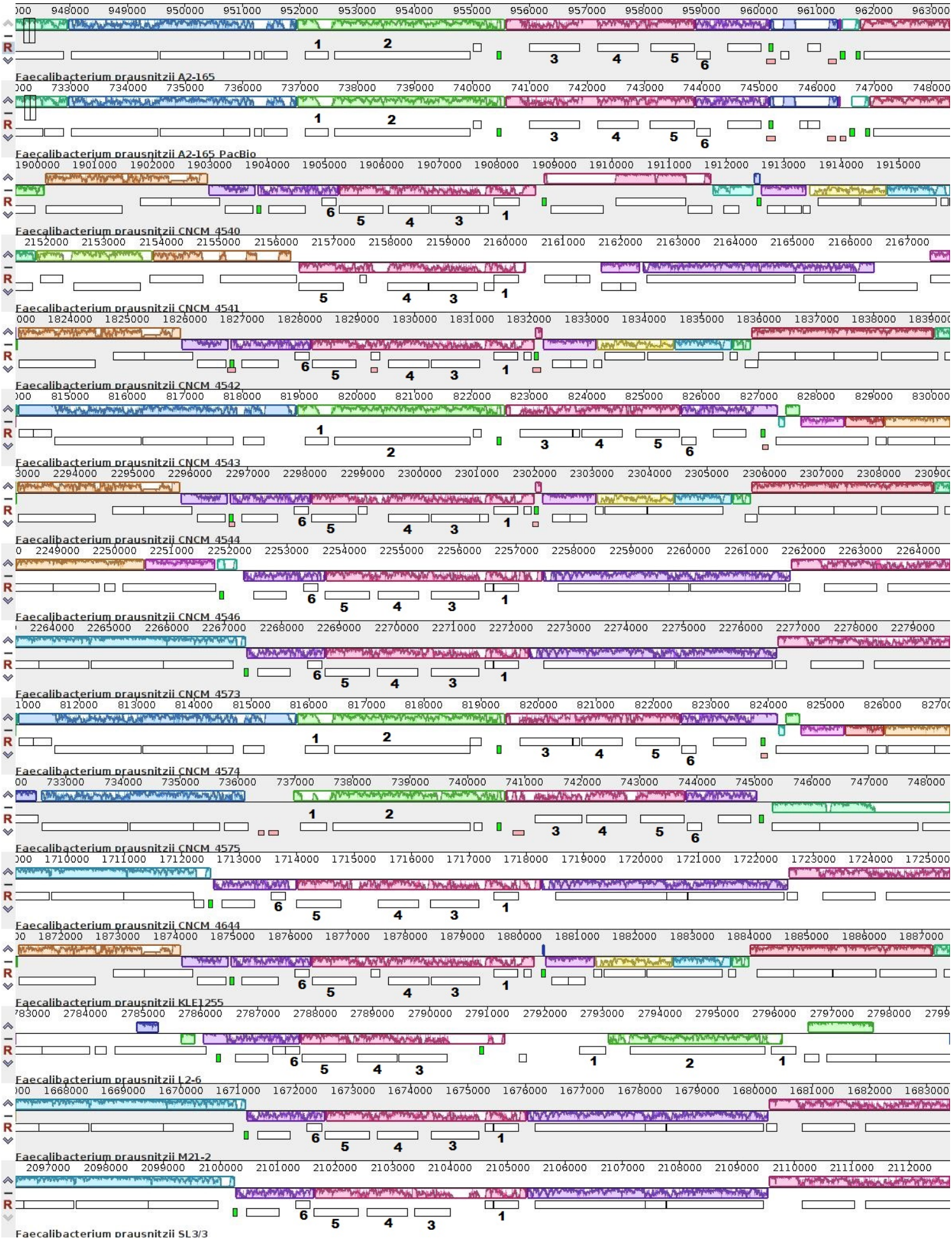

Figure 5 - The alignment display is organized into one horizontal "panel" per input genome sequence. Each genome's panel contains the name of the genome sequence, a scale showing the sequence coordinates for that genome, and a single black horizontal center line. Colored block (LCB) appear above and possibly below the center line. Each of these block surrounds a region of the genome sequence that aligned to part of another genome, and is presumably homologous and internally free from genomic rearrangement. Sequences outside colored blocks do not have homologs in the other genomes. The numbers indicate the following CDS: (1) MAM, (2) ABC-type bacteriocin transporter, (3) sporulation factor SpoIIGA, (4) sporulation specific sigma factor SigE, (5) sporulation specific sigma factor SigG, (6) Stage III sporulation protein D.

DISCUSSION

Prediction of horizontally acquired regions

To provide a better understanding about the genome plasticity of *Faecalibacterium* species, especially horizontally acquired regions, we have predicted putative phage regions and genomic islands of this genus. To investigate these horizontally acquired regions is very important in a safety assessment study because the potential probiotic strains can serve as a reservoir of virulence factors and antibiotic resistance genes (Oliveira et al., 2017). In this regard, the presence of these regions, and their potential mobility, deserves special attention.

In Figure 1, the Genomic Islands (GI) predicted for the strain A2-165_PacBio are shown. This strain belongs to the same genospecies that the strains CNCM_I_4543, CNCM_I_4574 and HMI-19 (Benevides et al., 2017). The result from BRIG software shows the regions that are present in the reference genome (in this case, the genome A2-165_PacBio) and absent in other genomes. The same is observed for the GI and phage regions predicted. We could not identify any island completely and exclusive to the genome of A2-165_PacBio or the other strains from the same genospecies, but in some cases we could observed that part of a GI or even other regions (some genes) are exclusive to one or more genomes. For example, the Resistance Island (RI) number 10 have some genes that are only found in the strain A2-165_PacBio and Indica. These genes can be related to characteristics unique to these strains, like mechanisms of antibiotic resistance. The phages number 1 and 2 are exclusive of this strain and the phage number 3 is only found in the genomes A2-165_PacBio, CNCM_I_4543 and CNCM_I_4574, which belongs to the same genospecies (Benevides et al., 2017). The phages 1 and 2 are classified as “incomplete” and “Intact”, respectively. These phages were predicted as 2 different regions, but other study (Cornuault et al., 2018) showed that this region corresponds to one phage called “Lagaffe” (GenBank: MG711461.1). The phage number 3 was classified as “questionable”, it is

16.8kb long and it partially overlaps the RI number 19. This phage was also described by Cornuault et al. (2018) as “Mushu” (GenBank: MG711460.1) and has 36,636 bps in length. After performing BLAST between the genomes of phage “Mushu” and A2-165_PacBio, we observed that the gapped area surrounding the Phage 3 corresponds to the phage “Mushu”, therefore the Phage 3 predicted here is part of the phage “Mushu” predicted by Cornuault et al. (2018) and it can be found in other strains of this genospecies (CNCM_I_4543 and CNCM_I_4574).

Figure 2 shows the BRIG representation of the genomes using the strain CNCM_I_4540, from the same genospecies of CNCM_I_4542, CNCM_I_4544 and KLE1255. Although many island and phage regions were identified in strain CNCM_I_4540, none of them is completely exclusive to its genome or the other strains from the same genospecies. However, in some cases, part of a GI or even other regions (some genes) is exclusive to one or more genomes. Here, we identified one region, between the positions ~850 kbp and ~1070 kbp that is almost completely limited to this strain. In this region, there are ~320 genes and, among them, 43 have associated EC numbers that will be further discussed in the appropriate section.

The GI and phage regions predicted to the strain CNCM_I_4546 are shown in figure 3. This strain belongs to the same genospecies that the strains CNCM_I_4573, CNCM_I_4644, M21-2 and SL3-3. 7 incomplete phage regions were predicted on this strain, some of them being partially exclusive to this strain or to other strains of the same genospecies (i.e Phages 2 and 4) and the same happens to the GI. Even without any exclusive horizontally acquired region, the similarity among the genomes of the same genospecies is higher than the similarity with the other genomes. It probably occurred because these probable horizontally acquired regions were incorporated to the genomes of the genus *Faecalibacterium* a long time ago in its phylogenetic history, resulting in different similarity percentages, but not removing these regions or acquiring new ones.

The strain *F. prausnitzii* L2-6 is one of the most distantly related strains of this genus (Benevides et al., 2017). In its genome, various regions are exclusive or present low similarity to the other *Faecalibacterium* genomes. For example, the

RI 8 and RI 16, the MI 7, MI 12 and MI 18 and the Phages 1, 2 and 4. These GI can be related to metabolic capacities or antibiotic resistance mechanisms unique to this strain or shared with some others. The Phages 1 and 2 were described by Cornuault et al. (2018) as one unique region, the phage “Taranis” (GenBank: MG711467.1). This phage is complete in the genomes of strains *F. prausnitzii* L2-6, *F. prausnitzii* CNCM_I_4543 and *F. prausnitzii* CNCM_I_4574. The Phage 4 corresponds to the phage “Lugh” and it is significantly more abundant in IBD patients than in healthy individuals (Cornuault et al., 2018).

Antibiotic resistance genes

The European Food Safety Authority (Binnendijk and Rijkers, 2013; Ricci et al., 2018) proposed that the nature of any antibiotic resistance determinant present in a candidate microorganism should be determined prior to approval for QPS status. Therefore, antibiotic resistance *per se* is not a safety issue; it only becomes such when the risk of resistance transfer is present.

Here, we have also investigated the genes related to antibiotic resistance in all *Faecalibacterium* strains. One of the genes found was the "Tetracycline resistance, ribosomal protection type Tet(W)” that encodes a ribosomal protection protein, which protects ribosome from the translation inhibition of tetracycline. It is associated with both conjugative and non-conjugative DNA and has been found in other genus like *Clostridium, Bifidobacterium, Lactobacillus* and *Streptococcus* (Jia et al., 2017; Liu and Pop, 2009; Scott et al., 2000). Here, 6 strains have this gene: M21_2, CNCM_4546 and CNCM_4644 from genospecies A, L2_6, CNCM_4542 and CNCM_4541. In the strains M21_2, CNCM_4546, L2-6 and CNCM_I_4644 the genes are inside Resistance Islands (RI). Of these, only L2-6, CNCM_4541 and CNCM_4546 demonstrated *in vitro* resistance to tetracycline.

The CDS encoding the “aminoglycoside N(3)-acetyltransferase” was found again in all the strains from group A (M21_2, CNCM_4546, CNCM_4644, SL3-3 and CNCM_I_4573). Of these, the only strain for which the gene is included in a RI is CNCM_I_4644. This protein is a chromosomal-encoded aminoglycoside

acetyltransferase that can confer resistance to gentamicin, streptomycin and other antibiotics (Ishikawa et al., 2000; Jia et al., 2017; Liu and Pop, 2009). Of these strains, M21_2, CNCM_4546, CNCM_4644 and CNCM_I_4573 showed *in vitro* resistance to gentamicin. The strain SL3-3 was not evaluated in antibiogram assays *in vitro*, yet (Martín et al., 2017).

Another protein that can confer resistance to streptomycin is the “SSU ribosomal protein S12p (S23e) “. This protein stabilizes the highly conserved pseudoknot structure formed by 16S rRNA. Some amino acid substitutions in its sequence may affect the higher-order structure of 16S rRNA and confer streptomycin resistance by disrupting interactions between 16S rRNA and streptomycin (Ballif et al., 2012; Jia et al., 2017; Liu and Pop, 2009). All genomes evaluated here have this gene, but *in vitro*, the strain CNCM_I_4575 did not demonstrate resistance to streptomycin. Furthermore, the strain A2-165_PacBio also present another gene related to streptomycin resistance: “aminoglycoside 6-adenylyltansferase”. This second gene evolved in streptomycin resistance is included in a RI and may explain the high Minimum Inhibitory Concentration evidenced by Martín et al. (2017).

One gene related to erythromycin resistance (ermB - 23S rRNA (adenine(2085)-N(6))-dimethyltransferase) was found in 7 genomes: CNCM_4546 and CNCM_4573 (genospecies A), HMI_19 and CNCM_4574 (genospecies B), CNCM_4540 and CNCM_4542 (genospecies C) and CNCM_4575. This gene can methylate adenine at a specific position of 23S rRNA, conferring resistance to erythromycin (Liu and Pop, 2009). Its expression is inducible by erythromycin. When erythromycin is present, it binds the leader peptide causing a change in conformation allowing for the expression of ErmB (Jia et al., 2017). Despite the presence of this gene in 7 genomes, none of them showed *in vitro* resistance. Meanwhile, the strain L2-6 have *in vitro* resistance to erythromycin (Martín et al., 2017), but does not present this gene, suggesting that other resistance mechanism should be involved.

The “ABC transporter, ATP-binding protein” evidenced in all genomes, is related to macrolides resistance. The gene *macB* observed here, encodes ATP-binding cassette (ABC) transporter that exports macrolides forming an antibiotic efflux complex with the proteins encoded by *macA* and *tolC* (Jia et al., 2017; Liu

and Pop, 2009). Despite the presence of this gene in all genomes, *in vitro* tests showed that only the strain CNCM_I_4541 is resistant to erythromycin, a macrolide type antibiotic, suggesting the absence of other genes necessary to cause resistance.

The gene *vanA*, that encodes “Vancomycin response regulator”, is one of the genes necessary to promote the resistance to vancomycin antibiotic. This gene belongs to an operon that can synthesize peptidoglycan with modified C-terminal D-Ala-D-Ala to D-alanine--D-lactate. Here, we observed this gene in 5 genomes: CNCM_4543 and A2_165_PacBio (genospecies B), CNCM_4540 and CNCM_4544 (genospecies C) and CNCM_4541. None of these strains present resistance to vancomycin, probably because the other genes from the operon are missing.

One gene that can confer resistance to Chloramphenicol was observed in 3 genomes: CNCM_I_4546, AHMP-21 and CNCM_I_4541. The “Chloramphenicol acetyltransferase (EC 2.3.1.28)” belongs to the Group A chloramphenicol acetyltransferase, which can inactivate chloramphenicol. The *in vitro* analysis demonstrated, among the strains tested, that only the strain L2-6 has resistance to this antibiotic (Martín et al., 2017). As we did not observe the gene in its genome, other mechanism should be used by this strain.

Metabolic pathways prediction

The metabolic pathways of all genomes were predicted to better understand its metabolic activities and metabolic diversity among the strains. We predicted genes related to 126 distinct metabolic pathways and, of these pathways, 18 are not shared by all genomes (table 3). Among the 18 pathways, 5 are related to Glycan Biosynthesis and Metabolism, 6 to Lipid Metabolism, 5 to Xenobiotics Biodegradation and Metabolism, 1 to Immune System and 1 to Metabolism of Cofactors and Vitamins.

From those pathways, some are further highlighted taking into account their absence in one or two of the genospecies proposed by Benevides et al. (2017)

like the “Bisphenol A degradation” metabolic pathway. The Bisphenol A (BPA) is a toxic monomer used in the manufacture of polycarbonate plastics, food cans, and other daily used chemicals (Zhang et al., 2013). BPA possesses estrogenic and genotoxic effects on human and other organisms. It could interact with the estrogen receptor and affect the reproductive behavior (vom Saal et al., 1999). There is evidence that some bacteria like *Bifidobacterium breve* and *Lactobacillus casei*, found in foods, might help to protect against BPA exposure (Solouki et al., 2018). Here, 12 genomes present genes involved in this pathway and could be tested for this important activity.

One well known characteristic of *Faecalibacterium spp* is its ability to produce butyrate (Duncan et al., 2002; Lopez-Siles et al., 2012; Martín et al., 2017). Butyrate has anti-inflammatory properties and is one of the molecules produced by *F.prausnitzii* explaining the protective effects of this bacterium in colitis model (Sokol et al., 2008). Butyrate is synthesized via pyruvate and acetyl-coenzyme A (CoA), mostly by the breakdown of complex polysaccharides that reach the colon (Louis and Flint, 2009; Vital et al., 2014). Here, were found genes involved in the “Butanoate metabolism” pathway in all genomes showing that all strains could transform the Pyruvate in Acetoacetyl-CoA and this one in (S)-3-Hydroxybutanoyl-CoA. The other genes necessary to complete this metabolic pathway were not found in all genomes, but accordingly to Vital et al. (2014): “direct functional predictions based on gene homology alone can commonly result in misannotations if genes with distinct function share regions of high similarity, as specifically described for both *but* and *buk.*”. Furthermore, the CoA transferases show activity with several different substrate combinations, and alternative terminal CoA transferases were proposed for this pathway (Eeckhaut et al., 2011).

Another important pathway predicted is the “Streptomycin biosynthesis” that was found in all strains. Martín et al. (2017) showed that only the strain CNCM_I_4575 is not resistant to this antibiotic, but there are no studies about the production of resistance proteins against this antibiotic by *Faecalibacterium* strains. In addition to streptomycin, the metabolic pathways related to the biosynthesis of tetracycline and vancomycin were found in all genomes. As the

only bacteriocin found in the genomes was the MAM protein, the possibility of antibiotic compounds being produced by these strains should be investigated.

Bacteriocin prediction

Probiotic strains often produce an array of antimicrobial compounds like bacteriocins. In *in vitro* assays performed by Martín et al (2017), they did not show antibacterial activities to the strains analyzed and here, the prediction of bacteriocins did not find any candidate. Nevertheless, our group identified, in other studies, a protein that have characteristics similar to bacteriocins: the Microbial Anti-Inflammatory Molecule (MAM) protein (Breyner et al., 2017; Quévrain et al., 2014). After analyzing the gene synteny of MAM protein we observed that in strains from group A the gene appear close to ABC-type bacteriocin transporter protein (Figure 5). As these proteins are not close together in all genomes, we first checked if the evolution of ABC-type bacteriocin transporter protein is similar to the full genomes phylogeny and MAM phylogeny. The MAM protein and ABC-type bacteriocin transporter have similar phylogenetics clusters structure (Supplementary Figures 1-2, respectively) and this structure is similar to the full genome cluster structure shown by Benevides et al. (2017). These results suggest that the two proteins had similar evolutionary processes and are not involved in recent lateral gene transfer. After that phylogenetic analysis, we checked, by BLASTp, the existence of ABC-type bacteriocin transporter protein in other species and checked its structure and conserved domains in NCBI Conserved Domains Database (CDD; https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Different from MAM, this protein can be found in other species like *Pseudobutyrivibrio sp*, *Butyrivibrio sp*, *Eubacterium rectale*, *Anaerobium acetethylicum*, *Clostridium clariflavum*, *Clostridium cellulovorans* and others.

The ABC-type bacteriocin transporter protein has three different conserved domains. The N-terminal domain belongs to the peptidase sub-family C39B, a sub-family of peptidase family C39. Peptidase family C39 mostly contains bacteriocin-processing endopeptidases from bacteria. The cysteine peptidases in family C39 cleave the "double-glycine" leader peptides from the precursors of various bacteriocins (mostly non-lantibiotic). The cleavage is mediated by the

transporter as part of the secretion process. Bacteriocins are antibiotic proteins secreted by some species of bacteria that inhibit the growth of other bacterial species. The bacteriocin is synthesized as a precursor with an N-terminal leader peptide, and processing involves removal of the leader peptide by cleavage at a Gly-Gly bond, followed by translocation of the mature bacteriocin across the cytoplasmic membrane. Most endopeptidases of family C39 are N-terminal domains in larger proteins (ABC transporters) that serve both functions. The proposed protease active site is conserved in this sub-family.

In the alignment of MAM protein sequences (Supplementary Figure 3), we can observe a Gly-Gly site at position 20 conserved in all sequences. The N-terminal part of these sequences is highly conserved suggesting that it could be a leader peptide.

CONCLUDING REMARKS

Faecalibacterium prausnitzii has been recognized as an important bacterium promoting the intestinal health and today is considered as a potential next generation probiotic (NGP) (Martín et al., 2018). The isolation and new proposition of reclassification into this genus (Benevides et al., 2017; Martín et al., 2017), increases the importance of newly and carefully performed studies of all strains to better understand the diversity, the interactions with the host and the safety aspects in its use as probiotic. Here we evaluated, *in silico*, the main safety aspects in order to perform a general screening, considering many of the species in this genus.

We identified Metabolic Islands, Antibiotic Resistance Islands and Phage regions, but none of them are exclusive to one strain or genospecies. These findings suggest that these regions were horizontally acquired by ancestral genomes and they are now integrated to these genomes, decreasing the probability of horizontal transference to other species.

The search for genes related to antibiotic resistance in all *Faecalibacterium* strains shows the existence of 8 genes related to resistance mechanism, but

the presence/absence of these genes inside RI seems not to interfere in the mechanisms of antibiotic resistance in the studied genomes. The antibiotic resistance *per se* is not a safety issue and it only becomes such when the risk of resistance transfer is present (Gueimonde et al., 2013). As the potential Genomic Islands predicted here have not been recently acquired and are already integrated to the genomes, we believe that none of the *Faecalibacterium* species have a high potential to transfer antibiotic resistance genes to other bacterial species.

Concerning the predicted metabolic pathways, we highlighted here some that should be investigated *in vitro*. First, the possibility of some strains to degrade the toxic compound Bisphenol A is a new beneficial capacity of these potential NGP that deserves attention; and second, the metabolic pathways related to the biosynthesis of antibiotics like streptomycin, tetracycline and vancomycin may open a new understanding about the competitive relationship among *Faecalibacterium* strains and other bacterial species.

Based on our analyses about MAM, we start to consider the possibility of its protein to be a bacteriocin. Recent *in vitro* assays performed by our group showed that the supernatant of a culture of strain A2-165 inhibit its own growth (JM Chatel, personal communication) supporting the idea that MAM could be a bacteriocin. More *in vitro* analyzes are being performed to test this hypothesis. Up to date, little was known about *Faecalibacterium spp* genomic characteristics, especially those related to probiotic features. Altogether, this work presents a step forward into the understanding of probiotic characteristics of *Faecalibacterium* species. Here, we did not identify any harmful aspect to the use of these species as probiotic.

REFERENCES

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12, 402. doi:10.1186/1471-2164-12-402.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16. doi:10.1093/nar/gkw387.

Ballif, M., Harino, P., Ley, S., Coscolla, M., Niemann, S., Carter, R., et al. (2012). Drug resistance-conferring mutations in Mycobacterium tuberculosis from Madang, Papua New Guinea. *BMC Microbiol.* 12, 191. doi:10.1186/1471-2180-12-191.

Barcenilla, A., Pryde, S. E., Martin, J. C., Duncan, S. H., Stewart, C. S., Henderson, C., et al. (2000). Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* 66, 1654–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10742256 [Accessed April 16, 2018].

Benevides, L., Burman, S., Martin, R., Robert, V., Thomas, M., Miquel, S., et al. (2017). New Insights into the Diversity of the Genus Faecalibacterium. *Front. Microbiol.* 8, 1790. doi:10.3389/fmicb.2017.01790.

Binnendijk, K. H., and Rijkers, G. T. (2013). What is a health benefit? An evaluation of EFSA opinions on health benefits with reference to probiotics. *Benef. Microbes* 4, 223–230. doi:10.3920/BM2013.0019.

Böcker, U., Nebe, T., Herweck, F., Holt, L., Panja, A., Jobin, C., et al. (2003). Butyrate modulates intestinal epithelial cell-mediated neutrophil migration. *Clin. Exp. Immunol.* 131, 53–60. doi:10.1046/J.1365-2249.2003.02056.X.

Breyner, N. M., Michon, C., de Sousa, C. S., Vilas Boas, P. B., Chain, F., Azevedo, V. A., et al. (2017). Microbial Anti-Inflammatory Molecule (MAM) from Faecalibacterium prausnitzii Shows a Protective Effect on DNBS and DSS-Induced Colitis Model in Mice through Inhibition of NF-κB Pathway.

Front. Microbiol. 8, 114. doi:10.3389/fmicb.2017.00114.

Cornuault, J. K., Petit, M.-A., Mariadassou, M., Benevides, L., Moncaut, E., Langella, P., et al. (2018). Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes. *Microbiome* 6, 65. doi:10.1186/s40168-018-0452-1.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–403. doi:10.1101/gr.2289704.

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., and Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13, 601–612. doi:10.1038/nrg3226.

Duncan, S. H., Hold, G. L., Harmsen, H. J. M., Stewart, C. S., and Flint, H. J. (2002). Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 52, 2141–6. doi:10.1099/00207713-52-6-2141.

Eeckhaut, V., Van Immerseel, F., Croubels, S., De Baere, S., Haesebrouck, F., Ducatelle, R., et al. (2011). Butyrate production in phylogenetically diverse Firmicutes isolated from the chicken caecum. *Microb. Biotechnol.* 4, 503–512. doi:10.1111/j.1751-7915.2010.00244.x.

Gueimonde, M., Sánchez, B., G. de los Reyes-Gavilán, C., and Margolles, A. (2013). Antibiotic resistance in probiotic bacteria. *Front. Microbiol.* 4, 202. doi:10.3389/fmicb.2013.00202.

Hippe, B., Remely, M., Aumueller, E., Pointner, A., Magnet, U., and Haslberger, A. G. (2016). Faecalibacterium prausnitzii phylotypes in type two diabetic, obese, and lean control subjects. *Benef. Microbes* 7, 511–517. doi:10.3920/BM2015.0075.

Ishikawa, J., Sunada, A., Oyama, R., and Hotta, K. (2000). Identification and characterization of the point mutation which affects the transcription level of the chromosomal 3-N-acetyltransferase gene of Streptomyces griseus SS-1198. *Antimicrob. Agents Chemother.* 44, 437–40. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/10639379 [Accessed March 28, 2018].

J.B. Prajapati and Suja Senan (2013). Omics: a novel tool for safety assessment of probiotic strains. *Asian J. Food Agro-Industry* 06, 208–221. Available at: http://www.ajofai.info/Abstract/Omics - a novel tool for safety assessment of probiotic strains.pdf [Accessed April 27, 2016].

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi:10.1093/nar/gkw1004.

Liu, B., and Pop, M. (2009). ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Res.* 37, D443–D447. doi:10.1093/nar/gkn656.

Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J., and Martinez-Medina, M. (2017). Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. *ISME J.* 11, 841–852. doi:10.1038/ismej.2016.176.

Lopez-Siles, M., Khan, T. M., Duncan, S. H., Harmsen, H. J. M., Garcia-Gil, L. J., and Flint, H. J. (2012). Cultured representatives of two major phylogroups of human colonic Faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl. Environ. Microbiol.* 78, 420–428. doi:10.1128/AEM.06858-11.

Lopez-Siles, M., Martinez-Medina, M., Surís-Valls, R., Aldeguer, X., Sabat-Mir, M., Duncan, S. H., et al. (2016). Changes in the Abundance of Faecalibacterium prausnitzii Phylogroups I and II in the Intestinal Mucosa of Inflammatory Bowel Disease and Patients with Colorectal Cancer. *Inflamm. Bowel Dis.* 22, 28–41. doi:10.1097/MIB.0000000000000590.

Louis, P., and Flint, H. J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol. Lett.* 294, 1–8. doi:10.1111/j.1574-6968.2009.01514.x.

Macfarlane, G. T., and Macfarlane, S. (2011). Fermentation in the Human Large Intestine. *J. Clin. Gastroenterol.* 45, S120–S127. doi:10.1097/MCG.0b013e31822fecfe.

Martín, R., Chain, F., Miquel, S., Lu, J., Gratadoux, J.-J., Sokol, H., et al. (2014). The Commensal Bacterium Faecalibacterium prausnitzii Is Protective in DNBS-induced Chronic Moderate and Severe Colitis Models. *Inflamm. Bowel Dis.* 20, 417–430. doi:10.1097/01.MIB.0000440815.76627.64.

Martín, R., Miquel, S., Benevides, L., Bridonneau, C., Robert, V., Hudault, S., et al. (2017). Functional Characterization of Novel Faecalibacterium prausnitzii Strains Isolated from Healthy Volunteers: A Step Forward in the Use of F. prausnitzii as a Next-Generation Probiotic. *Front. Microbiol.* 8, 1226. doi:10.3389/fmicb.2017.01226.

Martín, R., Miquel, S., Chain, F., Natividad, J. M., Jury, J., Lu, J., et al. (2015). Faecalibacterium prausnitzii prevents physiological damages in a chronic low-grade inflammation murine model. *BMC Microbiol.* 15, 67. doi:10.1186/s12866-015-0400-1.

Oliveira, L. C., Saraiva, T. D. L., Silva, W. M., Pereira, U. P., Campos, B. C., Benevides, L. J., et al. (2017). Analyses of the probiotic property and stress resistance-related genes of Lactococcus lactis subsp. lactis NCDO 2118 through comparative genomics and in vitro assays. *PLoS One* 12, e0175116. doi:10.1371/journal.pone.0175116.

Papadimitriou, K., Zoumpopoulou, G., Foligné, B., Alexandraki, V., Kazou, M., Pot, B., et al. (2015). Discovering probiotic microorganisms: in vitro, in vivo, genetic and omics approaches. *Front. Microbiol.* 6, 58. doi:10.3389/fmicb.2015.00058.

Pineiro, M., and Stanton, C. (2007). Probiotic bacteria: legislative framework--requirements to evidence basis. *J. Nutr.* 137, 850S–3S. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17311986 [Accessed March 12, 2018].

Quévrain, E., Maubert, M. A., Michon, C., Chain, F., Marquant, R., Tailhades, J., et al. (2014). Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii , a commensal bacterium deficient in Crohn's disease. *Gut* 65, 415–425. doi:10.1136/gutjnl-2014-307649.

Ricci, A., Allende, A., Bolton, D., Chemaly, M., Davies, R., Girones, R., et al. (2018). Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 7: suitability of taxonomic units notified to EFSA until September 2017. *EFSA J.* 16. doi:10.2903/j.efsa.2018.5131.

Scott, K. P., Melville, C. M., Barbosa, T. M., and Flint, H. J. (2000). Occurrence of the new tetracycline resistance gene tet(W) in bacteria from the human gut. *Antimicrob. Agents Chemother.* 44, 775–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10681357 [Accessed March 28, 2018].

Senan, S., Prajapati, J. B., and Joshi, C. G. (2015). Feasibility of Genome-Wide Screening for Biosafety Assessment of Probiotics: A Case Study of Lactobacillus helveticus MTCC 5463. *Probiotics Antimicrob. Proteins* 7, 249–258. doi:10.1007/s12602-015-9199-1.

Soares, S. C., Geyik, H., Ramos, R. T. J., de Sá, P. H. C. G., Barbosa, E. G. V, Baumbach, J., et al. (2015). GIPSy: Genomic island prediction software. *J. Biotechnol.* doi:10.1016/j.jbiotec.2015.09.008.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J.-J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci.* 105, 16731–16736. doi:10.1073/pnas.0804812105.

Solouki, S., Fazeli, M. R., and Solouki, S. (2018). Efficiency of Multispecies Probiotic Supplements in Bioremoval of Bisphenol A: An In Vitro Study. *Appl. Food Biotechnol.* 5, 37–45. doi:10.22037/afb.v5i1.18191.

Song, H., Yoo, Y., Hwang, J., Na, Y.-C., and Kim, H. S. (2016). Faecalibacterium prausnitzii subspecies–level dysbiosis in the human gut microbiome underlying atopic dermatitis. *J. Allergy Clin. Immunol.* 137, 852–860. doi:10.1016/j.jaci.2015.08.021.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-P., et al. (2009). Towards the human intestinal microbiota phylogenetic core.

Environ. Microbiol. 11, 2574–84. doi:10.1111/j.1462-2920.2009.01982.x.

van Heel, A. J., de Jong, A., Montalbán-López, M., Kok, J., and Kuipers, O. P. (2013). BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* 41, W448-53. doi:10.1093/nar/gkt391.

Vital, M., Howe, A. C., and Tiedje, J. M. (2014). Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)genomic Data. *MBio* 5, e00889-14-e00889-14. doi:10.1128/mBio.00889-14.

vom Saal, F. S., Howdeshell, K. L., Hotchkiss, A. K., Thayer, K. A., and Vandenbergh, J. G. (1999). Environmental toxins: Exposure to bisphenol A advances puberty. *Nature* 401, 763–764. doi:10.1038/44517.

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., et al. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* 5, 220–30. doi:10.1038/ismej.2010.118.

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* 45, D535–D542. doi:10.1093/nar/gkw1017.

Zhang, W., Yin, K., and Chen, L. (2013). Bacteria-mediated bisphenol A degradation. *Appl. Microbiol. Biotechnol.* 97, 5681–5689. doi:10.1007/s00253-013-4949-z.

Supplementary Files

Table S1 – Predicted metabolic pathways

Pathway Name	Pathway Class	Unique Genome Count	Unique Gene Count	Unique EC Count	EC Conservation
Glycolysis / Gluconeogenesis	Carbohydrate Metabolism	18	451	18	80.56
Citrate cycle (TCA cycle)	Carbohydrate Metabolism	18	162	9	57.41
Pentose phosphate pathway	Carbohydrate	18	343	16	92.01

	Metabolism				
Pentose and glucuronate interconversions	Carbohydrate Metabolism	18	384	22	70.2
Fructose and mannose metabolism	Carbohydrate Metabolism	18	399	16	70.83
Galactose metabolism	Carbohydrate Metabolism	18	503	17	68.3
Ascorbate and aldarate metabolism	Carbohydrate Metabolism	18	164	8	37.5
Fatty acid biosynthesis	Lipid Metabolism	18	266	9	89.51
Fatty acid elongation in mitochondria	Lipid Metabolism	18	22	4	30.56
Fatty acid metabolism	Lipid Metabolism	18	92	8	50.69
Synthesis and degradation of ketone bodies	Lipid Metabolism	18	19	2	52.78
Ubiquinone and other terpenoid-quinone biosynthesis	Metabolism of Cofactors and Vitamins	18	96	2	91.67
Oxidative phosphorylation	Energy Metabolism	18	372	7	66.67
Photosynthesis	Energy Metabolism	18	270	1	100
Purine metabolism	Nucleotide Metabolism	18	958	40	92.5
Puromycin biosynthesis	Biosynthesis of Secondary Metabolites	18	48	1	100
Pyrimidine metabolism	Nucleotide Metabolism	18	852	33	85.02
Alanine, aspartate and glutamate metabolism	Amino Acid Metabolism	18	441	17	88.24
Tetracycline biosynthesis	Biosynthesis of Secondary Metabolites	18	169	4	83.33
Glycine, serine and threonine metabolism	Amino Acid Metabolism	18	314	22	63.89
Cysteine and methionine metabolism	Amino Acid Metabolism	18	444	21	85.71
Valine, leucine and isoleucine degradation	Amino Acid Metabolism	18	67	10	46.67
Geraniol degradation	Xenobiotics Biodegradation and Metabolism	18	21	3	38.89
Valine, leucine and isoleucine biosynthesis	Amino Acid Metabolism	18	203	9	90.12
Lysine biosynthesis	Amino Acid Metabolism	18	183	12	84.26
Lysine degradation	Amino Acid Metabolism	18	96	9	46.3
Arginine and proline metabolism	Amino Acid Metabolism	18	532	27	84.36
Histidine metabolism	Amino Acid Metabolism	18	234	14	73.41
Tyrosine metabolism	Amino Acid Metabolism	18	206	6	90.74
Phenylalanine metabolism	Amino Acid Metabolism	18	211	8	76.39
Benzoate degradation via hydroxylation	Xenobiotics Biodegradation and Metabolism	18	87	7	69.84

Fluorobenzoate degradation	Xenobiotics Biodegradation and Metabolism	18	18	1	100
Tryptophan metabolism	Amino Acid Metabolism	18	101	6	38.89
Phenylalanine, tyrosine and tryptophan biosynthesis	Amino Acid Metabolism	18	286	19	74.56
Novobiocin biosynthesis	Biosynthesis of Secondary Metabolites	18	107	5	94.44
beta-Alanine metabolism	Metabolism of Other Amino Acids	18	27	4	37.5
Taurine and hypotaurine metabolism	Metabolism of Other Amino Acids	18	38	2	100
Phosphonate and phosphinate metabolism	Metabolism of Other Amino Acids	18	62	3	40.74
Selenoamino acid metabolism	Metabolism of Other Amino Acids	18	179	6	84.26
Cyanoamino acid metabolism	Metabolism of Other Amino Acids	18	84	4	83.33
D-Glutamine and D-glutamate metabolism	Metabolism of Other Amino Acids	18	55	3	100
D-Alanine metabolism	Metabolism of Other Amino Acids	18	37	2	100
Glutathione metabolism	Metabolism of Other Amino Acids	18	77	6	61.11
Starch and sucrose metabolism	Carbohydrate Metabolism	18	539	25	67.33
High-mannose type N-glycan biosynthesis	Glycan Biosynthesis and Metabolism	18	89	3	59.26
Amino sugar and nucleotide sugar metabolism	Carbohydrate Metabolism	18	776	32	73.26
Streptomycin biosynthesis	Biosynthesis of Secondary Metabolites	18	123	6	87.04
Biosynthesis of 12-, 14- and 16-membered macrolides	Biosynthesis of Polyketides and Nonribosomal Peptides	18	110	2	86.11
Polyketide sugar unit biosynthesis	Biosynthesis of Polyketides and Nonribosomal Peptides	18	109	5	86.67
Glycosaminoglycan degradation	Glycan Biosynthesis and Metabolism	18	161	4	68.06
Lipopolysaccharide biosynthesis	Glycan Biosynthesis and Metabolism	18	94	4	66.67
Peptidoglycan biosynthesis	Glycan Biosynthesis and Metabolism	18	447	13	96.15
Glycerolipid metabolism	Lipid Metabolism	18	128	7	65.08
Inositol phosphate metabolism	Carbohydrate Metabolism	18	38	4	44.44
Glycerophospholipid metabolism	Lipid Metabolism	18	279	14	71.43
Ether lipid metabolism	Lipid Metabolism	18	26	1	100
alpha-Linolenic acid metabolism	Lipid Metabolism	18	68	3	70.37
Sphingolipid metabolism	Lipid Metabolism	18	252	5	90
Glycosphingolipid biosynthesis - ganglio series	Glycan Biosynthesis and Metabolism	18	195	5	67.78

Pyruvate metabolism	Carbohydrate Metabolism	18	398	22	72.47
2,4-Dichlorobenzoate degradation	Xenobiotics Biodegradation and Metabolism	18	74	4	52.78
1- and 2-Methylnaphthalene degradation	Xenobiotics Biodegradation and Metabolism	18	88	2	80.56
Naphthalene and anthracene degradation	Xenobiotics Biodegradation and Metabolism	18	41	2	75
1,4-Dichlorobenzene degradation	Xenobiotics Biodegradation and Metabolism	18	133	6	60.19
Glyoxylate and dicarboxylate metabolism	Carbohydrate Metabolism	18	276	18	68.21
Trinitrotoluene degradation	Xenobiotics Biodegradation and Metabolism	18	61	3	48.15
Propanoate metabolism	Carbohydrate Metabolism	18	155	11	67.17
Ethylbenzene degradation	Xenobiotics Biodegradation and Metabolism	18	44	2	100
Butanoate metabolism	Carbohydrate Metabolism	18	199	14	54.76
C5-Branched dibasic acid metabolism	Carbohydrate Metabolism	18	37	1	100
One carbon pool by folate	Metabolism of Cofactors and Vitamins	18	218	14	82.94
Methane metabolism	Energy Metabolism	18	542	14	92.86
Carbon fixation in photosynthetic organisms	Energy Metabolism	18	303	13	92.74
Reductive carboxylate cycle (CO2 fixation)	Energy Metabolism	18	317	17	83.01
Thiamine metabolism	Metabolism of Cofactors and Vitamins	18	154	8	88.19
Riboflavin metabolism	Metabolism of Cofactors and Vitamins	18	62	8	70.83
Vitamin B6 metabolism	Metabolism of Cofactors and Vitamins	18	61	4	83.33
Nicotinate and nicotinamide metabolism	Metabolism of Cofactors and Vitamins	18	221	12	75
Pantothenate and CoA biosynthesis	Metabolism of Cofactors and Vitamins	18	301	12	99.54
Biotin metabolism	Metabolism of Cofactors and Vitamins	18	222	9	97.53
Folate biosynthesis	Metabolism of Cofactors and Vitamins	18	87	5	81.11
Porphyrin and chlorophyll metabolism	Metabolism of Cofactors and	18	417	24	80.79

	Vitamins				
Terpenoid backbone biosynthesis	Biosynthesis of Secondary Metabolites	18	201	13	77.35
Limonene and pinene degradation	Biosynthesis of Secondary Metabolites	18	30	3	40.74
Diterpenoid biosynthesis	Biosynthesis of Secondary Metabolites	18	26	1	100
Carotenoid biosynthesis	Biosynthesis of Secondary Metabolites	18	59	2	86.11
Zeatin biosynthesis	Biosynthesis of Secondary Metabolites	18	51	2	86.11
Nitrogen metabolism	Energy Metabolism	18	194	10	78.33
Sulfur metabolism	Energy Metabolism	18	119	4	100
Caprolactam degradation	Xenobiotics Biodegradation and Metabolism	18	22	4	30.56
Phenylpropanoid biosynthesis	Biosynthesis of Secondary Metabolites	18	114	2	88.89
Flavonoid biosynthesis	Biosynthesis of Secondary Metabolites	18	113	3	62.96
Anthocyanin biosynthesis	Biosynthesis of Secondary Metabolites	18	136	3	90.74
Flavone and flavonol biosynthesis	Biosynthesis of Secondary Metabolites	18	50	2	77.78
Stilbenoid, diarylheptanoid and gingerol biosynthesis	Biosynthesis of Secondary Metabolites	18	136	3	90.74
Isoquinoline alkaloid biosynthesis	Biosynthesis of Secondary Metabolites	18	233	6	88.89
Tropane, piperidine and pyridine alkaloid biosynthesis	Biosynthesis of Secondary Metabolites	18	138	6	87.96
Betalain biosynthesis	Biosynthesis of Secondary Metabolites	18	59	2	86.11
Aminoacyl-tRNA biosynthesis	Translation	18	413	23	92.27
Insect hormone biosynthesis	Biosynthesis of Secondary Metabolites	18	136	3	77.78
Drug metabolism - other enzymes	Xenobiotics Biodegradation and Metabolism	18	201	10	87.22
Biosynthesis of ansamycins	Biosynthesis of Polyketides and Nonribosomal Peptides	18	152	4	73.61
Biosynthesis of siderophore group nonribosomal peptides	Biosynthesis of Polyketides and	18	191	2	100

	Nonribosomal Peptides				
Biosynthesis of vancomycin group antibiotics	Biosynthesis of Polyketides and Nonribosomal Peptides	18	21	1	100
Biosynthesis of type II polyketide backbone	Biosynthesis of Polyketides and Nonribosomal Peptides	18	37	2	66.67
Biosynthesis of type II polyketide products	Biosynthesis of Polyketides and Nonribosomal Peptides	18	169	4	76.39
Phosphatidylinositol signaling system	Signal Transduction	18	22	2	61.11
mTOR signaling pathway	Signal Transduction	18	18	1	100
Glycosphingolipid biosynthesis - globo series	Glycan Biosynthesis and Metabolism	17	56	3	54.9
Glycosphingolipid biosynthesis - lacto and neolacto series	Glycan Biosynthesis and Metabolism	16	44	2	75
Retinol metabolism	Metabolism of Cofactors and Vitamins	16	32	3	43.75
T cell receptor signaling pathway	Immune System	14	20	2	71.43
O-Glycan biosynthesis	Glycan Biosynthesis and Metabolism	13	33	1	100
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	Glycan Biosynthesis and Metabolism	13	33	1	100
Steroid biosynthesis	Lipid Metabolism	12	12	1	100
Bisphenol A degradation	Xenobiotics Biodegradation and Metabolism	12	17	2	50
Tetrachloroethene degradation	Xenobiotics Biodegradation and Metabolism	12	28	3	47.22
Metabolism of xenobiotics by cytochrome P450	Xenobiotics Biodegradation and Metabolism	9	15	1	100
Drug metabolism - cytochrome P450	Xenobiotics Biodegradation and Metabolism	9	15	1	100
Primary bile acid biosynthesis	Lipid Metabolism	6	6	2	50
C21-Steroid hormone metabolism	Lipid Metabolism	6	11	1	100
Linoleic acid metabolism	Lipid Metabolism	6	11	1	100
Toluene and xylene degradation	Xenobiotics Biodegradation and Metabolism	6	11	1	100
Secondary bile acid biosynthesis	Lipid Metabolism	5	5	1	100
N-Glycan biosynthesis	Glycan Biosynthesis and Metabolism	5	8	1	100
Biosynthesis of unsaturated fatty acids	Lipid Metabolism	1	1	1	100

Figure S1

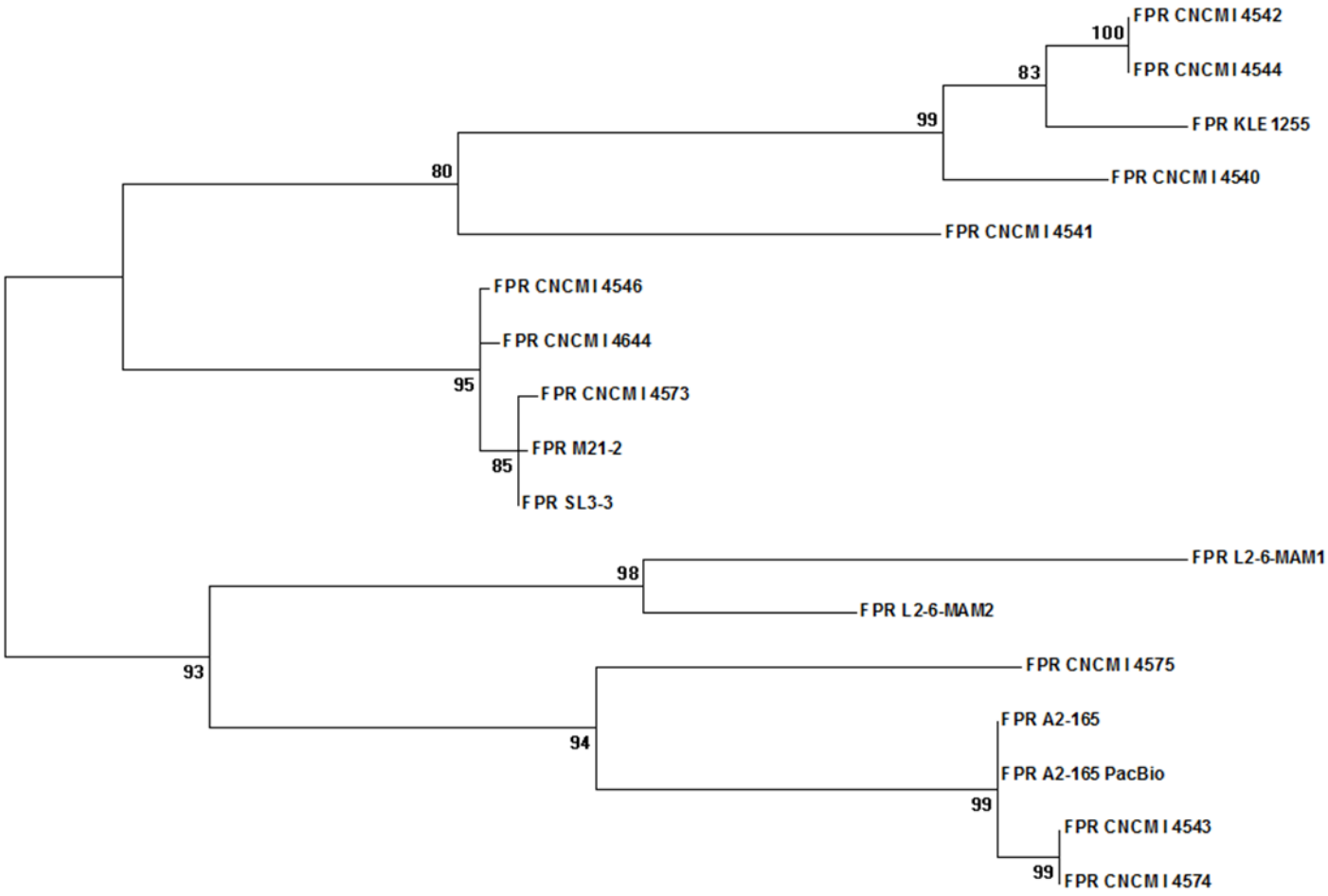

Figure S1 - Molecular Phylogenetic analysis of MAM protein. The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model. The tree with the highest log likelihood (-1876.5456) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.0530)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 17 amino acid sequences. Evolutionary analyses were conducted in MEGA7 software.

Figure S2

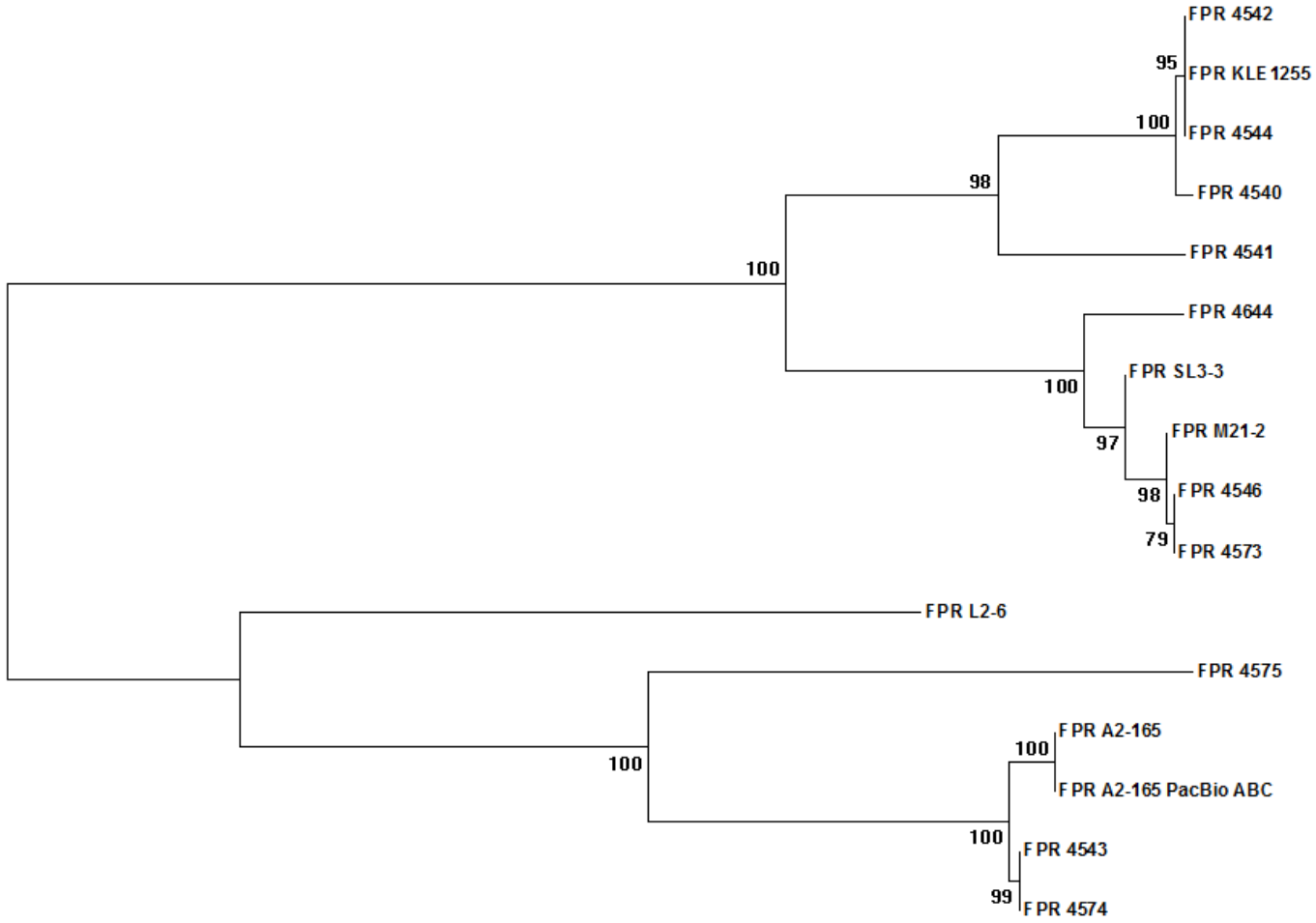

Figure S2 - Molecular Phylogenetic analysis of ABC-type bacteriocin transporter protein. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model. The tree with the highest log likelihood (-4573.2031) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.4510)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. Evolutionary analyses were conducted in MEGA7 software.

Figure S3

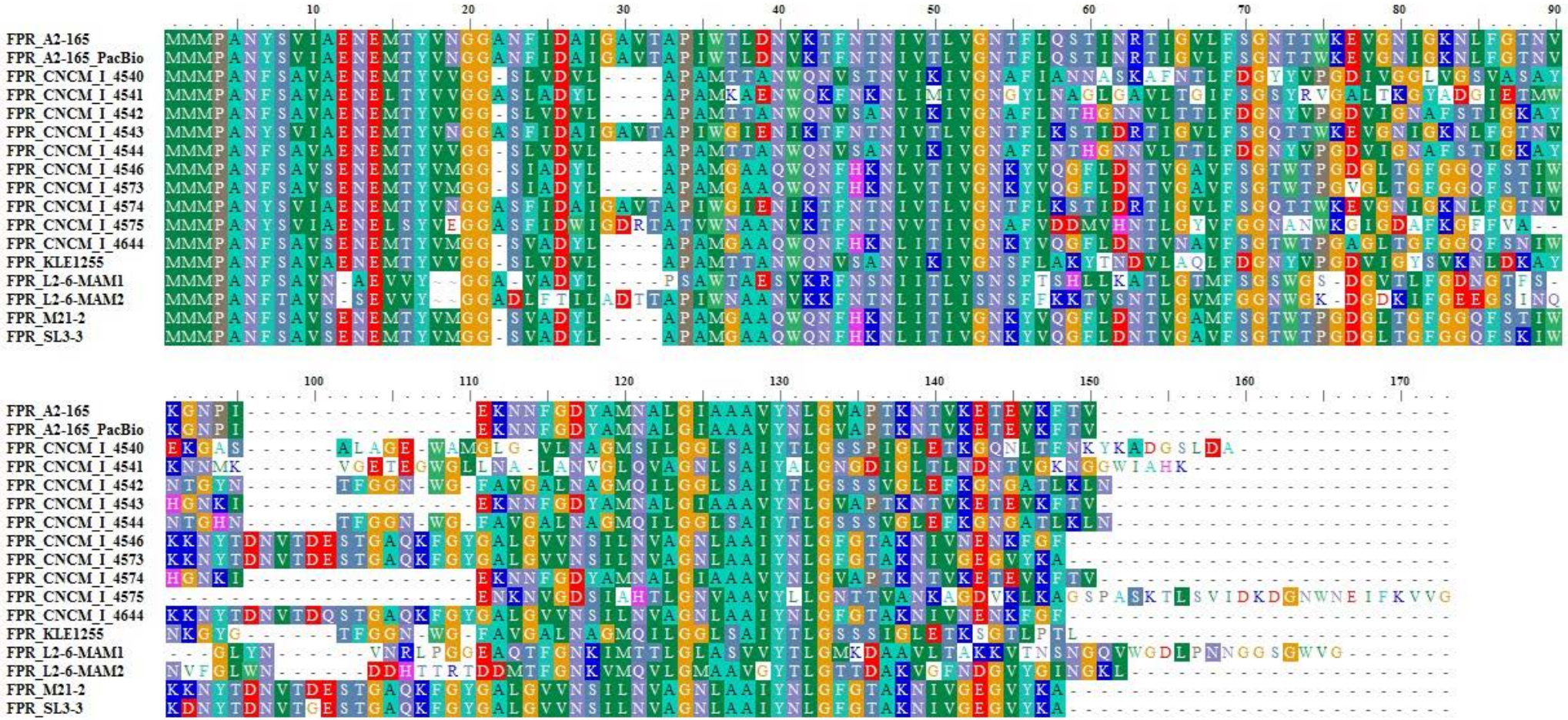

Figure S3 - Sequence alignment of MAM protein from different strains of Faecalibacterium.

VI CONCLUDING REMARKS

The overall aim of the current Ph.D. project was to study the genomics aspects of different strains of *Faecalibacterium* to provide a more global view of the genomic diversity of this genus. Here we evaluated two main aspects to the use of *Faecalibacterium* spp strains as probiotic: First, a precise phylogenetic classification of the available strains; and second, the *in silico* safety assessment.

This work gave us the opportunity to substantially improve the number of sequenced genomes of *Faecalibacterium* spp. deposited in public database with the inclusion of 10 new sequenced and assembled genomes.

In the first chapter we performed a deep phylogenetic analysis into the genus *Faecalibacterium* using different strategies which included: 16S rRNA gene phylogeny, phylogenomics, whole genome Multi-Locus Sequence Typing (wgMLST), Average Nucleotide Identity (ANI), gene synteny, and pangenome.

Among the 17 genomes studied, we could identify 3 clusters of genomes that, according to our analyzes, can be considered as 3 different species of the genus *Faecalibacterium* and other 4 genomes that do not cluster and are probably 4 distinct species of *Faecacelibacterium*. So, our results provide evidence for the phylogenetic reconsideration of this genus, showing that there are diverse species into the genus *Faecalibacterium* and not only the specie *Faecalibacterium prauznitii* proposed before. These findings highlight the need for lineage (strain-based) differentiation of this taxon to better define how specific members might be associated with positive or negative host interactions.

With the isolation and new proposition of reclassification into this genus, in the second chapter we performed a general screening of safety aspects, considering many of the species in this genus and we did not identify any harmful aspect to the use of these species as probiotic.

We have identified metabolic islands, antibiotic resistance islands and phage regions, that were probably horizontally acquired by ancestral genomes and are now integrated to these genomes, decreasing the probability of horizontal transference to other species.

Also, we identified important metabolic pathways, like “Bisphenol A degradation”, “Butanoate metabolism” and “Streptomycin biosynthesis”, that, with the other metabolic pathways, helped us to better understand the physiology of these species.

Were identified 8 genes related to antibiotic resistance mechanism that can help to explain the antibiotic resistance of these strains showed in *in vitro* analyzes.

At least, we studied the genomic context of the MAM protein and showed that this protein has characteristics similar to bacteriocins.

Altogether, this work presented pioneering studies of the genomic aspects to the genus *Faecalibacterium* and provided a step forward into the understanding of the diversity of the genus, its interaction with the host and its important probiotic characteristics. The results shown here open new perspectives to the future studies of these bacteria. Other studies showed that in some diseases the abundance of *Faecalibacterium* varies between the two phylogroups proposed before (Hippe et al., 2016; Lopez-Siles et al., 2016; Song et al., 2016). Then, the existence of distinct species and the genomic features showed here will help to improve our knowledge about how these bacteria can act as a sensor and promoter of human health and what is its rule in intestinal diseases. The studies of lineage-specific variations might not only explain the variable abundances of *Faecalibacterium spp.* linked with adverse health outcomes (e.g., atopic dermatitis, Crohn’s disease, and ulcerative colitis), but also provide new opportunities for the diagnosis and strain-specific treatment of gut inflammation and associated diseases. To that is important to perform more metagenomic analyzes considering various diseases and also to sequence more genomes, specially from sick volunteers, which will facilitate our understanding of the biology of these bacteria.

VII References

Abraham, C., and Cho, J. H. (2006). Functional consequences of NOD2 (CARD15) mutations. *Inflamm. Bowel Dis.* 12, 641–650. doi:10.1097/01.MIB.0000225332.83861.5f.

Abriouel, H., Pérez Montoro, B., Casado Muñoz, M. del C., Knapp, C. W., Gálvez, A., and Benomar, N. (2017). In silico genomic insights into aspects of food safety and defense mechanisms of a potentially probiotic Lactobacillus pentosus MP-10 isolated from brines of naturally fermented Aloreña green table olives. *PLoS One* 12, e0176801. doi:10.1371/journal.pone.0176801.

Agren, J., Sundström, A., Håfström, T., and Segerman, B. (2012). Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One* 7, e39107. doi:10.1371/journal.pone.0039107.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Ananthakrishnan, A. N. (2015). Epidemiology and risk factors for IBD. *Nat. Rev. Gastroenterol. Hepatol.* 12, 205–217. doi:10.1038/nrgastro.2015.34.

Anderson, J. L., Edney, R. J., and Whelan, K. (2012). Systematic review: Faecal microbiota transplantation in the management of inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 36, 503–516. doi:10.1111/j.1365-2036.2012.05220.x.

Arber, W. (2014). Horizontal Gene Transfer among Bacteria and Its Role in Biological Evolution. *Life (Basel, Switzerland)* 4, 217–24. doi:10.3390/life4020217.

Archer, S., Meng, S., Wu, J., Johnson, J., Tang, R., and Hodin, R. (1998). Butyrate inhibits colon carcinoma cell growth through two distinct pathways. *Surgery* 124, 248–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9706145 [Accessed April 16, 2018].

Artis, D. (2008). Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat. Rev. Immunol.* 8, 411–420. doi:10.1038/nri2316.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473,

174–180. doi:10.1038/nature09944.

Balamurugan, R., Janardhan, H. P., George, S., Chittaranjan, S. P., and Ramakrishna, B. S. (2008). Bacterial succession in the colon during childhood and adolescence: molecular studies in a southern Indian village. *Am. J. Clin. Nutr.* 88, 1643–1647. doi:10.3945/ajcn.2008.26511.

Barcenilla, A., Pryde, S. E., Martin, J. C., Duncan, S. H., Stewart, C. S., Henderson, C., et al. (2000). Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* 66, 1654–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10742256 [Accessed April 16, 2018].

Baugher, J. L., and Klaenhammer, T. R. (2011). Invited review: Application of omics tools to understanding probiotic functionality. *J. Dairy Sci.* 94, 4753–65. doi:10.3168/jds.2011-4384.

Benevides, L., Burman, S., Martin, R., Robert, V., Thomas, M., Miquel, S., et al. (2017). New Insights into the Diversity of the Genus Faecalibacterium. *Front. Microbiol.* 8, 1790. doi:10.3389/fmicb.2017.01790.

Bennett, P. M. (2004). Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods Mol. Biol.* 266, 71–113. doi:10.1385/1-59259-763-7:071.

Benus, R. F. J., van der Werf, T. S., Welling, G. W., Judd, P. A., Taylor, M. A., Harmsen, H. J. M., et al. (2010). Association between Faecalibacterium prausnitzii and dietary fibre in colonic fermentation in healthy human subjects. *Br. J. Nutr.* 104, 693–700. doi:10.1017/S0007114510001030.

Bernstein, C. N. (2015). Treatment of IBD: Where we are and where we are going. *Am. J. Gastroenterol.* 110, 114–126. doi:10.1038/ajg.2014.357.

Bernstein, C. N., Wajda, A., and Blanchard, J. F. (2005). The clustering of other chronic inflammatory diseases in inflammatory bowel disease: A population-based study. *Gastroenterology* 129, 827–836. doi:10.1053/j.gastro.2005.06.021.

Bonen, D. K., Ogura, Y., Nicolae, D. L., Inohara, N., Saab, L., Tanabe, T., et al. (2003). Crohn's disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. *Gastroenterology* 124, 140–146. doi:10.1053/gast.2003.50019.

Burisch, J., and Munkholm, P. (2015). The epidemiology of inflammatory bowel

disease. *Scand. J. Gastroenterol.* 50, 942–951. doi:10.3109/00365521.2015.1014407.

Carlsson, A. H., Yakymenko, O., Olivier, I., Håkansson, F., Postma, E., Keita, Å. V, et al. (2013). *Faecalibacterium prausnitzii* supernatant improves intestinal barrier function in mice DSS colitis. *Scand. J. Gastroenterol.* 48, 1136–1144. doi:10.3109/00365521.2013.828773.

Castillo, M., Skene, G., Roca, M., Anguita, M., Badiola, I., Duncan, S. H., et al. (2007). Application of 16S rRNA gene-targetted fluorescence in situ hybridization and restriction fragment length polymorphism to study porcine microbiota along the gastrointestinal tract in response to different sources of dietary fibre. *FEMS Microbiol. Ecol.* 59, 138–46. doi:10.1111/j.1574-6941.2006.00204.x.

Cato, E. P., Salmon, C. W., and Moore, W. E. C. (1974). Fusobacterium prausnitzii (Hauduroy et al.) Moore and Holdeman: Emended Description and Designation of Neotype Strain. *Int. J. Syst. Bacteriol.* 24, 225–229. Available at: http://www.microbiologyresearch.org/docserver/fulltext/ijsem/24/2/ijs-24-2-225.pdf?expires=1528896793&id=id&accname=guest&checksum=DA76C61654D57DA66E310CB179EB94C8 [Accessed June 13, 2018].

Chan, J. Z.-M., Halachev, M. R., Loman, N. J., Constantinidou, C., and Pallen, M. J. (2012). Defining bacterial species in the genomic era: insights from the genus Acinetobacter. *BMC Microbiol.* 12, 302. doi:10.1186/1471-2180-12-302.

Chang, C., and Lin, H. (2016). Dysbiosis in gastrointestinal disorders. *Best Pract. Res. Clin. Gastroenterol.* 30, 3–15. doi:10.1016/j.bpg.2016.02.001.

Chassaing, B., and Darfeuille–Michaud, A. (2011). The Commensal Microbiota and Enteropathogens in the Pathogenesis of Inflammatory Bowel Diseases. *Gastroenterology* 140, 1720–1728.e3. doi:10.1053/j.gastro.2011.01.054.

Cho, J. H. (2008). Inflammatory bowel disease: genetic and epidemiologic considerations. *World J. Gastroenterol.* 14, 338–47. doi:18200657.

Christl, S. U., Eisner, H. D., Dusel, G., Kasper, H., and Scheppach, W. (1996). Antagonistic effects of sulfide and butyrate on proliferation of colonic mucosa: a potential role for these agents in the pathogenesis of ulcerative

colitis. *Dig. Dis. Sci.* 41, 2477–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9011461 [Accessed April 16, 2018].

Chung, W. S. F., Walker, A. W., Louis, P., Parkhill, J., Vermeiren, J., Bosscher, D., et al. (2016). Modulation of the human gut microbiota by dietary fibres occurs at the species level. *BMC Biol.* 14, 3. doi:10.1186/s12915-015-0224-3.

Cobrin, G. M., and Abreu, M. T. (2005). Defects in mucosal immunity leading to Crohn's disease. *Immunol. Rev.* 206, 277–295. doi:10.1111/j.0105-2896.2005.00293.x.

COENYE, T., GEVERS, D., DEPEER, Y., VANDAMME, P., and SWINGS, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol. Rev.* 29, 147–167. doi:10.1016/j.femsre.2004.11.004.

Cosnes, J., Gowerrousseau, C., Seksik, P., and Cortot, A. (2011). Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology* 140, 1785–1794. doi:10.1053/j.gastro.2011.01.055.

Desai, D. (2014). Colorectal cancer surveillance in inflammatory bowel disease: A critical analysis. *World J. Gastrointest. Endosc.* 6, 541. doi:10.4253/wjge.v6.i11.541.

Dobrindt, U., and Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Curr. Opin. Microbiol.* 4, 550–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11587932 [Accessed April 16, 2018].

Dörffel, Y., Swidsinski, A., Loening-Baucke, V., Wiedenmann, B., and Pavel, M. (2012). Common biostructure of the colonic microbiota in neuroendocrine tumors and Crohn’s disease and the effect of therapy. *Inflamm. Bowel Dis.* 18, 1663–1671. doi:10.1002/ibd.21923.

Duerr, R. H. (2007). Genome-Wide Association Studies Herald a New Era of Rapid Discoveries in Inflammatory Bowel Disease Research. *Gastroenterology* 132, 2045–2049. doi:10.1053/j.gastro.2007.03.082.

Duncan, S. H., Hold, G. L., Harmsen, H. J. M., Stewart, C. S., and Flint, H. J. (2002). Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 52, 2141–6. doi:10.1099/00207713-52-6-2141.

Eckburg, P., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent,

M., et al. (2005). Diversity of the Human Intestinal Microbial Flora. *Science (80-.).* 86, 573–579. doi:10.1109/TMI.2012.2196707.Separate.

Emerson, D., Agulto, L., Liu, H., and Liu, L. (2008). Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics. *Bioscience* 58, 925–936. doi:10.1641/B581006.

Fallani, M., Rigottier-Gois, L., Aguilera, M., Bridonneau, C., Collignon, A., Edwards, C. A., et al. (2006). Clostridium difficile and Clostridium perfringens species detected in infant faecal microbiota using 16S rRNA targeted probes. *J. Microbiol. Methods* 67, 150–161. doi:10.1016/J.MIMET.2006.03.010.

Falvey, J. D., Hoskin, T., Meijer, B., Ashcroft, A., Walmsley, R., Day, A. S., et al. (2015). Disease activity assessment in IBD: Clinical indices and biomarkers fail to predict endoscopic remission. *Inflamm. Bowel Dis.* 21, 824–831. doi:10.1097/MIB.0000000000000341.

FAO/WHO (2001). Probiotics in food Health and nutritional properties and guidelines for evaluation FAO FOOD AND NUTRITION PAPER. Available at: http://www.fao.org/3/a-a0512e.pdf [Accessed July 6, 2018].

Figueiredo, H. C. P., Soares, S. C., Pereira, F. L., Dorella, F. A., Carvalho, A. F., Teixeira, J. P., et al. (2015). Comparative genome analysis of Weissella ceti, an emerging pathogen of farm-raised rainbow trout. *BMC Genomics* 16, 1095. doi:10.1186/s12864-015-2324-4.

Gareau, M. G., Sherman, P. M., and Walker, W. A. (2010). Probiotics and the gut microbiota in intestinal health and disease. *Nat. Rev. Gastroenterol. Hepatol.* 7, 503–514. doi:10.1038/nrgastro.2010.117.

Garneau, J. E., and Moineau, S. (2011). Bacteriophages of lactic acid bacteria and their impact on milk fermentations. *Microb. Cell Fact.* 10 Suppl 1, S20. doi:10.1186/1475-2859-10-S1-S20.

Gerritsen, J., Smidt, H., Rijkers, G. T., and de Vos, W. M. (2011). Intestinal microbiota in human health and disease: the impact of probiotics. *Genes Nutr.* 6, 209–40. doi:10.1007/s12263-011-0229-7.

Goldsmith, J. R., and Sartor, R. B. (2014). The role of diet on intestinal microbiota metabolism: Downstream impacts on host immune function and health, and therapeutic implications. *J. Gastroenterol.* 49, 785–798. doi:10.1007/s00535-014-0953-z.

Goodfellow, M., Sutcliffe, I., and Chun, J. (2014). *New Approaches to Prokaryotic Systematics*.

Guarner, F., Bourdet-Sicard, R., Brandtzaeg, P., Gill, H. S., McGuirk, P., van Eden, W., et al. (2006). Mechanisms of Disease: the hygiene hypothesis revisited. *Nat. Clin. Pract. Gastroenterol. Hepatol.* 3, 275–284. doi:10.1038/ncpgasthep0471.

Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. (2011). The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.* 7, e1002222. doi:10.1371/journal.pgen.1002222.

Guinane, C. M., and Cotter, P. D. (2013). Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therap. Adv. Gastroenterol.* 6, 295–308. doi:10.1177/1756283X13482996.

Hanauer, S. B. (2006). Inflammatory bowel disease: Epidemiology, pathogenesis, and therapeutic opportunities. *Inflamm. Bowel Dis.* 12, S3–S9. doi:10.1097/01.MIB.0000195385.19268.68.

Hansen, R., Russell, R. K., Reiff, C., Louis, P., McIntosh, F., Berry, S. H., et al. (2012). Microbiota of De-Novo Pediatric IBD: Increased Faecalibacterium Prausnitzii and Reduced Bacterial Diversity in Crohn's But Not in Ulcerative Colitis. *Am. J. Gastroenterol.* 107, 1913–1922. doi:10.1038/ajg.2012.335.

Hippe, B., Remely, M., Aumueller, E., Pointner, A., Magnet, U., and Haslberger, A. G. (2016). Faecalibacterium prausnitzii phylotypes in type two diabetic, obese, and lean control subjects. *Benef. Microbes* 7, 511–517. doi:10.3920/BM2015.0075.

Hopkins, M. J., Macfarlane, G. T., Furrie, E., Fite, A., and Macfarlane, S. (2005). Characterisation of intestinal bacteria in infant stools using real-time PCR and northern hybridisation analyses. *FEMS Microbiol. Ecol.* 54, 77–85. doi:10.1016/j.femsec.2005.03.001.

Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., et al. (2013). Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* 155, 1451–1463. doi:10.1016/j.cell.2013.11.024.

Ianiro, G., Bibbò, S., Scaldaferri, F., Gasbarrini, A., and Cammarota, G. (2014).

Fecal Microbiota Transplantation in Inflammatory Bowel Disease. *Medicine (Baltimore).* 93, e97. doi:10.1097/MD.0000000000000097.

J.B. Prajapati and Suja Senan (2013). Omics: a novel tool for safety assessment of probiotic strains. *Asian J. Food Agro-Industry* 06, 208–221. Available at: http://www.ajofai.info/Abstract/Omics - a novel tool for safety assessment of probiotic strains.pdf [Accessed April 27, 2016].

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60, 631–637. doi:10.1136/gut.2010.223263.

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491, 119–124. doi:10.1038/nature11582.

Khan, K. J., Ullman, T. A., Ford, A. C., Abreu, M. T., Abadir, A., Marshall, J. K., et al. (2011). Antibiotic therapy in inflammatory bowel disease: A systematic review and meta-analysis. *Am. J. Gastroenterol.* 106, 661–673. doi:10.1038/ajg.2011.72.

Khan, M. T., Duncan, S. H., Stams, A. J. M., van Dijl, J. M., Flint, H. J., and Harmsen, H. J. M. (2012). The gut anaerobe Faecalibacterium prausnitzii uses an extracellular electron shuttle to grow at oxic–anoxic interphases. *ISME J.* 6, 1578–1585. doi:10.1038/ismej.2012.5.

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi:10.1099/ijs.0.059774-0.

Klampfer, L., Huang, J., Sasazuki, T., Shirasawa, S., and Augenlicht, L. (2003). Inhibition of interferon gamma signaling by the short chain fatty acid butyrate. *Mol. Cancer Res.* 1, 855–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14517348 [Accessed April 16, 2018].

Koga, Y., Tokunaga, S., Nagano, J., Sato, F., Konishi, K., Tochio, T., et al. (2016). Age-associated effect of kestose on Faecalibacterium prausnitzii and symptoms in the atopic dermatitis infants. doi:10.1038/pr.2016.167.

Konstantinidis, K. T., Ramette, A., and Tiedje, J. M. (2006). The bacterial

species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1929–40. doi:10.1098/rstb.2006.1920.

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., et al. (2013). A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. *PLoS Comput. Biol.* 9. doi:10.1371/journal.pcbi.1002863.

Lay, C., Sutren, M., Rochet, V., Saunier, K., Doré, J., and Rigottier-Gois, L. (2005). Design and validation of 16S rRNA probes to enumerate members of the Clostridium leptum subgroup in human faecal microbiota. *Environ. Microbiol.* 7, 933–946. doi:10.1111/j.1462-2920.2005.00763.x.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., and Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci.* 102, 11070–11075. doi:10.1073/pnas.0504978102.

Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848. doi:10.1016/j.cell.2006.02.017.

Liu, T.-C., and Stappenbeck, T. S. (2016). Genetics and Pathogenesis of Inflammatory Bowel Disease. *Annu. Rev. Pathol. Mech. Dis.* 11, 127–148. doi:10.1146/annurev-pathol-012615-044152.

Loftus, E. V (2004). Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* 126, 1504–1517. doi:10.1053/j.gastro.2004.01.063.

Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J., and Martinez-Medina, M. (2017). Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. *ISME J.* 11, 841–852. doi:10.1038/ismej.2016.176.

Lopez-Siles, M., Khan, T. M., Duncan, S. H., Harmsen, H. J. M., Garcia-Gil, L. J., and Flint, H. J. (2012). Cultured representatives of two major phylogroups of human colonic Faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl. Environ. Microbiol.* 78, 420–8. doi:10.1128/AEM.06858-11.

Lopez-Siles, M., Martinez-Medina, M., Surís-Valls, R., Aldeguer, X., Sabat-Mir, M., Duncan, S. H., et al. (2016). Changes in the Abundance of Faecalibacterium prausnitzii Phylogroups I and II in the Intestinal Mucosa of Inflammatory Bowel Disease and Patients with Colorectal Cancer.

Inflamm. Bowel Dis. 22, 28–41. doi:10.1097/MIB.0000000000000590.

Louis, P., Scott, K. P., Duncan, S. H., and Flint, H. J. (2007). Understanding the effects of diet on bacterial metabolism in the large intestine. *J. Appl. Microbiol.* 102, 1197–1208. doi:10.1111/j.1365-2672.2007.03322.x.

Lovasz, B. D., Golovics, P. A., Vegh, Z., and Lakatos, P. L. (2013). New trends in inflammatory bowel disease epidemiology and disease course in Eastern Europe. *Dig. Liver Dis.* 45, 269–276. doi:10.1016/j.dld.2012.08.020.

Maccaferri, S., Vitali, B., Klinder, A., Kolida, S., Ndagijimana, M., Laghi, L., et al. (2010). Rifaximin modulates the colonic microbiota of patients with Crohn's disease: an in vitro approach using a continuous culture colonic model system. *J. Antimicrob. Chemother.* 65, 2556–2565. doi:10.1093/jac/dkq345.

Macfarlane, G. T., and Macfarlane, S. (2011). Fermentation in the Human Large Intestine. *J. Clin. Gastroenterol.* 45, S120–S127. doi:10.1097/MCG.0b013e31822fecfe.

Maiden, M. C. J., van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013a). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi:10.1038/nrmicro3093.

Maiden, M. C. J., van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013b). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi:10.1038/nrmicro3093.

Manichanh, C., Borruel, N., Casellas, F., and Guarner, F. (2012). The gut microbiota in IBD. *Nat. Rev. Gastroenterol. Hepatol.* 9, 599–608. doi:10.1038/nrgastro.2012.152.

Manrique, P., Dills, M., and Young, M. (2017). The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses* 9, 141. doi:10.3390/v9060141.

Mariat, D., Firmesse, O., Levenez, F., Guimarães, V., Sokol, H., Doré, J., et al. (2009). The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol.* 9, 123. doi:10.1186/1471-2180-9-123.

Marteau, P. (2006). Probiotics, prebiotics, synbiotics: Ecological treatment for inflammatory bowel disease? *Gut* 55, 1692–1693. doi:10.1136/gut.2004.051458.

Martín, R., Bermúdez-Humarán, L. G., and Langella, P. (2018). Searching for the Bacterial Effector: The Example of the Multi-Skilled Commensal Bacterium Faecalibacterium prausnitzii. *Front. Microbiol.* 9, 346. doi:10.3389/fmicb.2018.00346.

Martín, R., Chain, F., Miquel, S., Lu, J., Gratadoux, J.-J., Sokol, H., et al. (2014a). The Commensal Bacterium Faecalibacterium prausnitzii Is Protective in DNBS-induced Chronic Moderate and Severe Colitis Models. *Inflamm. Bowel Dis.* 20, 417–430. doi:10.1097/01.MIB.0000440815.76627.64.

Martín, R., Martín, R., Chain, F., Chain, F., Miquel, S., Miquel, S., et al. (2014b). Effects in the use of a genetically engineered strain of Lactococcus lactis delivering in situ IL-10 as a therapy to treat low-grade colon inflammation. *Hum. Vaccin. Immunother.* 10, 1611–1621. doi:10.4161/hv.28549.

Martín, R., Miquel, S., Benevides, L., Bridonneau, C., Robert, V., Hudault, S., et al. (2017). Functional Characterization of Novel Faecalibacterium prausnitzii Strains Isolated from Healthy Volunteers: A Step Forward in the Use of F. prausnitzii as a Next-Generation Probiotic. *Front. Microbiol.* 8, 1226. doi:10.3389/fmicb.2017.01226.

Martín, R., Miquel, S., Chain, F., Natividad, J. M., Jury, J., Lu, J., et al. (2015). Faecalibacterium prausnitzii prevents physiological damages in a chronic low-grade inflammation murine model. *BMC Microbiol.* 15, 67. doi:10.1186/s12866-015-0400-1.

Martín, R., Miquel, S., Ulmer, J., Kechaou, N., Langella, P., and Bermúdez-Humarán, L. G. (2013). Role of commensal and probiotic bacteria in human health: a focus on inflammatory bowel disease. *Microb. Cell Fact.* 12, 71. doi:10.1186/1475-2859-12-71.

Martinez, C., Antolin, M., Santos, J., Torrejon, A., Casellas, F., Borruel, N., et al. (2008). Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *Am. J. Gastroenterol.* 103, 643–648. doi:10.1111/j.1572-0241.2007.01592.x.

Mazmanian, S. K., Round, J. L., and Kasper, D. L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 453, 620–625. doi:10.1038/nature07008.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat. Rev.*

Genet. 11, 31–46. doi:10.1038/nrg2626.

Mintz, R., Feller, E. R., Bahr, R. L., and Shah, S. a (2004). Ocular manifestations of inflammatory bowel disease. *Inflamm. Bowel Dis.* 10, 135–139. doi:http://dx.doi.org/10.1097/00054725-200403000-00012.

Miquel, S., Leclerc, M., Martin, R., Chain, F., Lenoir, M., Raguideau, S., et al. (2015). Identification of Metabolic Signatures Linked to Anti-Inflammatory Effects of Faecalibacterium prausnitzii. *MBio* 6, e00300-15. doi:10.1128/mBio.00300-15.

Miquel, S., Martín, R., Bridonneau, C., Robert, V., Sokol, H., Bermúdez-Humarán, L. G., et al. (2014). Ecology and metabolism of the beneficial intestinal commensal bacterium Faecalibacterium prausnitzii. *Gut Microbes* 5, 146–151. doi:10.4161/gmic.27651.

Miquel, S., Martín, R., Lashermes, A., Gillet, M., Meleine, M., Gelot, A., et al. (2016). Anti-nociceptive effect of Faecalibacterium prausnitzii in non-inflammatory IBS-like models. *Sci. Rep.* 6, 19399. doi:10.1038/srep19399.

Miquel, S., Martín, R., Rossi, O., Bermúdez-Humarán, L. G., Chatel, J. M., Sokol, H., et al. (2013). Faecalibacterium prausnitzii and human intestinal health. *Curr. Opin. Microbiol.* 16, 255–61. doi:10.1016/j.mib.2013.06.003.

Molodecky, N. A., Soon, I. S., Rabi, D. M., Ghali, W. A., Ferris, M., Chernoff, G., et al. (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 142, 46–54.e42. doi:10.1053/j.gastro.2011.10.001.

Moreno-Indias, I., Cardona, F., Tinahones, F. J., and Queipo-Ortuño, M. I. (2014). Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus. *Front. Microbiol.* 5, 1–10. doi:10.3389/fmicb.2014.00190.

Mowat, A. M., and Agace, W. W. (2014). Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* 14, 667–685. doi:10.1038/nri3738.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., et al. (2017). Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* 45, D446–D456. doi:10.1093/nar/gkw992.

Nava, G. M., and Stappenbeck, T. S. (2011). Diversity of the autochthonous

colonic microbiota. *Gut Microbes* 2, 99–104. doi:10.4161/gmic.2.2.15416.

Nielsen, O. H., and Munck, L. K. (2007). Drug Insight: Aminosalicylates for the treatment of IBD. *Nat. Clin. Pract. Gastroenterol. Hepatol.* 4, 160–170. doi:10.1038/ncpgasthep0696.

Noguchi, E., Homma, Y., Kang, X., Netea, M. G., and Ma, X. (2009). A Crohn's disease-associated NOD2 mutation suppresses transcription of human IL10 by inhibiting activity of the nuclear ribonucleoprotein hnRNP-A1. *Nat. Immunol.* 10, 471–479. doi:10.1038/ni.1722.

O'Toole, P. W., Marchesi, J. R., and Hill, C. (2017). Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nat. Microbiol.* 2, 17057. doi:10.1038/nmicrobiol.2017.57.

Oikonomou, G., Teixeira, A. G. V., Foditsch, C., Bicalho, M. L., Machado, V. S., and Bicalho, R. C. (2013). Fecal microbial diversity in pre-weaned dairy calves as described by pyrosequencing of metagenomic 16S rDNA. Associations of Faecalibacterium species with health and growth. *PLoS One* 8, e63157. doi:10.1371/journal.pone.0063157.

Orlicka, K., Barnes, E., and Culver, E. L. (2013). Prevention of infection caused by immunosuppressive drugs in gastroenterology. *Ther. Adv. Chronic Dis.* 4, 167–185. doi:10.1177/2040622313485275.

Ott, S. J., Plamondon, S., Hart, A., Begun, A., Rehman, A., Kamm, M. A., et al. (2008). Dynamics of the mucosa-associated flora in ulcerative colitis patients during remission and clinical relapse. *J. Clin. Microbiol.* 46, 3510–3. doi:10.1128/JCM.01512-08.

Papadimitriou, K., Zoumpopoulou, G., Foligné, B., Alexandraki, V., Kazou, M., Pot, B., et al. (2015). Discovering probiotic microorganisms: in vitro, in vivo, genetic and omics approaches. *Front. Microbiol.* 6, 58. doi:10.3389/fmicb.2015.00058.

Qiu, X., Zhang, M., Yang, X., Hong, N., and Yu, C. (2013). Faecalibacterium prausnitzii upregulates regulatory T cells and anti-inflammatory cytokines in treating TNBS-induced colitis. *J. Crohn's Colitis* 7, e558–e568. doi:10.1016/j.crohns.2013.04.002.

Quévrain, E., Maubert, M. A., Michon, C., Chain, F., Marquant, R., Miquel, S., et al. (2017). Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii, a commensal bacterium deficient in Crohn's

disease. *Gut* 65, 415–425. doi:10.1136/gutjnl-2014-307649.Identification.

Quévrain, E., Maubert, M. A., Michon, C., Chain, F., Marquant, R., Tailhades, J., et al. (2014). Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii , a commensal bacterium deficient in Crohn's disease. *Gut* 65, 415–425. doi:10.1136/gutjnl-2014-307649.

Rahimi, R., Nikfar, S., Rezaie, A., and Abdollahi, M. (2007). A meta-analysis of antibiotic therapy for active ulcerative colitis. *Dig. Dis. Sci.* 52, 2920–2925. doi:10.1007/s10620-007-9760-1.

Rajilić-Stojanović, M., Biagi, E., Heilig, H. G. H. J., Kajander, K., Kekkonen, R. A., Tims, S., et al. (2011). Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* 141, 1792–801. doi:10.1053/j.gastro.2011.07.043.

Ramirez-Farias, C., Slezak, K., Fuller, Z., Duncan, A., Holtrop, G., and Louis, P. (2009). Effect of inulin on the human gut microbiota: stimulation of Bifidobacterium adolescentis and Faecalibacterium prausnitzii. *Br. J. Nutr.* 101, 533. doi:10.1017/S0007114508019880.

Rehman, A., Rausch, P., Wang, J., Skieceviciene, J., Kiudelis, G., Bhagalia, K., et al. (2016). Geographical patterns of the standing and active human gut microbiome in health and IBD. *Gut* 65, 238–248. doi:10.1136/gutjnl-2014-308341.

Reinoso Webb, C., Koboziev, I., Furr, K. L., and Grisham, M. B. (2016). Protective and pro-inflammatory roles of intestinal bacteria. *Pathophysiology* 23, 67–80. doi:10.1016/j.pathophys.2016.02.002.

Ricci, A., Allende, A., Bolton, D., Chemaly, M., Davies, R., Girones, R., et al. (2018). Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 7: suitability of taxonomic units notified to EFSA until September 2017. *EFSA J.* 16. doi:10.2903/j.efsa.2018.5131.

Rodríguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., et al. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.* 26, 26050. doi:10.3402/MEHD.V26.26050.

Rossi, O., van Berkel, L. A., Chain, F., Tanweer Khan, M., Taverne, N., Sokol,

H., et al. (2016). Faecalibacterium prausnitzii A2-165 has a high capacity to induce IL-10 in human and murine dendritic cells and modulates T cell responses. *Sci. Rep.* 6, 18507. doi:10.1038/srep18507.

Sangal, V., Goodfellow, M., Jones, A. L., Schwalbe, E. C., Blom, J., Hoskisson, P. A., et al. (2016). Next-generation systematics: An innovative approach to resolve the structure of complex prokaryotic taxa. *Sci. Rep.* 6, 38392. doi:10.1038/srep38392.

Sartor, R. B. (2010). Genetics and environmental interactions shape the intestinal microbiome to promote inflammatory bowel disease versus mucosal homeostasis. *Gastroenterology* 139, 1816–1819. doi:10.1053/j.gastro.2010.10.036.

Schmidt, H., and Hensel, M. (2004). Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* 17, 14–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14726454 [Accessed April 16, 2018].

Scupham, A. J. (2007). Succession in the intestinal microbiota of preadolescent turkeys. *FEMS Microbiol. Ecol.* 60, 136–147. doi:10.1111/j.1574-6941.2006.00245.x.

SEKSIK, P., SOKOL, H., LEPAGE, P., VASQUEZ, N., MANICHANH, C., MANGIN, I., et al. (2006). Review article: the role of bacteria in onset and perpetuation of inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 24, 11–18. doi:10.1111/j.1365-2036.2006.03053.x.

Shivananda, S., Lennard-Jones, J., Logan, R., Fear, N., Price, A., Carpenter, L., et al. (1996). Incidence of inflammatory bowel disease across Europe: Is there a difference between north and south? Results of the European collaborative study on inflammatory bowel disease (EC-IBD). *Gut* 39, 690–697. doi:10.1136/gut.39.5.690.

Soares, S. C., Abreu, V. a C., Ramos, R. T. J., Cerdeira, L., Silva, A., Baumbach, J., et al. (2012). PIPS: pathogenicity island prediction software. *PLoS One* 7, e30848. doi:10.1371/journal.pone.0030848.

Sokol, H., Lay, C., Seksik, P., and Tannock, G. W. (2008a). Analysis of bacterial bowel communities of IBD patients: What has it revealed? *Inflamm. Bowel Dis.* 14, 858–867. doi:10.1002/ibd.20392.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J.-J., et al. (2008b). Faecalibacterium prausnitzii is an anti-

inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci.* 105, 16731–16736. doi:10.1073/pnas.0804812105.

Song, H., Yoo, Y., Hwang, J., Na, Y.-C., and Kim, H. S. (2016). Faecalibacterium prausnitzii subspecies–level dysbiosis in the human gut microbiome underlying atopic dermatitis. *J. Allergy Clin. Immunol.* 137, 852–860. doi:10.1016/j.jaci.2015.08.021.

Stepaniuk, P., Bernstein, C. N., Targownik, L. E., and Singh, H. (2015). Characterization of inflammatory bowel disease in elderly patients: A review of epidemiology, current practices and outcomes of current management strategies. *Can. J. Gastroenterol. Hepatol.* 29, 327–33. doi:10.1155/2015/136960.

Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M., and Doerffel, Y. (2008). Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. *Inflamm. Bowel Dis.* 14, 147–161. doi:10.1002/ibd.20330.

Takahashi, K., Nishida, A., Fujimoto, T., Fujii, M., Shioya, M., Imaeda, H., et al. (2016). Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease. *Digestion* 93, 59–65. doi:10.1159/000441768.

Tamboli, C. P., Neut, C., Desreumaux, P., and Colombel, J. F. (2004). Dysbiosis in inflammatory bowel disease. *Gut* 53, 1–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14684564 [Accessed July 6, 2018].

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-P., et al. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* 11, 2574–84. doi:10.1111/j.1462-2920.2009.01982.x.

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi:10.1016/j.mib.2008.09.006.

Thia, K. T., Loftus, E. V., Sandborn, W. J., and Yang, S. K. (2008). An update on the epidemiology of inflammatory bowel disease in Asia. *Am. J. Gastroenterol.* 103, 3167–3182. doi:10.1111/j.1572-0241.2008.02158.x.

Thursby, E., and Juge, N. (2017). Introduction to the human gut microbiota. *Biochem. J.* 474, 1823–1836. doi:10.1042/BCJ20160510.

Tontini, G. E., Vecchi, M., Pastorelli, L., Neurath, M. F., and Neumann, H. (2015). Differential diagnosis in inflammatory bowel disease colitis: State of the art and future perspectives. *World J. Gastroenterol.* 21, 21–46. doi:10.3748/wjg.v21.i1.21.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484. doi:10.1038/nature07540.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031. doi:10.1038/nature05414.

Underwood, M. A., German, J. B., Lebrilla, C. B., and Mills, D. A. (2015). Bifidobacterium longum subspecies infantis: Champion colonizer of the infant gut. *Pediatr. Res.* 77, 229–235. doi:10.1038/pr.2014.156.

Vangay, P., Ward, T., Gerber, J. S., and Knights, D. (2015). Antibiotics, pediatric dysbiosis, and disease. *Cell Host Microbe* 17, 553–564. doi:10.1016/j.chom.2015.04.006.

Velasquez-Manoff, M. (2015). Gut microbiome: The peacekeepers. *Sci. Am.* 312, S3–S11. doi:10.1038/scientificamerican0315-S3.

Ventura, M., O'Flaherty, S., Claesson, M. J., Turroni, F., Klaenhammer, T. R., van Sinderen, D., et al. (2009). Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat. Rev. Microbiol.* 7, 61–71. doi:10.1038/nrmicro2047.

Ventura, M., Sozzi, T., Turroni, F., Matteuzzi, D., and van Sinderen, D. (2011). The impact of bacteriophages on probiotic bacteria and gut microbiota diversity. *Genes Nutr.* 6, 205–207. doi:10.1007/s12263-010-0188-4.

Villanueva-Millán, M. J., Pérez-Matute, P., and Oteo, J. A. (2015). Gut microbiota: a key player in health and disease. A review focused on obesity. *J. Physiol. Biochem.* 71, 509–525. doi:10.1007/s13105-015-0390-3.

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., et al. (2011a). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* 5, 220–30. doi:10.1038/ismej.2010.118.

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N.,

Rayment, N., et al. (2011b). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol.* 11. doi:10.1186/1471-2180-11-7.

Wallace, K. L., Zheng, L. B., Kanazawa, Y., and Shih, D. Q. (2014). Immunopathology of inflammatory bowel disease. *World J. Gastroenterol.* 20, 6–21. doi:10.3748/wjg.v20.i1.6.

WANG, R.-F., CAO, W.-W., and CERNIGLIA, C. E. (1996). Phylogenetic Analysis of Fusobacterium prausnitzii Based upon the 16S rRNA Gene Sequence and PCR Confirmation. *Int. J. Syst. Bacteriol.* 46, 341–343. doi:10.1099/00207713-46-1-341.

Ward, M. A., Pierre, J. F., Leal, R. F., Huang, Y., Shogan, B., Dalal, S. R., et al. (2016). Insights into the pathogenesis of ulcerative colitis from a murine model of stasis-induced dysbiosis, colonic metaplasia, and genetic susceptibility. *Am. J. Physiol. - Gastrointest. Liver Physiol.* 310, G973–G988. doi:10.1152/ajpgi.00017.2016.

Watanabe, T., Kitani, A., Murray, P. J., and Strober, W. (2004). NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. *Nat. Immunol.* 5, 800–808. doi:10.1038/ni1092.

Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schäffeler, E., Schlee, M., et al. (2004). NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal α-defensin expression. *Gut* 53, 1658–1664. doi:10.1136/gut.2003.032805.

Wrzosek, L., Miquel, S., Noordine, M.-L., Bouet, S., Joncquel Chevalier-Curt, M., Robert, V., et al. (2013). Bacteroides thetaiotaomicron and Faecalibacterium prausnitzii influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC Biol.* 11, 61. doi:10.1186/1741-7007-11-61.

Xiong, J. (2006). *Essential bioinformatics*. 1st ed. Cambridge: Cambridge University Press Available at: http://books.google.com/books?hl=en&lr=&id=AFsu7_goA8kC&oi=fnd&pg=PA14&dq=ESSENTIAL+BIOINFORMATICS&ots=hGtoxR7ulh&sig=Maaw64FCUvKDXeaya_g41AzP_-g [Accessed May 28, 2014].

Youngster, I., Russell, G. H., Pindar, C., Ziv-Baran, T., Sauk, J., and Hohmann,

E. L. (2014). Oral, Capsulized, Frozen Fecal Microbiota Transplantation for Relapsing Clostridium difficile Infection. *JAMA - J. Am. Med. Assoc.* 312, 1772–1778. doi:10.1001/jama.2014.13875.

Zhang, Y. Z., and Li, Y. Y. (2014). Inflammatory bowel disease: Pathogenesis. *World J. Gastroenterol.* 20, 91–99. doi:10.3748/wjg.v20.i1.91.

Zhi, X.-Y., Zhao, W., Li, W.-J., and Zhao, G.-P. (2012). Prokaryotic systematics in the genomics era. *Antonie Van Leeuwenhoek* 101, 21–34. doi:10.1007/s10482-011-9667-x.

Zoetendal, E. G., Akkermans, A. D., Akkermans-van Vliet, W. M., de Visser, A. G. M., and de Vos, W. M. (2001). The Host Genotype Affects the Bacterial Community in the Human Gastrointestinal Tract. *Microb. Ecol. Health Dis.* 13, 129–134.

VIII Appendices

VIII.1 *Curriculum Vitae*

Address to this CV:	http://lattes.cnpq.br/0886258515265632
Full Name	Leandro de Jesus Benevides
Parental information	Jofre da Silva Benevides and Clicia Maria de Jesus Benevides
Birth information	05/07/1987 - Feira de Santana/BA - Brazil
Identification document	0953164934 SSP - BA - 05/05/2005
CPF Number	028.401.495-83
Passport	FM065659
Professional Address	Universidade Federal de Minas Gerais Instituto de Ciências Biológicas Departamento de Biologia Geral Av. Antonio Carlos. ICB bloco Q3 259 Belo Horizonte 31270-215, MG - Brazil
Eletronic Adress	ljbenevides@gmail.com

Formal Education

2014 – current date	Doctorate in Genetics. Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brazil with Sandwich Doctorate at “Institut National de la Recherche Agronomique”, Jouy-en-Josas, France and Joint supervision at “Université

Paris-Saclay, Paris, France (Advisor: Philippe Langella)

Title: Comparative Genomics of *Faecalibacterium* spp.

Advisor: Vasco Ariston de Carvalho Azevedo

Co- Advisor: Siomar de Castro Soares

2011 - 2013 Master's in Genetics and Biodiversity

Universidade Federal da Bahia, UFBA, Salvador, Brazil

Title: Estudo Evolutivo de Vertebrados Baseado em Sequências de Nucleotídeos e de Aminoácidos das Apolipoproteínas A1, A4 e E, Year of degree: 2013

Advisor: Flora Maria de Campos Fernandes

Co- Advisor: Gilberto Cafezeiro Bomfim

2007 - 2010 Bachelor's in Biomedicina.

Escola Bahiana de Medicina e Saúde Pública, EBMSP, Brazil

Title: Ferramentas Básicas da Bioinformática Úteis nas Pesquisas Biomédicas

Advisor: Flora Maria de Campos Fernandes

Bibliographic Production

Articles Published in Scientific Journals

1. CORNUAULT, J. K., PETIT, M.-A., MARIADASSOU, M., **BENEVIDES, L**., MONCAUT, E., LANGELLA, P., SOKOL, H., De PAEPE, M.

Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes. *Microbiome* 6, 65. doi:10.1186/s40168-018-0452-1, 2018

2. JAMAL, SYED BABAR; HASSAN, SYED SHAH; TIWARI, SANDEEP; VIANA, MARCUS V.; **BENEVIDES, LEANDRO DE JESUS**; ULLAH, ASAD; TURJANSKI, ADRIÁN G.; BARH, DEBMALYA; GHOSH, PREETAM; COSTA, DANIELA ARRUDA; SILVA, ARTUR; RÖTTGER, RICHARD; BAUMBACH, JAN; AZEVEDO, VASCO A. C.

An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae. PLoS One. , v.12, p.e0186401 - , 2017.

3. OLIVEIRA, LETÍCIA C.; SARAIVA, TESSÁLIA D. L.; SILVA, WANDERSON M.; PEREIRA, ULISSES P.; CAMPOS, BRUNO C.; **BENEVIDES, LEANDRO J.**; ROCHA, FLÁVIA S.; FIGUEIREDO, HENRIQUE C. P.; AZEVEDO, VASCO; SOARES, SIOMAR C.

Analyses of the probiotic property and stress resistance-related genes of Lactococcus lactis subsp. lactis NCDO 2118 through comparative genomics and in vitro assays. PLoS One. , v.12, p.e0175116 - , 2017.

4. BARAÚNA, RAFAEL A.; RAMOS, ROMMEL T. J.; VERAS, ADONNEY A. O.; PINHEIRO, KENNY C.; **BENEVIDES, LEANDRO J.**; VIANA, MARCUS V. C.; GUIMARÃES, LUÍS C.; EDMAN, JUDY M.; SPIER, SHARON J.; AZEVEDO, VASCO; SILVA, ARTUR

Assessing the Genotypic Differences between Strains of Corynebacterium pseudotuberculosis biovar equi through Comparative Genomics. Plos One. , v.12, p.e0170676 - , 2017.

5. HURTADO, RAQUEL ENMA; ABURJAILE, FLAVIA; MARIANO, DIEGO; CANÁRIO, MARCUS VINICIUS; **BENEVIDES, LEANDRO**; FERNANDEZ, DANIEL ANTONIO; ALLASI, NATALY OLIVIA; RIMAC, ROCIO; JUSCAMAYTA, JULIO EDUARDO; MAXIMILIANO, JORGE ENRIQUE; ROSADIO, RAUL HECTOR; AZEVEDO, VASCO; MATURRANO, LENIN

Draft Genome Sequence of a Virulent Strain of Pasteurella Multocida Isolated From Alpaca. JOURNAL OF GENOMICS. , v.5, p.68 - 70, 2017.

6. MARTÍN, REBECA; MIQUEL, SYLVIE; **BENEVIDES, LEANDRO**; BRIDONNEAU, CHANTAL; ROBERT, VÉRONIQUE; HUDAULT, SYLVIE; CHAIN, FLORIAN; BERTEAU, OLIVIER; AZEVEDO, VASCO; CHATEL, JEAN M.; SOKOL, HARRY; BERMÚDEZ-HUMARÁN, LUIS G.; THOMAS, MURIEL; LANGELLA, PHILIPPE Functional Characterization of Novel Faecalibacterium prausnitzii Strains Isolated from Healthy Volunteers: A Step Forward in the Use of F. prausnitzii as a Next-Generation Probiotic. Frontiers in Microbiology. , v.8, p.1

- , 2017.

7. OLIVEIRA, ALBERTO; OLIVEIRA, LETICIA C.; ABURJAILE, FLAVIA; **BENEVIDES, LEANDRO**; TIWARI, SANDEEP; JAMAL, SYED B.; SILVA, ARTHUR; FIGUEIREDO, HENRIQUE C. P.; GHOSH, PREETAM; PORTELA, RICARDO W.; DE CARVALHO AZEVEDO, VASCO A.; WATTAM, ALICE R.

Insight of Genus Corynebacterium: Ascertaining the Role of Pathogenic and Non-pathogenic Species. Frontiers in Microbiology. , v.8, p.1 - , 2017.

8. **BENEVIDES, LEANDRO**; BURMAN, SRITI; MARTIN, REBECA; ROBERT, VÉRONIQUE; THOMAS, MURIEL; MIQUEL, SYLVIE; CHAIN, FLORIAN; SOKOL, HARRY; BERMUDEZ-HUMARAN, LUIS G.; MORRISON, MARK; LANGELLA, PHILIPPE; AZEVEDO, VASCO A.; CHATEL, JEAN-MARC; SOARES, SIOMAR

New Insights into the Diversity of the Genus Faecalibacterium. Frontiers in Microbiology. , v.8, p.1 - , 2017.

9. DE AGUIAR, EDGAR LACERDA; MARIANO, DIEGO CÉSAR BATISTA; VIANA, MARCUS VINÍCIUS CANÁRIO; **BENEVIDES, LEANDRO DE JESUS**; DE SOUZA ROCHA, FLÁVIA; DE CASTRO OLIVEIRA, LETÍCIA; PEREIRA, FELIPE LUIZ; DORELLA, FERNANDA ALVES; LEAL, CARLOS AUGUSTO GOMES; DE CARVALHO, ALEX FIORINI; SANTOS, GABRIELA SILVA; MATTOS-GUARALDI, ANA LUIZA; NAGAO, PRESCILLA EMY; DE CASTRO SOARES, SIOMAR; HASSAN, SYED SHAH; PINTO, ANNE CYBELE; FIGUEIREDO, HENRIQUE CÉSAR PEREIRA; AZEVEDO, VASCO

Complete genome sequence of Streptococcus agalactiae strain GBS85147 serotype of type Ia isolated from human oropharynx. Standards in Genomic Sciences. , v.11, p.1 - , 2016.

10. GUIMARÃES, LUIS C.; VIANA, MARCUS V. C.; **BENEVIDES, LEANDRO J.**; MARIANO, DIEGO C. B.; VERAS, ADOONEY A. O.; SÁ, PABLO H. C.; ROCHA, FLÁVIA S.; VILAS BOAS, PRISCILLA C. B.; SOARES, SIOMAR C.; BARBOSA, MARIA S.; GUISO, NICOLE; BADELL, EDGAR; CARNEIRO, ADRIANA R.; AZEVEDO, VASCO; RAMOS, ROMMEL T. J.; SILVA, ARTUR

Draft Genome Sequence of Corynebacterium ulcerans Strain 04-3911, Isolated from Humans. Genome Announcements. , v.4, p.e00171-16 - , 2016.

11. GUIMARÃES, LUIS C.; VIANA, MARCUS V. C.; **BENEVIDES, LEANDRO J.**; MARIANO, DIEGO C. B.; VERAS, ADONNEY A. O.; SÁ, PABLO H. C.; ROCHA, FLÁVIA S.; VILAS BOAS, PRISCILLA C. B.; SOARES, SIOMAR C.; BARBOSA, MARIA S.; GUISO, NICOLE; BADELL, EDGAR; AZEVEDO, VASCO; RAMOS, ROMMEL T. J.; SILVA, ARTUR

Draft Genome Sequence of Toxigenic Corynebacterium ulcerans Strain 03-8664 Isolated from a Human Throat. Genome Announcements. , v.4, p.e00719-16 - , 2016.

12. GUIMARÃES, LUIS C.; VIANA, MARCUS V. C.; **BENEVIDES, LEANDRO J.**; MARIANO, DIEGO C. B.; VERAS, ADOONEY A. O.; SÁ, PABLO H. C.; ROCHA, FLÁVIA S.; VILAS BOAS, PRISCILLA C. B.; SOARES, SIOMAR C.; BARBOSA, MARIA S.; GUISO, NICOLE; BADELL, EDGAR; CARNEIRO, ADRIANA R.; AZEVEDO, VASCO; RAMOS, ROMMEL T. J.; SILVA, ARTUR

Draft Genome Sequence of Toxigenic Corynebacterium ulcerans Strain 04-7514, Isolated from a Dog in France. Genome Announcements. , v.4, p.e00172-16 - , 2016.

13. GUIMARÃES, LUIS C.; VIANA, MARCUS V. C.; **BENEVIDES, L. J**.; MARIANO, D. C. B.; VERAS, ADONNEY A. O.; SÁ, PABLO H. C.; ROCHA, FLÁVIA S.; VILAS BOAS, PRISCILLA C. B.; SOARES, SIOMAR C.; BARBOSA, MARIA S.; GUISO, NICOLE; BADELL, EDGAR; AZEVEDO, VASCO; RAMOS, ROMMEL T. J.; SILVA, ARTUR

Draft Genome Sequence of Toxigenic Strain 03-8664 Isolated from a Human Throat. Genome Announcements. , v.4, p.e00719-16 - , 2016.

14. **BENEVIDES, LEANDRO DE JESUS**; CARVALHO, DANIEL SANTANA DE; ANDRADE, ROBERTO FERNANDES SILVA; BOMFIM, GILBERTO CAFEZEIRO; FERNANDES, FLORA MARIA DE CAMPOS Evolutionary analysis of apolipoprotein E by Maximum Likelihood and complex network methods. Genetics and Molecular Biology (online version). , v.39, p.665 - 673, 2016.

15. MARIANO, DIEGO C. B.; PEREIRA, FELIPE L.; AGUIAR, EDGAR L.; OLIVEIRA, LETÍCIA C.; **BENEVIDES, LEANDRO**; GUIMARÃES, LUÍS C.; FOLADOR, EDSON L.; SOUSA, THIAGO J.; GHOSH, PREETAM; BARH, DEBMALYA; FIGUEIREDO, HENRIQUE C. P.; SILVA, ARTUR; RAMOS, ROMMEL T. J.; AZEVEDO, VASCO A. C.

SIMBA: a web tool for managing bacterial genome assembly generated by Ion PGM sequencing technology. BMC Bioinformatics. , v.17, p.65 - 72, 2016.

16. SOUSA, THIAGO JESUS; MARIANO, DIEGO; PARISE, DOGLAS; PARISE, MARIANA; VIANA, MARCUS VINICIUS CANÁRIO; GUIMARÃES, LUIS CARLOS; **BENEVIDES, LEANDRO JESUS**; ROCHA, FLÁVIA; BAGANO, PRISCILLA; RAMOS, ROMMEL; SILVA, ARTUR; FIGUEIREDO, HENRIQUE; ALMEIDA, SINTIA; AZEVEDO, VASCO

Complete Genome Sequence of Corynebacterium pseudotuberculosis Strain 12C. Genome Announcements. , v.3, p.e00759-15 - , 2015.

17. **BENEVIDES, LEANDRO DE JESUS**; VIANA, MARCUS VINICIUS CANÁRIO; MARIANO, DIEGO CÉSAR BATISTA; ROCHA, FLÁVIA DE SOUZA; BAGANO, PRISCILLA CAROLINNE; FOLADOR, EDSON LUIZ; PEREIRA, FELIPE LUIZ; DORELLA, FERNANDA ALVES; LEAL, CARLOS AUGUSTO GOMES; CARVALHO, ALEX FIORINI; SOARES, SIOMAR DE CASTRO; CARNEIRO, ADRIANA; RAMOS, ROMMEL; BADELL-OCANDO, EDGAR; GUISO, NICOLE; SILVA, ARTUR; FIGUEIREDO, HENRIQUE; AZEVEDO, VASCO; GUIMARÃES, LUIS CARLOS

Genome Sequence of Corynebacterium ulcerans Strain FRC11. Genome Announcements. , v.3, p.e00112-15 - , 2015.

18. CARLOS GUIMARAES, LUIS; **BENEVIDES DE JESUS, LEANDRO**; VINICIUS CANARIO VIANA, MARCUS; SILVA, ARTUR; THIAGO JUCA RAMOS, ROMMEL; DE CASTRO SOARES, SIOMAR; AZEVEDO, VASCO

Inside of the Pan-Genome - Methods and Software Overview. Current Genomics. , v.16, p.1 - 1, 2015.

19. BARAUNA, R. A.; GUIMARAES, L. C.; VERAS, A. A. O.; DE SA, P. H. C. G.; GRACAS, D. A.; PINHEIRO, K. C.; SILVA, A. S. S.; FOLADOR, E. L.; **BENEVIDES, L. J.**; VIANA, M. V. C.; CARNEIRO, A. R.; SCHNEIDER, M. P. C.; SPIER, S. J.; EDMAN, J. M.; RAMOS, R. T. J.; AZEVEDO, V.; SILVA, A. Genome Sequence of Corynebacterium pseudotuberculosis MB20 bv. equi Isolated from a Pectoral Abscess of an Oldenburg Horse in California. Genome Announcements. , v.2, p.e00977-14 - e00977-14, 2014.

20. VIANA, M. V. C.; **DE JESUS BENEVIDES, L.**; BATISTA MARIANO, D. C.; DE SOUZA ROCHA, F.; BAGANO VILAS BOAS, P. C.; FOLADOR, E. L.; PEREIRA, F. L.; ALVES DORELLA, F.; GOMES LEAL, C. A.; FIORINI DE CARVALHO, A.; SILVA, A.; DE CASTRO SOARES, S.; PEREIRA FIGUEIREDO, H. C.; AZEVEDO, V.; GUIMARAES, L. C.

Genome Sequence of Corynebacterium ulcerans Strain 210932. GENOME ANNOUNCEMENTS. , v.2, p.e01233-14 - e01233-14, 2014.

21. **BENEVIDES, L. J**.; NETO, A.M.S.; DIAS, V.S; BENEVIDES,C M J; LOPES, M. V.; BRAVO, I. S. J.; OLIVEIRA, G.B.; NASCIMENTO A.S.

Low-cost artificial diets for the Mediterranean fruit fly, Ceratitis capitata Wiedemann (Diptera: Tephritidae). Magistra. , v.24, p.1 - 7, 2012.

22. SILVA NETO, ALBERTO MOREIRA DA; SANTOS, TATIANA RODRIGUES DE OLIVEIRA; DIAS, VANESSA SIMÕES; JOACHIM-BRAVO, IARA SORDI; **BENEVIDES, LEANDRO DE JESUS**; BENEVIDES, CLICIA MARIA DE JESUS; SILVA, MARIANGELA VIEIRA LOPES; DOS SANTOS, DEISE CRISTINA CAMPOS; VIRGÍNIO, JAIR; OLIVEIRA, GIVANILDO BEZERRA; WALDER, JULIO MARCOS MELGES; PARANHOS, BEATRIZ AGUIAR JORDÃO; NASCIMENTO, ANTÔNIO SOUZA DO

Mass-rearing of Mediterranean fruit fly using low-cost yeast products produced in Brazil. Scientia Agricola (USP. Impresso). , v.69, p.364 - 369, 2012.

23. NASCIMENTO A.S.; NETO, A.M.S.; **BENEVIDES, L. J**.; BRAVO, I. S. J.; BENEVIDES,C M J; LOPES,

M. V.; SANTOS, D. C. C.; VIRGINIO, J.; OLIVEIRA, G.B.

Dieta Artificial para criação massal de adultos de moscas-das-frutas (Dip.: Tephritidae) em escala industrial. Documentos (Embrapa Mandioca e Fruticultura. Impresso). , v.202, p.09 - , 2011.

Book chapters published

1. OLIVEIRA, A.; **BENEVIDES, L. J.**; MARIANO, D. C. B.; AGUIAR, E. L.; SOUSA, T. J.; SILVA, A.; AZEVEDO, V.

Bioinformatics In: A Textbook of Biotechnology.1 ed. : SM Online Publishers LLC, 2015, p. 1-.

VIII.2 Research Articles

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 June 2017
doi: 10.3389/fmicb.2017.01226

Functional Characterization of Novel *Faecalibacterium prausnitzii* Strains Isolated from Healthy Volunteers: A Step Forward in the Use of *F. prausnitzii* as a Next-Generation Probiotic

Rebeca Martín[1†], Sylvie Miquel[1,2†], Leandro Benevides[1,3], Chantal Bridonneau[1], Véronique Robert[1], Sylvie Hudault[1], Florian Chain[1], Olivier Berteau[1], Vasco Azevedo[3], Jean M. Chatel[1], Harry Sokol[1,4,5], Luis G. Bermúdez-Humarán[1], Muriel Thomas[1] and Philippe Langella[1*]

[1] Commensals and Probiotics-Host Interactions Laboratory, Micalis Institute, Institut National de la Recherche Agronomique, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, [2] Université Clermont Auvergne, Centre National de la Recherche Scientifique UMR 6023 Laboratoire Microorganismes: Génome et Environnement, Clermont-Ferrand, France, [3] Department of General Biology, Federal University of Minas Gerais, Belo Horizonte, Brazil, [4] AVENIR Team Gut Microbiota and Immunity Equipe de Recherche Labélisée (ERL), Institut National de la Santé et de la Recherche Médicale U1157/UMR7203, Faculté de Médecine Saint-Antoine, Université Pierre et Marie Curie, Paris, France, [5] Service de Gastroentérologie, Hôpital Saint-Antoine, Assistance Publique—Hôpitaux de Paris, Paris, France

Faecalibacterium prausnitzii is a major member of the Firmicutes phylum and one of the most abundant bacteria in the healthy human microbiota. *F. prausnitzii* depletion has been reported in several intestinal disorders, and more consistently in Crohn's disease (CD) patients. Despite its importance in human health, only few microbiological studies have been performed to isolate novel *F. prausnitzii* strains in order to better understand the biodiversity and physiological diversity of this beneficial commensal species. In this study, we described a protocol to isolate novel *F. prausnitzii* strains from feces of healthy volunteers as well as a deep molecular and metabolic characterization of these isolated strains. These *F. prausnitzii* strains were classified in two phylogroups and three clusters according to 16S rRNA sequences and results support that they would belong to two different genomospecies or genomovars as no genome sequencing has been performed in this work. Differences in enzymes production, antibiotic resistance and immunomodulatory properties were found to be strain-dependent. So far, all *F. prausnitzii* isolates share some characteristic such as (i) the lack of epithelial cells adhesion, plasmids, anti-microbial, and hemolytic activity and (ii) the presence of DNAse activity. Furthermore, Short Chain Fatty Acids (SCFA) production was assessed for the novel isolates as these products influence intestinal homeostasis. Indeed, the butyrate production has been correlated to the capacity to induce IL-10, an anti-inflammatory

OPEN ACCESS

Edited by:
Andrea Gomez-Zavaglia, Center for Research and Development in Food Cryotechnology (CIDCA, National Council for Scientific and Technological Research (CONICET)-Argentina-Capital), Argentina

Reviewed by:
Mireia Lopez Siles, University of Girona, Spain
Maria de los Angeles Serradell, CONICET La Plata (CCT) and Instituto de Ciencias de la Salud-UNAJ, Argentina

***Correspondence:**
Philippe Langella
philippe.langella@infra.fr

[†] *These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *03 April 2017*
Accepted: *16 June 2017*
Published: *30 June 2017*

Citation:
Martín R, Miquel S, Benevides L, Bridonneau C, Robert V, Hudault S, Chain F, Berteau O, Azevedo V, Chatel JM, Sokol H, Bermúdez-Humarán LG, Thomas M and Langella P (2017) Functional Characterization of Novel Faecalibacterium prausnitzii Strains Isolated from Healthy Volunteers: A Step Forward in the Use of F. prausnitzii as a Next-Generation Probiotic. Front. Microbiol. 8:1226. doi: 10.3389/fmicb.2017.01226

cytokine, in peripheral blood mononuclear cells (PBMC) but not to the ability to block IL-8 secretion in TNF-α-stimulated HT-29 cells, reinforcing the hypothesis of a complex anti-inflammatory pathway driven by *F. prausnitzii*. Altogether, our results suggest that some *F. prausnitzii* strains could represent good candidates as next-generation probiotic.

Keywords: probiotic, commensal, *Faecalibacterium*, molecular and metabolic characterization, immune-modulatory properties

INTRODUCTION

Despite a large number of bacteria, archaea, viruses, and unicellular eukaryotes inhabit the human body, only a few bacterial genera (*Bacteroides, Clostridium, Bifidobacterium,* and *Faecalibacterium*) predominate in the human gut microbiome (Schmidt, 2013). Nowadays it is recognized that *Faecalibacterium prausnitzii* represents around 5% from the total fecal microbiota in healthy adults (Hold et al., 2003). Furthermore, this bacterium has been proposed to be a sensor and an actor of the human intestinal health. Indeed, the levels of *F. prausnitzii* have been found to be decreased in patients suffering from intestinal and metabolic disorders such as inflammatory bowel diseases (IBD), irritable bowel syndrome (IBS), colorectal cancer (CRC), obesity, and celiac disease among others (Balamurugan et al., 2008; Sokol et al., 2008; Neish, 2009; De Palma et al., 2010; Furet et al., 2010; Rajilic-Stojanovic et al., 2011) as well as in frail elderly (van Tongeren et al., 2005). Moreover, this species may be a biomarker of choice to assist in Ulcerative colitis (UC) and Crohn's disease (CD) discrimination (Lopez-Siles et al., 2017).

F. prausnitzii has been only described in detail recently probably because it is very difficult to grow as it is an Extremely Oxygen Sensitive (EOS) bacterium (Duncan et al., 2002). Similar to other EOS bacteria, little is known about the biology of *F. prausnitzii* despite its relevance in the human gut ecosystem (Miquel et al., 2014). Most of the data referring *F. prausnitzii* are based on metagenomic studies (Miquel et al., 2013), with only few studies with isolated strains and functional approach (Duncan et al., 2002; Ramirez-Farias et al., 2009; Lopez-Siles et al., 2012; Foditsch et al., 2014). This gap between metagenomic and microbiological data is striking for microbiota-derived EOS bacteria. To reduce this gap, it is now essential to increase the knowledge of several commensal bacterial strains in order to better understand the beneficial effect of this species.

Most of the commercial probiotics do not include dominant commensal human isolates. This is a reason why these probiotic strains do not colonize the human gut and their effects persist only during a short period of time (Schmidt, 2013). Nowadays, there is an increasing interest in the use of commensal bacteria as potential probiotic agents. The reasons are multiple and the most evident is that the role of commensal bacteria in homeostatic crosstalk has started to be unraveled in the last decade (Wrzosek et al., 2013). The domestic probiotic market, with a turnover approaching $7 billion in Europe and $1.7 billion in the US in 2013 (Schmidt, 2013), is expected to grow in the next years. However, these next-generation probiotic-commensal candidates must meet the same criteria than the conventional ones. It means that they should (i) be isolated and well-characterized, (ii) achieve safety requirements, such as the acceptable resistance to antibiotics or the lack of lytic and adhesion capacities, and (iii) show beneficial effects on the host before being considerate as a probiotic. In this sense, the Food and Agriculture Organization of the United Nations (FAO) and the European Food Safe Administration (EFSA) have established several guidelines for the correct definition and evaluation of probiotics on food (FAO/WHO, 2002; Pineiro and Stanton, 2007; Binnendijk and Rijkers, 2013). Regarding *F. prausnitzii*, although little is known about its safety, there is a clear potential of this species as a next-generation probiotic. This was already proposed for livestock animals with the isolation and characterization of *F. prausnitzii* strains from stool of calves and piglets (Foditsch et al., 2014) but also for patients with intestinal dysbiosis-associated illness with the development of specific formulation keeping this EOS bacteria alive at ambient air (Khan et al., 2014). Besides, its beneficial anti-inflammatory effect has been only analyzed *in vitro* and *in vivo* with the reference strain *F. prausnitzii* A2-165 (Sokol et al., 2008) and the biofilm forming strain HTF-F (Rossi et al., 2015). As the probiotic properties are usually strain-specific ones (Pineiro and Stanton, 2007), individual studies are required to assess the anti-inflammatory properties of other *F. prausnitzii* isolated strains.

The aim of this work is to isolate a collection of novel *F. prausnitzii* strains from healthy volunteers in order to characterize them as potential probiotic bacteria in accordance with Novel Food regulatory (Miquel et al., 2015a). We have also validated the collection of viable isolated strains by metabolic and safety tests in order to better understand their biology especially in the gastrointestinal tract. Furthermore, the anti-inflammatory properties of all these strains were validated *in vitro* in order to identify the best potential *F. prausnitzii* strain to be used as a next-generation probiotic.

MATERIALS AND METHODS

Isolation of Novel Extremely Oxygen Sensitive (EOS) Strains

A cohort of healthy volunteers was first established (**Table 1**) to collect freshly emitted fecal samples used as inocula. All volunteers signed informed consent to provide the samples and an agreement of confidentiality. The complete isolation of EOS strain procedure was performed in an anaerobic chamber (N_2 = 90%, CO_2 = 5% and H_2 = 5%). Briefly, fecal samples were homogenized and serial dilutions performed in order to plate dilutions 10^{-8} and 10^{-9} on YBHI [Brain–heart infusion medium supplemented with 0.5% yeast extract (Difco)] agar

TABLE 1 | Studied cohort of healthy humans' volunteers and new *F. prausnitzii* strain identified.

Subject	Sex	Age (years)	Fecal SCFA (mM)			CFU/g	% EOS	Identified *F. prausnitzii* strains	Cultivability of the strain
			Butyrate	Propionate	Acetate				
A	M	81	nd	nd	nd	4.4×10^9	51	CNCM-I4540	Yes
B	F	59	nd	nd	nd	8.7×10^9	30.7	X	
C	M	54	nd	nd	nd	8.0×10^9	67.7	CNCM-I4541	Yes
								CNCM-I4542	Yes
								S3C12	No
								S3G1	No
D	M	54	21.7	14.8	65.4	8.0×10^9	69	X	
E	F	60	nd	nd	nd	3.5×10^{10}	40	X	
F	M	53	3.7	4.4	11.6	3.0×10^9	35.4	X	
G	F	26	3	3.9	14	2.0×10^9	37.5	X	
H	F	56	15.8	10.2	27.6	4.8×10^9	56.2	CNCM-I4574	Yes
								CNCM-I4543	Yes
I	M	59	10.1	9.5	27.3	7.8×10^9	33	S9G3	No
								S9D8	No
J	M	34	9.9	12.3	26.8	7.7×10^9	30.7	CNCM-I4644	Yes
								CNCM-I4544	Yes
								S10H3	No
K	F	60	1.1	2	5.3	2.0×10^9	28.1	X	
L	M	40	1.7	2	6.9	7.4×10^9	29.6	CNCM-I4575	Yes
								CNCM-I4573	Yes
								S13A12	No
								S13E3	No
M	F	51	2.5	3.1	10.8	4.6×10^9	53.1	CNCM-I4546	Yes

All the isolates were obtained from human fecal samples of healthy volunteers consuming omnivorous diets. F, female; M, male; nd, not determined, X, no identified F. prausnitzii strain.

supplemented with rumen fluid 20%. After 4 days of incubation at 37°C, single colonies were obtained on plates and 96 varied colonies were selected and isolated in duplicate on YHBHI supplemented with rumen fluid 20% agar plate. A group of plates was placed brought out of the anaerobic chamber for 1 h to eliminate EOS strains and after a long period of incubation (usually between 48 h and 4 days), we performed a negative screening. The EOS colonies were further re-isolated and a specific *F. prausnitzii* PCR (primers Fprau07/Fprau02) was done to identify strains of this specie (**Table 2**). Finally, a 16S rRNA gene sequencing was performed after complete 16S rRNA amplification using primer FP1 to FP5 (**Table 2**; MWG France). The viable isolates were stocked at −80°C with 16% of glycerol.

Bacterial Strains, Cell Culture, and Growth Conditions

The reference strains A2-165 (DSM 17677; Duncan et al., 2002), L2/6 (Barcenilla et al., 2000) and M21/2 (Louis et al., 2004) and the *F. prausnitzii* isolated strains (**Table 1**) were grown at 37°C in YBHI medium supplemented with cellobiose (1 mg/ml; Sigma), maltose (1 mg/ml; Sigma), and cysteine (0.5 mg/ml; Sigma) in an anaerobic chamber filled with $N_2 = 90\%$, $CO_2 = 5\%$ and $H_2 = 5\%$.

HT-29 (ATCC HTB-38) (LGC-Standars) cell line was grown in Dulbecco's Modified Eagle's minimal essential medium (DMEM) (Sigma-Aldrich) supplemented with 10% (w/v) heat-inactivated fetal bovine serum (FBS) (GibcoBRL, Eragny, France) and with penicillin G/ streptomycin (5,000 IU/mL, 5,000 μg/mL) (Sigma-Aldrich). Cultures were incubated in 25 cm^2 tissue culture flasks (Nunc, Roskilde, Denmark) at 37°C in a 5% (v/v) CO_2 atmosphere until confluence.

16S rRNA Gene Analysis

DNA was extracted from isolated colonies of the different *F. prausnitzii* strains by alkaline lysis in 50 μL of NaOH 0.5 M during 30 min and 50 μL of Tris 1M pH7 and 100 μL H_2O were added. 16S rRNA sequences were amplified using FP1 and FP2 primers (**Table 2**) and PCR products purified with the Wizard SV Gel. PCR Clean-Up system (Promega) was used to obtain bidirectional partial 16S rRNA gene sequences by using primers FP1, FP2, FP3, FP4, and FP5 (**Table 2**). All DNA sequences were confirmed by sequencing (Eurofins MWG Operon, Ebersberg, Germany). Sequences for the novel isolates were deposited in the NCBI database under the accession numbers MF185398 to MF186168.

Phylogenetic analysis based on 16S rRNA were performed using the multiple sequence alignment—CLUSTALW (Thompson et al., 1994) integrated in MEGA6 software (Tamura et al., 2013). After that, the most appropriate evolutionary model was defined and the evolutionary history was inferred

TABLE 2 | Oligonucleotides used in this study and PCR product sizes.

Primer	Oligonucleotide sequence (5′–3′)	PCR product size (bp)	Use	References
Fprau07	CCATGAATTGCCTTCAAAACTGTT	141	PCR *F. prausnitzii* specific	Sokol et al., 2008
Fprau02	GAGCCTCAGCGTCAGTTGGT			
FP1	AGAGTTTGATCCTGGCTCAG	1,474	16S rRNA complete sequence amplification and sequencing	This study
FP2	ACGGCTACCTTGTTACGACTT			
FP3	GTTGCGGGACTTAACCCAACATC		16S rRNA sequencing	This study
FP4	GTTTTTCTTGAGTAGTGCAGAGG		16S rRNA sequencing	This study
FP5	GATGTTGGGTTAAGTCCCGCAAC		16S rRNA sequencing	This study

using the Maximum likelihood (ML) criterion, based on the Kimura 2-parameter model (Kimura, 1980), with 1,000 bootstrap replicates. A discrete Gamma distribution was used to model evolutionary rate differences among sites [five categories (+G, parameter = 0.1846)]. The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 64.70% sites). Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach and all positions containing gaps and missing data were eliminated. The tree with the highest log likelihood (−3073.67) is shown (**Figure 3**). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 36 nucleotide sequences. There were a total of 1090 positions in the final dataset. In this analysis, sequences used by Lopez-Siles et al. (Duncan et al., 2002; Ramirez-Farias et al., 2009; Lopez-Siles et al., 2012) were included with the objective of compare the new strains to the two phylogroups proposed by that study. *Eubacterium desmolans* was used to root the tree.

Plasmid Presence

The presence of plasmids in the isolated strains were determined following Wizard® Plus SV Minipreps DNA Purification System (Promega) with modifications to adapt it for use with Gram positive bacteria. Briefly, an extra lysis step was performed after centrifugation of liquid overnight (ON) cultures by incubation for 1 h at 37°C with lysozyme (Sigma; 10 mg/ml) in the cell resuspension solution.

Scanning Electron Microscopy

Scanning electron microscopy analyses were performed on the MIMA2 platform (INRA, France) with pure pellet of bacterial culture suspended and fixed in 200 µL of glutaraldehyde and 3% ruthenium red during 2 h in an anaerobic chamber and stored at 4°C. Scanning electron microscopy was performed as previously reported (Joly et al., 2010).

Determination of Antibiotics Resistance

The minimum inhibitory concentrations (MIC) for 13 antibiotics (including tetracycline, kanamycin, chloranphenicol, linezolid, nupri/dalfopri, trimethoprim, gentamicin, erythromycin, cefpirome, clindamycin, streptomycin, vanomycin, and ampicillin) were determined on Wilkins-Chalgren agar (Difco) according to the E-test procedure, in accordance with the conditions recommended by the supplier (Biomerieux, France). The results were recorded after 48 h of incubation.

Anti-bacterial Assays

The anti-bacterial effect of *F. prausnitzii* supernatants were investigated *in vitro* using the bacteriocin activity assay as previously described (Ramirez-Farias et al., 2009). This anti-bacterial effect was tested on six different bacterial species: three aerobic bacteria (*E. coli* Nissle 1917, *E. coli* DH10B, and *Listeria monocytogenes* 11765), one facultative anaerobic bacterium (*Lactococcus* subsp *cremoris* MG1363), and two obligate anaerobic bacteria (*Clostridium perfringens* ATCC13124 and *Bifidobacterium infantis* DSM20088/ATCC15697). YBHI liquid medium alone was used as negative control.

Metabolic Activities

To determine the metabolic activities of the cultivable strains, API-20A galleries and the gelatin degradation test of API-20E galleries were used according to manufacturer's instructions. For detection of DNase and hemolytic activity, the strains were grown ON and then plated into Methyl green-DNA agar plates (Difco) or blood agar plates (Biomérieux) respectively. The results were recorded after 48 h of incubation. The capacity to grow in presence of mucin was assayed using a defined medium (KH_2PO_4: 5.236 g/L, $(NH_4)_2SO_4$: 4 g/L, NaCl: 4 g/L, $CaCl_2$: 30 mg/L, $MgCl_2$: 300 mg/L, $MnCl_2$: 30 mg/L, $FeCl_2$: 8 mg/L, Vitamin B12: 5 mg/L, Vitamin B1: 1 mg/L, Biotin: 1 mg/L, PABA: 1 mg/L, Folic acid: 1 mg/L, Vitamin K: 2 mg/L, cystein 0.5 mg/mL) supplemented with 1.5% mucin (Type II, Sigma-Aldrich).

Short Chain Fatty Acid (SCFA) Analysis

Supernatant concentrations of propionate, acetate, and butyrate were analyzed using gas liquid chromatography (Nelson 1020, Perkin-Elmer, St Quentin en Yvelines, France) as previously described (Lan et al., 2008). Overnight culture (20 h) of *F. prausnitzii* strains were used and culture media as negative control; each measurement for performed at least in triplicate except for fecal samples. SCFA concentrations are expressed in mM.

Dosage of D- and L-Lactate

D- and L-lactate was measured in supernatant of bacterial cultures. This supernatant was precipitated with trichloroacetic acid (10%) and centrifuged at 4,500 g for 20 min at 4°C. Lactate

was then measured in the supernatants with the Biosentec D/L lactic acid enzymatic kits according to the manufacturer instructions (Biosentec, Toulouse, France). Overnight culture (20 h) of *F. prausnitzii* strains were used and culture media as negative control; each measurement was performed at least in triplicate.

Adhesion Assays

Monolayers of HT-29 cells were seeded in 24-well tissue culture plates (Nunc) with 1.83×10^5 HT-29 cells/well and cultivated until confluence, culture medium was changed daily. Monolayers were then infected in 1 ml of the cell culture medium without antibiotics and with heat-inactivated FBS at a multiplicity of infection (MOI) of 100 bacteria per epithelial cell. After, 3 h of incubation at 37°C in anaerobic conditions (as describe above), monolayers were washed three times in phosphate-buffered saline (PBS; pH 7.2). The epithelial cells were then lysed with 1% Triton X-100 (Sigma Chemical Company, St Louis, Mo.) in water. Samples were plated onto YHBHI supplemented agar plates to determine the number of CFU corresponding to the total number of cell-associated bacteria. Adhesion to mucin has been performed as previously described by Radziwill-Bienkowska et al. (2014, 2016) Briefly, after an overnight coating of 96 plates (Nunc) with a solution of 10 mg/ml of mucin [Type III mucin from porcine stomach (lyophilized powder, Sigma-Aldrich)] a bacterial suspension ($OD_{600nm} = 1$) in PBS of each strain was incubated 3-h at 37°C in the anaerobic chamber. Bound cells were stained with crystal violet. Stained bacteria were suspended in 96% ethanol and optical density was determined at 583 nm. All the experiments were performed in triplicate. The adhesion values have been normalized using *Lactobacillus rhamnosus* GG (LGG) a positive control know by their good adhesion properties to mucin (Martin et al., 2015). Results are presented by the mean and the standard deviation.

Immuno-Modulatory Properties Using HT-29 Cells

Anti-inflammatory assays were done following the procedure described by Kechaou et al. (2012). Briefly, 50,000 HT-29 cells per well were seeded in 24-well culture plates (Nunc). Twenty-four h before bacterial co-culture (day 6), the culture medium was changed for a medium with 5% heat-inactivated FBS and 1% glutamine. On the day of co-culture, 10% of bacterial supernatant or bacterial medium (YBHI) were added in DMEM in a total volume of 500 µL. Cells were stimulated simultaneously with human TNF-α (5 ng/ml; Peprotech, NJ) for 6 h at 37°C in 10% CO_2. All samples were analyzed in triplicate. After co-incubation, cell supernatants were collected and stocked at −80°C until further analysis of interleukin-8 (IL-8) concentrations by ELISA (Biolegend, San Diego, CA). Total protein was determined by Bradford Reagent test (Sigma-Aldrich). Experiments have been done at least in triplicate. Results are expressed as IL-8/protein (pg/mg) and have been normalized using as reference value the IL-8 produced after the co-incubation with PBS as a negative control.

Experiments on Peripheral Blood Mononuclear Cells (PBMCs)

The protocol used in this study was adapted from Kechaou et al. (2012). Commercial PBMCs (StemCell Technologies, France) from five healthy donors were used in this assay. Donors presented the following characteristics: men, age under 65, body mass index <30, non-smoking, no drugs with anti-inflammatory known effects taken during the 15 days prior to sampling, and tested negative for HIV, hepatitis A and B viruses. After reception, cells were stored in liquid nitrogen until use. To prepare PBMCs for co-culture experiments with bacteria, the vial were thawed at 37°C in a water bath and then transferred into a medium containing RPMI-1640 medium supplemented with 10% heat-inactivated FCS, 1% L-glutamine and 0.1% Penicillin/Streptavidin (medium components were bought from Lonza, Switzerland). DNase (100 µg/mL, Roche Applied Science, France) was added to this mix to avoid cell clumping. Cells were then centrifuged at 200 g for 15 min, counted using trypan blue and spread on 24-well plates at 1×10^6 cells/well. Supernatants were added in triplicates (three wells per donor) at 10% in a total volume of 1 ml. Plates were incubated for 24 h at 37°C with 10% CO_2. Culture supernatant were collected, mixed with an antiprotease cocktail according to manufacturer's instructions (Complete EDTA-Free protease inhibitor, Roche Applied Bioscience) and stored at −80°C until further analysis of IL-10 concentrations by ELISA (Mabtech, Sweden).

Statistical Analysis

GraphPad software (GraphPad Sofware, La Jolla, CA, USA) was used for statistical analysis. Results are presented as bar graphs ±SEM. Comparisons were realized with the non-parametric Kruskal–Wallis test followed by a Dunn's Multiple Comparison test. Correlation test were performed using spearman test. A $p < 0.05$ was considered significant.

RESULTS AND DISCUSSION

Construction of EOS and *F. prausnitzii* Libraries

The vast majority of intestinal bacteria are EOS and thus mostly very difficult to culture (Qin et al., 2010). Although metagenomic approaches recently allow identifying some uncultivable organisms, the use of cultivable strains is requested to determine their biological activities. In this study, we report a method for isolation of novel EOS strains from human fecal samples on a complete medium (**Figure 1**). For this, a negative screening was performed through the exposition of bacterial isolates to oxygen and in parallel, these same strains were cultivated in an anaerobic chamber, which maintains a consistent anaerobic environment to ensure proper conditions for optimal EOS growth. We identified between 28.1 and 67.7% of EOS strains in the microbiota of healthy volunteers (**Table 1**). Interestingly, the proportion of EOS strains in the human microbiota was positively and significantly correlated to the amount of fecal acetate ($r = 0.7$; $p = 0.0433$) and tend to be correlated to the amount of fecal butyrate ($r =$

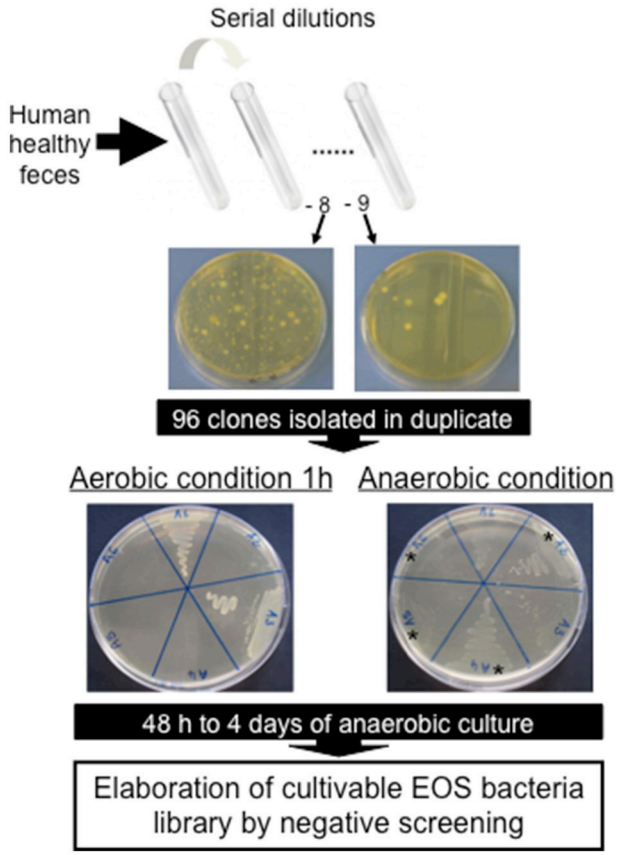

FIGURE 1 | Negative screening for isolation of new Extremely Oxygen Sensitive (EOS) strains from human healthy feces.

0.6833; $p = 0.0503$). These observations suggest that EOS population has an important metabolic impact that could participate to intestinal homeostasis (Wrzosek et al., 2013). The EOS isolates were identified by 16S rRNA gene sequencing and among them *F. prausnitzii* candidate strains were selected for further characterization. These isolation and screening set up can have a small inspecificity rate and no-*F. prausnitzii* strains can be recovered as well as the strain S13E3. After three subcultures, cultivable strains were stored at −80°C in 16% glycerol. Among 17 identified *F. prausnitzii* strains, only 10 were cultivable in the tested conditions (**Table 1**) with an $OD_{600\ nm}$ lower than 2 corresponding to $>1 \times 10^8$ CFU/mL (**Figure 2**). There was no direct correlation between CFU counts and OD_{600nm} due to difference of viability between strains. We substantially increased the number of cultured *F. prausnitzii* isolates from human origin and provided new tools for a better understanding of the diversity and microbial ecology of the colon.

Phylogenetic Diversity of *Faecalibacterium prausnitzii*

Full-length 16S rRNA gene sequences were determined for the 17 isolates of *F. prausnitzii* from healthy individuals (**Table 1**). The sequences from the literature (Barcenilla et al., 2000; Duncan et al., 2002; FEEDAP, 2012; Lopez-Siles et al., 2012) were included in order to classify the new isolates in the two phylogroups proposed by Lopez-Siles et al. (Barcenilla et al., 2000; Duncan et al., 2002; FEEDAP, 2012; Lopez-Siles et al., 2012; **Figure 3**). Each of these 16S rRNA sequences were unique, came from a different colony, and share >97% 16S rRNA sequences similarity. Cultivability of strains was not linked to phylogroups affiliation (**Figure 3**). Of note, all strains have a similar morphotype with cell wall extensions, like "swellings" (**Figure 4**) already described but with yet unknown function (Miquel et al., 2013). The average nucleotide identity between strains of the two phylogroups (S3L/3 and L2/6 = 94%) supports the hypothesis of the existence of two genomospecies without phenotypic properties defined yet (Lopez-Siles et al., 2017). Although, as was previously described for another library, there was a tendency for some sequences to group by isolation and individual with a clustering of strains (subgroup B of the phylogroup II; Lopez-Siles et al., 2012). For example, CNCM I-4574 and CNCM I-4543 strains were isolated from the same volunteer and present 99.8% of homology at 16S rRNA level.

Interestingly, the existence of strains that do not fit in any phylogroup (as CNCM I-4541) suggest that biodiversity of *F. prausnitzii* remains poorly known, maybe since only few strains have been isolated. Moreover, the strain S13E3, could be not an *F. prausnitzii* stain.

Resistance to Antibiotics

The MIC for the different antibiotics tested are represented in the **Table 3**. Concerning the breakpoints for Gram positive bacteria from EFSA (Duncan et al., 2004) which classify bacteria as resistant or not to a specific antibiotic, all *F. prausnitzii* isolates were susceptible to clindamycin, vancomycin, ampicillin, quinupristin+dalfopristin, and chloramphenicol (MICs lower than 0.25, 2, 1, 0.5, and 2 mg/L respectively). Only one isolate, the CNCM I-4541 strain was resistant to erythromycin (MICs > 0.5mg/L). Surprisingly, all tested strains were resistant to streptomycin (MICs ranging from 14 to 50 mg/L) excepted for the CNCM I-4575 isolate. Regarding gentamicin, kanamycin, and tetracycline, different results were obtained for the different isolates: with up to 5 isolates displaying resistance to higher concentrations of the tested antibiotics than the determined breakpoint. Finally, three antibiotics (not included in the EFSA guidance) were also analyzed due to their importance in the clinical treatments: trimetroprim, linezolid, and cefpirome. All strains were resistant to trimethoprim, as expected for an anaerobic bacteria (MICs >32 mg/L; data not shown), while they tended to be susceptible to linezolid (MICs ranging from 0.032 to 3.3 mg/L) and resistant to cefpirome (from 4.66 to >256 mg/L) which, when linked to the general susceptibility to ampicillin, might indicate that the penicillin binding proteins of *Faecalibacterium* are poorly recognized by cephalosporins. Remarkably, CNCM I-4543 and CNCM I-4574 isolates were resistant to cefpirome, a fourth-generation cephalosporin

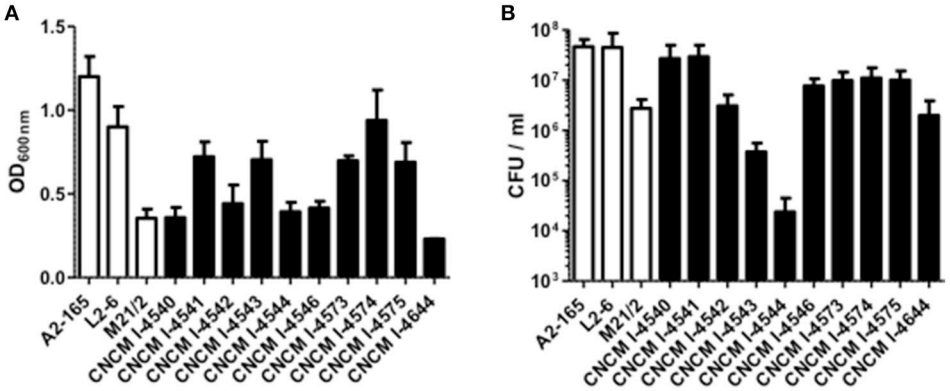

FIGURE 2 | Growth profile of *F. prausnitzii* strains. **(A)** $OD_{600\,nm}$ determination after 20 h growth in YBHI supplemented medium and **(B)** determination of viable bacteria: the CFU/mL numeration in the same cultures. Each measurement have been done at least in triplicate.

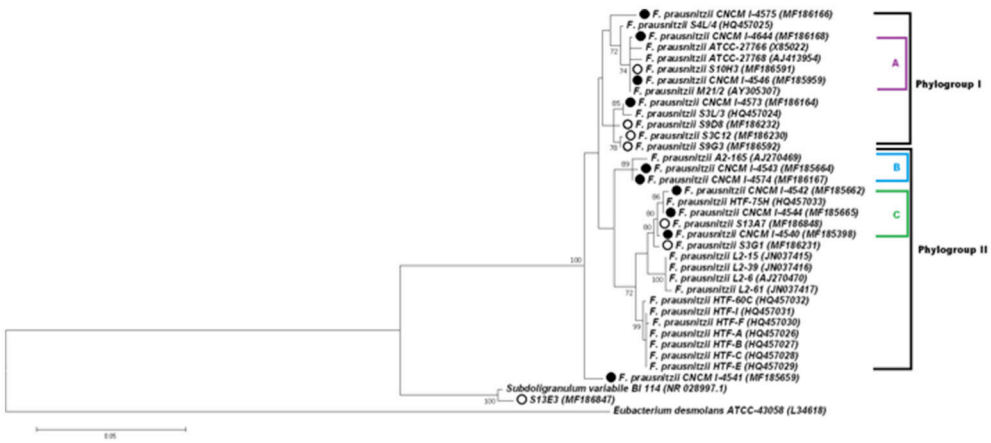

FIGURE 3 | Phylogenetic tree of *F. prausnitzii* strains based on 16S rRNA gene sequences. The tree was constructed with the MEGA6 software package using the Maximum Likelihood method. The bootstrap values above 70% are shown next to the branches. The *F. prausnitzii* isolates incorporated in this study have circles besides. The black circles represent the cultured strains and white circles represent uncultured isolates. Colors (purple, blue, and green) and letters (A, B, and C) indicate the tree groups with high bootstrap values, formed by our cultured strains.

stable against most plasmid- and chromosome-mediated beta-lactamases (Wiseman et al., 1997), with a MIC higher than 256 mg/L.

The analysis of antimicrobial resistance is of major importance due to the fast evolution of antibiotic resistance in response to the extensive use of antimicrobials. However, the microbiological breakpoints marked by the EFSA for most of Gram positive bacteria is probably not the most correct for the analysis of *F. prausnitzii* isolates as no so many information about their natural or acquired resistance patters is reported, to our knowledge, up to day in the literature. Foditsch et al. (2014) have identify that more of the 50% of the *F. prausnitzii* strains that they isolated from fecal samples of healthy calves and piglets were resistant to tetracycline, amikacin, cefepime, and cefoxitin comparing the MIC values with the standard values determined by CLSI for *Bacteroides fragilis* ATCC 25285. This fact highlights the need of more microbiological studies of antibiotic resistance in this species in order to determine a correct standard values for *Faecalibacterium* as well as the search for genes codifying for the most important resistance mechanisms for, at least, some of the antibiotics tested in this study.

Metabolic Activities

Enzymatic activities detected by API-20A gallery system are reported in **Table 4**. Interestingly, only one enzyme was detected and active in all the tested strains: the beta-galactosidase. Otherwise, all the strains were not able to ferment mannose or raffinose, to reduce nitrate and to produce indole (data not shown). Furthermore, all the isolates were negative for the presence of urease, arginine dihydrolase, beta-glucosidase, alpha-arabinosidase, N-acetyl-beta-glucosaminidase, glutamic acid decarboxylase, alkaline phosphatase, phenylalanine arylamidase, leucine arylamidase, pyroglutamic acid arylamidase, tyrosin arylamidase, alanine arylamidase, glutamyl glutamic acid arylamidase, and serin arylamidase (data not shown). These results confirm previous observations where no strain was able

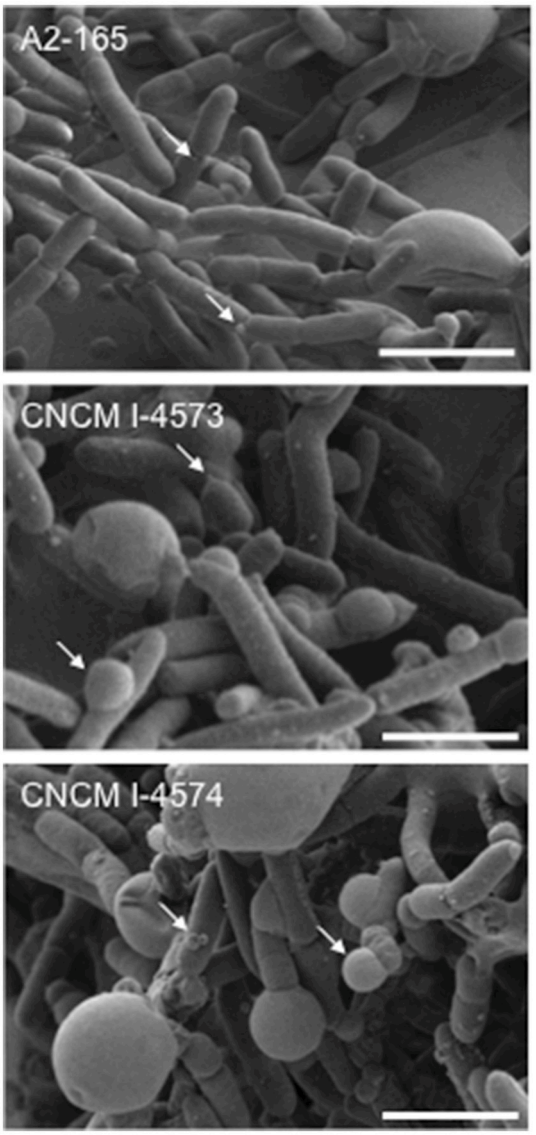

FIGURE 4 | Scanning electron microscopy images of *F. prausnitzii* strains phylogroup II. Strains were grown in YBHI liquid medium 20 h. Scale bars indicate 2 μm. Arrows indicates "swelling."

to metabolize arabinose and raffinose among others as the sole energy source (Duncan et al., 2002; Lopez-Siles et al., 2012).

For all the other enzymes (6 phospho-beta galactosidase, alpha-glucosidase, beta-glucuronidase, arginine arylamiase, leucyl glycerine-arylamidase, glycine-arylamidaseycine, and histidine-arylamidase), differences inter-strains were detected (**Table 4**). Beta-glucuronidase activity has been previously reported in some *F. prausnitzii* isolates (Lopez-Siles et al., 2012). While six strains showed individual profiles, the other seven are included in three different profiles. Two of them corresponds to the group A from phylogroup I (CNCM I-4546 and M21/2). The strains CNCM I-4543 and CNCM I-4574 (group B, phylogroup II), which are the only ones resistant to cefpirome, share also the same metabolic profile and donor. And the third metabolic

TABLE 3 | Minimum inhibitory concentrations (MIC) (mg/L) for the different antibiotics tested.

EFSA Breakpoint (other Gram +)	**GEN**	**STR**	**KM**	**ERY**	**CLI**	**VAN**	**TET**	**QD**	**CM**	**AMP**	**CPO**	**LZD**
	4	**8**	**16**	**0.5**	**0.25**	**2**	**2**	**0.5**	**2**	**1**	**nd**	**nd**
A2-165	1.37 ± 0.12	**96 ± 18.47**	**8.67 ± 1.76**	0.20 ± 0.08	0.016 ± 0	0.27 ± 0.05	0.016 ± 0	0.03 ± 0.01	0.08 ± 0.02	0.11 ± 0.02	9 ± 1	1.25 ± 0.25
L2-6	**4.33 ± 0.88**	**32 ± 7.15**	0.42 ± 0.08	0.10 ± 0.02	0.023 ± 0	0.47 ± 0.12	**4.21 ± 0.62**	0.58 ± 0.46	**20 ± 5.66**	0.06 ± 0.03	24 ± 4.62	0.62 ± 0.12
M21/2	**4.75 ± 1.49**	**21 ± 4.43**	**51 ± 45**	0.05 ± 0.01	0.016 ± 0	0.25 ± 0	1.09 ± 0.63	0.016 ± 0	0.15 ± 0.03	0.04 ± 0.03	16.67 ± 7.86	0.75 ± 0
CNCM I-4540	1.25 ± 0.25	**21.33 ± 2.67**	**122.67 ± 66.83**	0.11 ± 0.02	0.016 ± 0	0.71 ± 0.18	0.016 ± 0	0.011 ± 0.004	0.05 ± 0.04	0.11 ± 0.02	24 ± 4.62	0.04 ± 0.06
CNCM I-4541	1.62 ± 0.47	**24 ± 0**	**14 ± 2**	**20 ± 4**	0.016 ± 0	0.125 ± 0	**8 ± 0**	0.27 ± 0.24	0.17 ± 0.07	0.06 ± 0	40 ± 8	1.25 ± 0.25
CNCM I-4542	2.87 ± 1.12	**23.67 ± 3.26**	**20 ± 4**	0.11 ± 0.07	0.016 ± 0	0.67 ± 0.17	0.016 ± 0	0.016 ± 0	0.29 ± 0.40	0.08 ± 0.03	20.67 ± 7.69	3.17 ± 1.42
CNCM I-4543	2 ± 0.40	**16 ± 0**	**20 ± 4**	0.08 ± 0.02	0.016 ± 0	0.29 ± 0.04	0.2 ± 0.002	0.07 ± 0.03	0.31 ± 0.09	0.12 ± 0	**≥256 ± 0**	1.25 ± 0.38
CNCM I-4544	**6 ± 1.15**	**21.2 ± 7.31**	**100 ± 52.41**	0.12 ± 0	0.016 ± 0	0.92 ± 0.08	0.016 ± 0	0.023 ± 0	0.11 ± 0.02	0.09 ± 0.02	53.33 ± 5.33	0.5 ± 0
CNCM I-4546	**10 ± 66**	**50.67 ± 13.33**	**256 ± 0**	0.22 ± 0.61	0.018 ± 0.002	0.56 ± 0.31	**2.6 ± 0.6**	0.16+0.07	1.25 ± 1.51	0.07 ± 0.02	24.8 ± 6.24	3 ± 1
CNCM I-4573	**7 ± 1**	**32 ± 0**	**234 ± 21.33**	0.01 ± 0.03	0.024 ± 0.007	0.28 ± 0.05	0.028 ± 0.003	0.04 ± 0.005	0.38 ± 0	0.25 ± 0	22 ± 3.83	3.3 ± 0.8
CNCM I-4574	1.75 ± 0.25	**14 ± 0**	6 ± 2	0.07 ± 0.02	0.016 ± 0	0.25 ± 0	0.016 ± 0	0.03 ± 0.01	0.09 ± 0.02	0.22 ± 0.03	**≥256 ± 0**	0.5 ± 0.14
CNCM I-4575	1.25 ± 0.25	5 ± 1	4 ± 1	0.07 ± 0.01	0.016 ± 0	0.5 ± 0	0.032 ± 0	0.03 ± 0.03	0.016 ± 0	0.084 ± 0.02	4.67 ± 1.67	0.03 ± 0.01
CNCM I-4644	0.91 ± 0.08	**9.33 ± 1.33**	**135 ± 69.84**	0.10 ± 0.05	0.026 ± 0.01	0.23 ± 0.02	0.83 ± 0.32	0.04 ± 0.01	0.079 ± 0.04	0.026 ± 0.01	9.333 ± 1.33	0.58 ± 0.08

Gentamicin (GEN), streptomycin (STR), kanamycin (KM), erythromycin (ERY), clindamycin (CLI), vancomycin (VAN), tetracycline (TET), quinupristin/dalfopristin (QD), chloramphenicol (CM), ampicillin (AMP), cefpirome (CPO), and linezolid (LZD). Experiments have been done in triplicate and the results are expressed as the media ± SEM. nd, Not defined. In bold, Resistances.

TABLE 4 | Metabolic capacities of *F. prausnitzii* strains detected by API 32A galleries.

	bGal	bGP	αGLU	bGUR	ArgA	Lga	GlyA	HisA
	beta−galactosidase	beta Galactosidase 6 phosphate	Alpha glucosidase	beta glucuronidase	Arginine Arylamidase	Leucyl Glycine Arylamidase	Glycine Arylamidaseycine	Histidine Arylamidase
A2−165	+	+	+	+	+	+	+	+
L2−6	+	−	+	+	+	+	+	+
M21/2	+	−	−	−	+	−	+	+
CNCM I−4540	+	+	−	−	−	−	−	−
CNCM I−4541	+	+	−	−	+	−	−	−
CNCM I−4542	+	+	−	−	−	−	−	−
CNCM I−4543	+	+	−	+	+	+	+	−
CNCM I−4544	+	−	−	−	−	−	−	−
CNCM I−4546	+	−	−	−	+	−	+	+
CNCM I−4573	+	−	−	−	+	−	+	+
CNCM I−4574	+	+	−	+	+	+	+	−
CNCM I−4575	+	+	−	+	+	−	+	+
CNCM I−4644	+	−	−	−	+	+	+	+

+, Presence; −, absence. Experiments have been done in triplicate.

profile is shared by strains CNCM I-4540 and CNCM I-4542 that belong to the group C of phylogroup II.

It is now well-establish that *F. prausnitzii* is an acetate-consumer and butyrate-producer species (Duncan et al., 2002; Lopez-Siles et al., 2012). Here, we report that in pure cultures, our new isolated strains are also able to produce butyrate and this production is significantly and positively correlated to their growth (OD_{600nm}; $r = 0.8462$; $p = 0.003$; **Figures 5A,B**). It is interesting to highlight that the production level of butyrate was not linked to a particular phylogroup (Phylogroup I 3.91 mM ± 0.43 and Phylogroup II 4.89 mM ± 0.62). Moreover, all strains could metabolize acetate present in the culture medium at around the same level (**Figure 5A**). This consumption was not directly correlated to bacterial growth ($r = -0.3132$, $p = 0.2975$) and tended to be more correlated to butyrate production ($r = -0.544$, $p = 0.0546$). This observation is in agreement with the literature which describes that most of the carbon present in the butyrate produced (around 85%) is derived from external acetate, with only 15% provided directly from glucose (FEEDAP, 2012).

F. prausnitzii can also produce a few amount of D-lactate (FEEDAP, 2012). Indeed, among our strain collection, no L-lactate was detected and only few amounts of D-lactate were detected (1.09 mM ± 0.15 and 1.07 mM ± 0.39 phylogroup I and II respectively; data not shown). This production, not correlated with phylogroup affiliation, was correlated to the OD_{600nm}($r = 0.6209$, $p = 0.0235$). Bacterial D-Lactate production can be viewed as harmful since accumulation of this metabolite into the blood may be neurotoxic and leads to acidosis (Mack, 2004). In particular, humans with short bowel syndrome (in which small intestine has been surgically removed), the D/L fecal lactate ratio seems to be the most relevant index with a higher D-encephalopathy risk (Mayeur et al., 2013). However, in healthy adults, there is no lactate detectable in fecal samples, because lactobacilli (main producer of D-lactate) are minor groups in microbiota and lactate is degraded by other major bacterial groups (36, He, 2008 #41). This observation also suggested that the weak production of D-lactate by *F. prausnitzii* strains, major component of the microbiota, could not have metabolic deleterious impact on the host.

All strains were unable to growth in the presence of mucin as the only carbon source in a defined medium (data not shown). This data agrees with previous results where no evidence of fermentation of porcine gastric mucin by *F. prausnitzii* was detected (Lopez-Siles et al., 2012). Nevertheless, SCFA concentrations and OD_{600nm} measures taken after 2 days of incubation showed the ability of the different strains to survive but metabolically inactive as it could be deduced by the absence of butyrate in the supernatants of the cultures and the almost minimal OD_{600nm} recorded (data no shown). A decrease in butyrate production due to non-optimal growth conditions have been already reported for *F. prausnitzii* A2-165 strain (Lopez-Siles et al., 2012). This characteristic pointed out the intrinsic growth requirements of this species which, in addition to be an EOS, needs strain specific nutritional environment and has the ability to switch between substrates derived from the diet or the host (Lopez-Siles et al., 2017).

Lytic Activities

Gelatin is a heterogeneous mixture of water-soluble protein that is usually used in microbiological procedures to detect the presence of proteolytic activities. None of the strains were able to degrade gelatin in the conditions recommended by the API gallery supplier (data not shown). However, when the strains were inoculated in the gallery in a defined medium instead of API suspension medium, they were able to degrade partially this compound after 3 days of incubation. This fact suggests that the strains are able to hydrolyze gelatin although, maybe due to the growth limitations present in this culture media, the existence of this compound is not enough to allow the metabolic development of this activity in the strains.

The presence of hemolytic activity was tested using blood agar plates. None of the strains showed hemolytic activity under the conditions tested. In contrast, all the strains reveal a DNAse activity in green methyl-DNA medium (data not

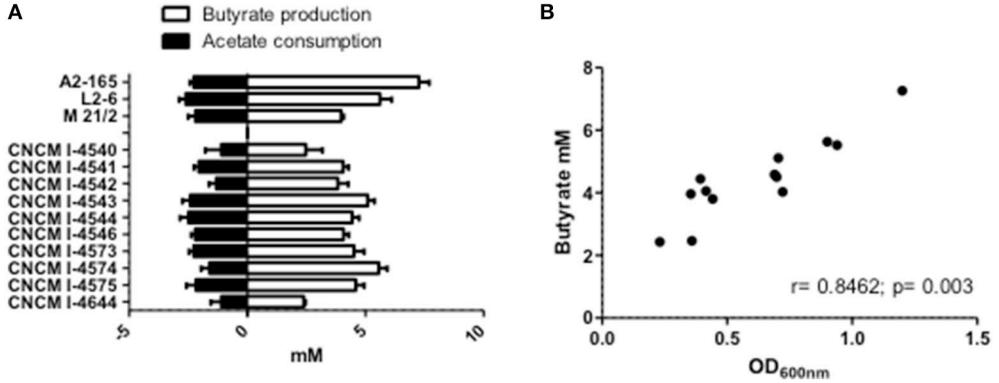

FIGURE 5 | SCFA metabolism of *F. prausnitzii* strains *in vitro*. **(A)** Acetate and butyrate concentrations after 20 h growth of strain in YBHI supplemented medium. The concentration of control media was subtracted for each measurement have been done at least in triplicate. **(B)** Correlation between $OD_{600\,nm}$ and butyrate production.

shown). Furthermore, the presence of a magnesium dependent DNase activity has been previously reported in at least three of five strains already sequenced [A2-165 (gi:257439194), SL3/3 (gi:295105207), and L2/6 (gi:295102777)].

The presence of these extracellular activities is often linked to a virulence status in some bacterial species such as *Enterococcus* spp. (Eaton and Gasson, 2001). However, these factors also contribute to the survival of microorganisms in the mammalian gut being characteristic of several members of the natural microbiota (Sanders et al., 2010). This can be the case of *Faecalibacterium* isolates, which are extremely well-adapted to the gut environment (Lopez-Siles et al., 2012).

Antibacterial Activities

We investigated antibacterial properties of *F. prausnitzii* supernatants, using the bacteriocin activity assay. We did not reveal any antibacterial effect on several anaerobic and aerobic bacterial species under the conditions tested. This fact is a desirable characteristic of a strain to be considered as a probiotic candidate.

Ability to Stimulate the Immune Response

The reference strain *F. prausnitzii* A2-165 is well-known for its immuno-modulatory properties and more specifically for its anti-inflammatory effects both *in vitro* and *in vivo* in different murine models of colitis (Sokol et al., 2008; Martin et al., 2014). To determine whether the newly isolated *F. prausnitzii* strains are able to modulate the immune response, we tested *in vitro* the immuno-modulatory properties of the supernatants from all the isolates in two different cellular models: HT-29 and PBMC. The first one is based on the capacity to block IL-8 production (a pro-inflammatory cytokine) induced by TNF-α stimulation in HT-29 epithelial cells and the second is based on the stimulation of PBMC cells and the measure of the anti-inflammatory cytokine IL-10. As shown in **Figure 6A**, all the strains tend to decrease IL-8 concentrations. However, this decrease was not equivalent in all the strains and does not correlate either with growth ratio ($r = -0.2857$, $p = 0.344$) or butyrate production ($r = -0.3357$, $p = 0.2869$).

For the PBMC assay, although all the strains tend to increase the production of IL-10 cytokine, only four strains (two controls and two new isolates from this study) were able to induce statistically significant increase production of this cytokine (**Figure 6B**) The two most performing strains (A2-165 and 4543) belong to the phylogroup II, group B. Notably, the IL-10 production was correlated with both growth ratio ($r = 0.6813$, $p = 0.0103$) and butyrate production ($r = -0.6923$, $p = 0.0126$). This different phenotype may suggest the presence of different molecule(s) responsible of the anti-inflammatory effects *in vitro*. The anti-inflammatory properties of butyrate have been already reported in the literature (Fusunyan et al., 1999; Kamitani et al., 1999) and its ability to block IL-8 production under the conditions tested in this study were confirmed *in vitro* in similar concentrations to those founds in *F. prausnitzii* supernatants (data not shown). However, its role remains controversial as its effects seems to be dose- and time-dependent as well as depended on the cellular model used (Martin et al., 2013). For instance, regarding cells from intestinal origin, butyrate has been found to decrease the secretion of IL-8 in Caco-2 and HIPEC cells and, in contrast to this study, to enhance IL-8 production in HT-29 and HT-29 MTX cells (Bocker et al., 2003).

However, several authors have found different candidate molecules/structures responsible for *F. prausnitzii* anti-inflammatory effects. MAM protein, found in *F. prausnitzii* supernatant, has been found to block NF-κB activation and the production of the pro-inflammatory cytokine IL-8 (Quevrain et al., 2016). *F. prausnitzii* is also able to produce bioactive anti-inflammatory molecules such as shikimic and salicylic acids (Miquel et al., 2015b). Besides, Rossi and co-workers showed the ability of *F. prausnitzii* strain HTF-F and its extracellular polymeric matrix to develop immunomodulatory effects through the TLR2 dependent modulation of IL-12 and IL-10 cytokine production in human monocyte-derived dendritic cells (Rossi et al., 2015) and *F. prausnitzii* has been found to be a strong inducer of regulatory T cells secreting IL-10 (Sarrabayrouse et al., 2014). All these results point out the complex anti-inflammatory mechanisms underlying this species.

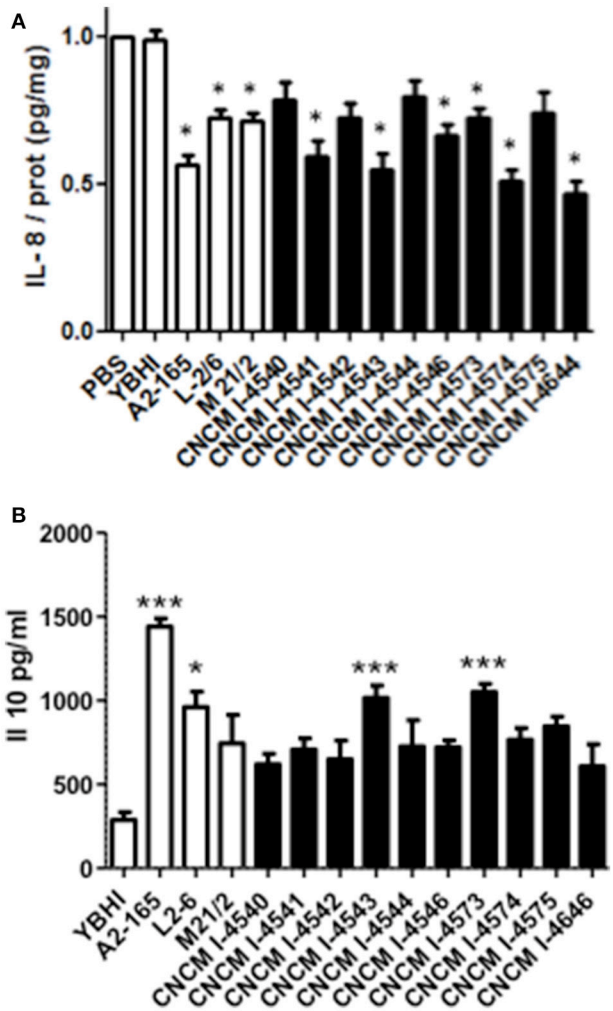

FIGURE 6 | Immuno-modulation capacities of *F. prausnitzii* strains *in vitro*. **(A)** IL-8 production in HT-29 TNF-α stimulated cells. Experiments have been done at least in triplicate. Results are expressed as IL-8/ protein (pg/mg) and have been normalized using as reference value the IL-8 produced after the co-incubation with PBS as a negative control. **(B)** IL-10 production in peripheral blood mononuclear cells. Experiments have been done at least in triplicate. Results are expressed as IL-10 concentration (pg/mL). Significant differences from the control (YBHI) was specified as: $*p < 0.05$ and $***p < 0.001$.

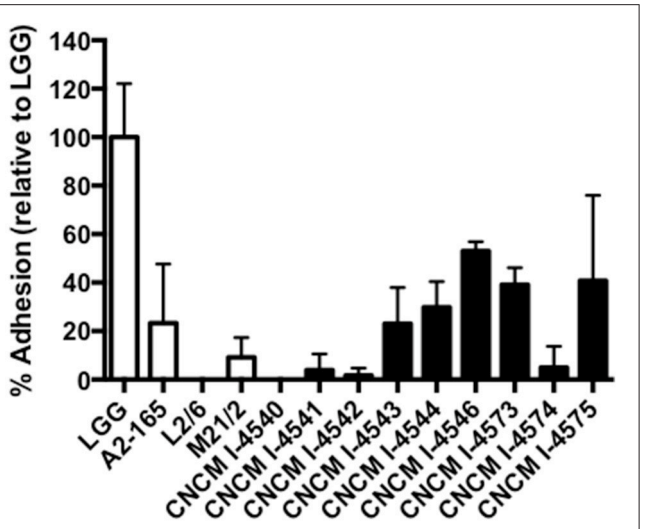

FIGURE 7 | Adhesion to mucin of *F. prausnitzii* strains. Experiments have been done in triplicate. The adhesion values have been normalized using *Lactobacillus rhamnosus* GG (LGG) a positive control know by their good adhesion properties to mucin (50). Results are presented by the mean and the standard deviation.

Adhesion to Epithelial Cells *In vitro*

In parallel, we also sought for the adhesion capacities of the new *F. prausnitzii* isolates to the intestinal epithelial cells HT-29 and mucin. All the tested strains were not able to adhere to HT-29 cells *in vitro* (data not shown) in anaerobic conditions. Regarding mucin, some of the strains were able to adhere to this compound after 3 h of incubation in the anaerobic chamber (**Figure 7**), Even if our conditions were not representative of physiological conditions (death of our eukaryotic cells), this result gives ecological clues about the processes of colonization of the gastro-intestinal tract by *F. prausnitzii*. In fact, this species is a late but major commensal colonizer of the gut which implantation requires a likely copro-cooperation maybe for the establishment of a trophic chain (Wrzosek et al., 2013).

CONCLUDING REMARKS

The development of new probiotic products containing human isolated strains with beneficial properties for the host requires the development of new techniques in order to: (i) isolate strains belonging to the major groups of the intestinal microbiota, (ii) determinate their safe status and (iii) infer in their potential beneficial effects. This study meets these entire three requests. Work with anaerobic and more precisely EOS bacteria are a prerequisite to succeed in the isolation of representative strains that can impact on intestinal homeostasis. For this reason, in this study, we have used a new procedure to isolate EOS strains from feces that has enabled us to build a collection of *F. prausnitzii* strains. The lack of knowledge about this species prompts us to further analyze their genetic diversity by comparing the new isolates with those already available in the databases. This has allowed us to point out the high diversity of our collection ranged on two different phylogroups with different clusters. *F. prausnitzii* strain genomes should be established or/and a metabolic comparison of several strains in the same culture conditions whether the phylogroups belong to genomovars or genomospecies.

Regarding safety concerns, this study is the first step toward a better understanding of *F. prausnitzii* properties. Up to date, little was known about *F. prausnitzii* resistance to antibiotics, lytic activities or adhesion properties. Here, we have shown for the first time the profile of all these characteristics in a collection of human *Faecalibacterium* strains. A positive remark is that all the strains were not antibacterial producers, not hemolytic and weak

producer of D-lactate. Furthermore, although some of the strains were able to adhere to mucin, this trait can be considered as factor favoring durable implantation and a highly effective probiotic (Miquel et al., 2015a). However, further analyses are required to better determine the presence of acquired or natural resistances as well as to distinguish between the pathogenic or adaptative nature of some of the properties detected such as the presence of DNase activity.

Finally, the anti-inflammatory properties of all the strains have been analyzed. There is a well-known correlation between *F. prausnitzii* dysbiosis and a large set of human diseases such as IBD and IBS (Miquel et al., 2013). Recent studies using *F. prausnitzii* strains in *in vivo* models provide arguments concerning its beneficial effect on the host (Sokol et al., 2008; Wrzosek et al., 2013; Martin et al., 2014). The presence of the anti-inflammatory properties of these strains also opens the possibility to test them in murine models in order to further determine their beneficial effects before testing them in human clinical trials.

AUTHOR CONTRIBUTIONS

RM, SM, JC, HS, LGBH, MT, and PL participate in the design of the project. RM, SM, JC, HS, OB, VA, LGBH, MT, and PL designed the experiments. RM, SM, LB, CB, VR, SH, and FC performed the experiments and analysis. RM and SM draft the manuscript. VR, CB, FC, JC, HS, LGBH, MT, and PL revised the manuscript critically. All the authors have read and approved the last version of the manuscript.

FUNDING

This paper was a part of FPARIS collaborative project selected and supported by the Vitagora Competitive Cluster and funded by the French FUI (Fond Unique Interministériel; FUI: n°F1010012D), the FEDER (Fonds Européen de Développement Régional; Bourgogne: 34606), the Burgundy Region, the Conseil Général 21 and the Grand Dijon. This work was also supported by Merck Médication Familiale (Dijon, France) and Biovitis (Saint Etienne de Chomeil, France). RM and SM receive a salary from the same grants.

ACKNOWLEDGMENTS

Authors thank Prof. Juan Evaristo Suárez for the critical reading of the manuscript and Stéphanie Courau and Pascal Molimard for fruitful discussions during the project. We gratefully acknowledge T. Meylheuc for scanning electron microscopy (MIMA2 platform, INRA, France) and Harry Flint for the reference *Faecalibacterium* strains.

REFERENCES

Balamurugan, R., Rajendiran, E., George, S., Samuel, G. V., and Ramakrishna, B. S. (2008). Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, Desulfovibrio and *Enterococcus faecalis* in the feces of patients with colorectal cancer. *J. Gastroenterol. Hepatol.* 23(8 Pt 1), 1298–1303. doi: 10.1111/j.1440-1746.2008.05490.x

Barcenilla, A., Pryde, S. E., Martin, J. C., Duncan, S. H., Stewart, C. S., Henderson, C., et al. (2000). Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* 66, 1654–1661. doi: 10.1128/AEM.66.4.1654-1661.2000

Binnendijk, K. H., and Rijkers, G. T. (2013). What is a health benefit? An evaluation of EFSA opinions on health benefits with reference to probiotics. *Benef. Microbes* 4, 223–230. doi: 10.3920/BM2013.0019

Bocker, U., Nebe, T., Herweck, F., Holt, L., Panja, A., Jobin, C., et al. (2003). Butyrate modulates intestinal epithelial cell-mediated neutrophil migration. *Clin. Exp. Immunol.* 131, 53–60. doi: 10.1046/j.1365-2249.2003.02056.x

De Palma, G., Nadal, I., Medina, M., Donat, E., Ribes-Koninckx, C., Calabuig, M., et al. (2010). Intestinal dysbiosis and reduced immunoglobulin-coated bacteria associated with coeliac disease in children. *BMC Microbiol.* 10:63. doi: 10.1186/1471-2180-10-63

Duncan, S. H., Hold, G. L., Harmsen, H. J., Stewart, C. S., and Flint, H. J. (2002). Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 52(Pt 6), 2141–2146. doi: 10.1099/00207713-52-6-2141

Duncan, S. H., Holtrop, G., Lobley, G. E., Calder, A. G., Stewart, C. S., and Flint, H. J. (2004). Contribution of acetate to butyrate formation by human faecal bacteria. *Br. J. Nutr.* 91, 915–923. doi: 10.1079/BJN20041150

Eaton, T. J., and Gasson, M. J. (2001). Molecular screening of Enterococcus virulence determinants and potential for genetic exchange between food and medical isolates. *Appl. Environ. Microbiol.* 67, 1628–1635. doi: 10.1128/AEM.67.4.1628-1635.2001

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) (2012). Guidance on the assessment of bacterial susceptibility to antimicrobials of human and veterinary importance. *EFSA J.* 10:2740. doi: 10.2903/j.efsa.2012.2740

FAO/WHO (2002). *FAO/WHO and Agriculture Organization Guidelines for the Evaluation of Probiotics in Food*. London: Food and Agriculture Organization, 11.

Foditsch, C., Santos, T. M., Teixeira, A. G., Pereira, R. V., Dias, J. M., Gaeta, N., et al. (2014). Isolation and characterization of *Faecalibacterium prausnitzii* from calves and piglets. *PLoS ONE* 9:e116465. doi: 10.1371/journal.pone.0116465

Furet, J. P., Kong, L. C., Tap, J., Poitou, C., Basdevant, A., Bouillot, J. L., et al. (2010). Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes* 59, 3049–3057. doi: 10.2337/db10-0253

Fusunyan, R. D., Quinn, J. J., Fujimoto, M., MacDermott, R. P., and Sanderson, I. R. (1999). Butyrate switches the pattern of chemokine secretion by intestinal epithelial cells through histone acetylation. *Mol. Med.* 5, 631–640.

Hold, G. L., Schwiertz, A., Aminov, R. I., Blaut, M., and Flint, H. J. (2003). Oligonucleotide probes that detect quantitatively significant groups of butyrate-producing bacteria in human feces. *Appl. Environ. Microbiol.* 69, 4320–4324. doi: 10.1128/AEM.69.7.4320-4324.2003

Joly, F., Mayeur, C., Bruneau, A., Noordine, M. L., Meylheuc, T., Langella, P., et al. (2010). Drastic changes in fecal and mucosa-associated microbiota in adult patients with short bowel syndrome. *Biochimie* 92, 753–761. doi: 10.1016/j.biochi.2010.02.015

Kamitani, H., Ikawa, H., Hsi, L. C., Watanabe, T., DuBois, R. N., and Eling, T. E. (1999). Regulation of 12-lipoxygenase in rat intestinal epithelial cells during differentiation and apoptosis induced by sodium butyrate. *Arch. Biochem. Biophys.* 368, 45–55. doi: 10.1006/abbi.1999.1284

Kechaou, N., Chain, F., Gratadoux, J. J., Blugeon, S., Bertho, N., Chevalier, C., et al. (2012). Identification of one novel candidate probiotic *Lactobacillus plantarum* strain active against influenza virus infection in mice by a large-scale screening. *Appl. Environ. Microbiol.* 79, 1491–1499. doi: 10.1128/AEM.03075-12

Khan, M. T., van Dijl, J. M., and Harmsen, H. J. (2014). Antioxidants keep the potentially probiotic but highly oxygen-sensitive human gut bacterium *Faecalibacterium prausnitzii* alive at ambient air. *PLoS ONE* 9:e96097. doi: 10.1371/journal.pone.0096097

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Lan, A., Bruneau, A., Bensaada, M., Philippe, C., Bellaud, P., Rabot, S., et al. (2008). Increased induction of apoptosis by *Propionibacterium freudenreichii* TL133 in colonic mucosal crypts of human microbiota-associated rats treated with 1,2-dimethylhydrazine. *Br. J. Nutr.* 100, 1251–1259. doi: 10.1017/S0007114508978284

Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J., and Martinez-Medina, M. (2017). *Faecalibacterium prausnitzii*: from microbiology to diagnostics and prognostics. *ISME J.* 11, 841–852. doi: 10.1038/ismej.2016.176

Lopez-Siles, M., Khan, T. M., Duncan, S. H., Harmsen, H. J., Garcia-Gil, L. J., and Flint, H. J. (2012). Cultured representatives of two major phylogroups of human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl. Environ. Microbiol.* 78, 420–428. doi: 10.1128/AEM.06858-11

Louis, P., Duncan, S. H., McCrae, S. I., Millar, J., Jackson, M. S., and Flint, H. J. (2004). Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J. Bacteriol.* 186, 2099–2106. doi: 10.1128/JB.186.7.2099-2106.2004

Mack, D. R. (2004). D(-)-lactic acid-producing probiotics, D(-)-lactic acidosis and infants. *Can. J. Gastroenterol.* 18, 671–675. doi: 10.1155/2004/342583

Martin, R., Chain, F., Miquel, S., Lu, J., Gratadoux, J. J., Sokol, H., et al. (2014). The commensal bacterium *Faecalibacterium prausnitzii* is protective in DNBS-induced chronic moderate and severe colitis models. *Inflamm. Bowel Dis.* 20, 417–430. doi: 10.1097/01.MIB.0000440815.76627.64

Martin, R., Miquel, S., Ulmer, J., Kechaou, N., Langella, P., and Bermudez-Humaran, L. G. (2013). Role of commensal and probiotic bacteria in human health: a focus on inflammatory bowel disease. *Microb. Cell Fact.* 12:71. doi: 10.1186/1475-2859-12-71

Martin, R., Sanchez, B., Urdaci, M. C., Langella, P., Suarez, J. E., and Bermudez-Humaran, L. G. (2015). Effect of iron on the probiotic properties of the vaginal isolate *Lactobacillus jensenii* CECT 4306. *Microbiology* 161(Pt 4), 708–718. doi: 10.1099/mic.0.000044

Mayeur, C., Gratadoux, J. J., Bridonneau, C., Chegdani, F., Larroque, B., Kapel, N., et al. (2013). Faecal D/L lactate ratio is a metabolic signature of microbiota imbalance in patients with short bowel syndrome. *PLoS ONE* 8:e54335. doi: 10.1371/journal.pone.0054335

Miquel, S. R., Martìn, Bridonneau, C., Robert, V., Sokol, H., Bermúdez-Humarán, L., Thomas, M., et al. (2014). Ecology and metabolism of the beneficial intestinal commensal bacterium *Faecalibacterium prausnitzii*. *Gut Microbes* 5, 146–151. doi: 10.4161/gmic.27651

Miquel, S., Beaumont, M., Martin, R., Langella, P., Braesco, V., and Thomas, M. (2015a). A proposed framework for an appropriate evaluation scheme for microorganisms as novel foods with a health claim in Europe. *Microb. Cell Fact.* 14, 48. doi: 10.1186/s12934-015-0229-1

Miquel, S., Leclerc, M., Martin, R., Chain, F., Lenoir, M., Raguideau, S., et al. (2015b). Identification of metabolic signatures linked to anti-inflammatory effects of *Faecalibacterium prausnitzii*. *Mbio* 6:e00300-15. doi: 10.1128/mBio.00300-15

Miquel, S., Martin, R., Rossi, O., Bermudez-Humaran, L., Chatel, J., Sokol, H., et al. (2013). *Faecalibacterium prausnitzii* and human intestinal health. *Curr. Opin. Microbiol.* 16, 255–261. doi: 10.1016/j.mib.2013.06.003

Neish, A. S. (2009). Microbes in gastrointestinal health and disease. *Gastroenterology* 136, 65–80. doi: 10.1053/j.gastro.2008.10.080

Pineiro, M., and Stanton, C. (2007). Probiotic bacteria: legislative framework–requirements to evidence basis. *J. Nutr.* 137(3 Suppl. 2), 850S–853S.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464, 59–65. doi: 10.1038/nature08821

Quevrain, E., Maubert, M. A., Michon, C., Chain, F., Marquant, R., Tailhades, J., et al. (2016). Identification of an anti-inflammatory protein from *Faecalibacterium prausnitzii*, a commensal bacterium deficient in Crohn's disease. *Gut* 65, 415–425. doi: 10.1136/gutjnl-2014-307649

Radziwill-Bienkowska, J. M., Le, D. T., Szczesny, P., Duviau, M. P., Aleksandrzak-Piekarczyk, T., Loubiere, P., et al. (2016). Adhesion of the genome-sequenced *Lactococcus lactis* subsp. *cremoris* IBB477 strain is mediated by specific molecular determinants. *Appl. Microbiol. Biotechnol.* 100, 9605–9617. doi: 10.1007/s00253-016-7813-0

Radziwill-Bienkowska, J. M., Zochowska, D., Bardowski, J., Mercier-Bonin, M., and Kowalczyk, M. (2014). *Lactococcus lactis* IBB477 presenting adhesive and muco-adhesive properties as a candidate carrier strain for oral vaccination against influenza virus. *Acta Biochim. Pol.* 61, 603–607.

Rajilic-Stojanovic, M., Biagi, E., Heilig, H. G., Kajander, K., Kekkonen, R. A., Tims, S., and de Vos, W. M. (2011). Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* 141, 1792–1801. doi: 10.1053/j.gastro.2011.07.043

Ramirez-Farias, C., Slezak, K., Fuller, Z., Duncan, A., Holtrop, G., and Louis, P. (2009). Effect of inulin on the human gut microbiota: stimulation of Bifidobacterium adolescentis and *Faecalibacterium prausnitzii*. *Br. J. Nutr.* 101, 541–550. doi: 10.1017/S0007114508019880

Rossi, O., Khan, M. T., Schwarzer, M., Hudcovic, T., Srutkova, D., Duncan, S. H., et al. (2015). *Faecalibacterium prausnitzii* strain HTF-F and its extracellular polymeric matrix attenuate clinical parameters in DSS-induced colitis. *PLoS ONE* 10:e0123013. doi: 10.1371/journal.pone.0123013

Sanders, M. E., Akkermans, L. M., Haller, D., Hammerman, C., Heimbach, J., Hormannsperger, G., et al. (2010). Safety assessment of probiotics for human use. *Gut Microbes* 1, 164–185. doi: 10.4161/gmic.1.3.12127

Sarrabayrouse, G., Bossard, C., Chauvin, J. M., Jarry, A., Meurette, G., Quevrain, E., et al. (2014). CD4CD8αα lymphocytes, a novel human regulatory T cell subset induced by colonic bacteria and deficient in patients with inflammatory bowel disease. *PLoS Biol.* 12:e1001833. doi: 10.1371/journal.pbio.1001833

Schmidt, C. (2013). The startup bugs. *Nat. Biotechnol.* 31, 279–281. doi: 10.1038/nbt.2544

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J. J., et al. (2008). *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16731–16736. doi: 10.1073/pnas.0804812105

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

van Tongeren, S. P., Slaets, J. P., Harmsen, H. J., and Welling, G. W. (2005). Fecal microbiota composition and frailty. *Appl. Environ. Microbiol.* 71, 6438–6442. doi: 10.1128/AEM.71.10.6438-6442.2005

Wiseman, L. R., Lamb, H., and Cefpirome, M. (1997). A review of its antibacterial activity, pharmacokinetic properties and clinical efficacy in the treatment of severe nosocomial infections and febrile neutropenia. *Drugs* 54, 117–140. doi: 10.2165/00003495-199754010-00013

Wrzosek, L., Miquel, S., Noordine, M. L., Bouet, S., Chevalier-Curt, M. J., Robert, V., et al. (2013). Bacteroides thetaiotaomicron and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC Biol.* 11:61. doi: 10.1186/1741-7007-11-61

Conflict of Interest Statement: PL and HS are co-founders of the start-up NextBiotiX aiming to use next-generation probiotics to fight and to prevent IBD.

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

Cornuault et al. Microbiome
https://doi.org/10.1186/s40168-018-0452-1

Microbiome

RESEARCH Open Access

Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes

Jeffrey K. Cornuault[1], Marie-Agnès Petit[1], Mahendra Mariadassou[2], Leandro Benevides[1,3], Elisabeth Moncaut[1], Philippe Langella[1], Harry Sokol[1,4,5] and Marianne De Paepe[1*]

Abstract

Background: Viral metagenomic studies have suggested a role for bacteriophages in intestinal dysbiosis associated with several human diseases. However, interpretation of viral metagenomic studies is limited by the lack of knowledge of phages infecting major human gut commensal bacteria, such as *Faecalibacterium prausnitzii*, a bacterial symbiont repeatedly found depleted in inflammatory bowel disease (IBD) patients. In particular, no complete genomes of phages infecting *F. prausnitzii* are present in viral databases.

Methods: We identified 18 prophages in 15 genomes of *F. prausnitzii*, used comparative genomics to define eight phage clades, and annotated the genome of the type phage of each clade. For two of the phages, we studied prophage induction in vitro and in vivo in mice. Finally, we aligned reads from already published viral metagenomic data onto the newly identified phages.

Results: We show that each phage clade represents a novel viral genus and that a surprisingly large fraction of them (10 of the 18 phages) codes for a diversity-generating retroelement, which could contribute to their adaptation to the digestive tract environment. We obtained either experimental or in silico evidence of activity for at least one member of each genus. In addition, four of these phages are either significantly more prevalent or more abundant in stools of IBD patients than in those of healthy controls.

Conclusion: Since IBD patients generally have less *F. prausnitzii* in their microbiota than healthy controls, the higher prevalence or abundance of some of its phages may indicate that they are activated during disease. This in turn suggests that phages could trigger or aggravate *F. prausnitzii* depletion in patients. Our results show that prophage detection in sequenced strains of the microbiota can usefully complement viral metagenomic studies.

Keywords: Bacteriophages, *Faecalibacterium prausnitzii*, Inflammatory bowel disease, Comparative genomics, Prophages

Background

Faecalibacterium prausnitzii is one of the most abundant bacterial species of the human gastrointestinal tract (GIT) microbiota. This species is an important producer of butyrate, a short-chain fatty acid with a broad range of beneficial properties on human health [1, 2]. *F. prausnitzii* also exerts anti-inflammatory activity through the production of the MAM protein [3] and provides protection against inflammation in different animal models [3–6]. In humans, a low abundance of this species has been regularly reported in the gut microbiota of inflammatory bowel disease (IBD) patients [6, 7]. Depletion of *F. prausnitzii* in patients has been associated with a higher risk of recurrence of flares [4, 8], possibly in relation with the capacity of this species to modulate T cell responses [6]. Bacteriophages, or phages, the viruses infecting bacteria, might play a role in this depletion [9]. But, although viral metagenomics studies have provided indications that *F. prausnitzii* phages are present in human fecal samples [10], no complete *F.*

* Correspondence: marianne.depaepe@inra.fr
[1]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
Full list of author information is available at the end of the article

prausnitzii phage has been characterized to date. Phages can be either virulent or temperate: virulent phages perform only lytic cycles, during which new phage particles are produced and the infected bacteria are killed. Temperate phages can similarly replicate through lytic cycles, but they are also able to enter a latent lysogenic state. During lysogeny, the repressed phage, known as a prophage, is generally integrated into the bacterial chromosome, and the expression of most of its genes is repressed. Nevertheless, a few genes continue to be expressed, including the repressor of the lytic cycle and a class of genes known as morons whose expression often modifies the physiology of the host bacterium. In this way, prophages can modify numerous phenotypes of their bacterial host, such as resistance against phage attack, stress resistance, pathogenicity, or ability to form biofilms [11, 12]. In response to a specific signal, active prophages can switch back to lytic multiplication, a process known as induction. Because of this, active prophages can be highly detrimental to their host in some inducing conditions [13].

Therefore, not only virulent phages but also temperate ones have been suspected to play a role in dysbiosis of intestinal microbiota [14, 15]. To examine this possibility, several viral metagenomic studies of the intestinal microbiota have been performed in recent years [10, 16–18]. However, a major limitation of these studies is the difficulty of viral sequence analysis. Indeed, annotation of viral sequences, and in particular the prediction of the bacterial host, is mainly based on alignment to reference viral sequences in databases [19], and limitations in the viral databases render annotation of most viral sequences difficult if not impossible. Limitations are particularly flagrant concerning phages infecting important groups such as *Ruminococcaceae*, previously known as *Clostridium* cluster IV. There is, to our knowledge, no phage infecting members of this family whose complete genome has been deposited in viral databases. Isolating and describing phages infecting *Clostridiaceae* species such as *F. prausnitzii* is thus a prerequisite for investigation of their potential role in dysbiosis. In this study, we characterize for the first time complete *F. prausnitzii* temperate phages and show that they belong to eight new proposed viral genera. In addition, we show that some of them are either more prevalent or more abundant in the fecal samples of IBD patients compared to healthy controls, suggesting these phages might play a role in the disease.

Results

Identification of 23 prophages corresponding to novel proposed genera

Fifteen *F. prausnitzii* genomes, among which 10 were recently published [20], were analyzed with the PHASTER interface (http://phaster.ca), dedicated to prophage detection [21, 22]. Expert inspection retained 23 complete prophages out of the 50 predicted regions, whose borders were manually refined, notably by detecting integrases (Additional file 1). Most integrases belonged to one of the two classical families of phage integrase: tyrosine integrase and large serine recombinase (LSR) (Table 1). The integrase of prophage Lagaffe, however, is atypical since it belongs to the IS30 transposase family, found almost exclusively in transposons. Enzymes of this family were recently described in small temperate single-stranded DNA filamentous phages, and mediate site-specific recombination in a chimeric prophage [23], but had not yet been described in large prophages.

The 23 retained prophages were compared pairwise at the nucleotide level, allowing the definition of six phylogenetic clades (Fig. 1), within which gene synteny was well conserved, and two singletons. The clusters/singletons are well separated from each other, mostly with no gene shared between clusters. None of the predicted prophages had significant nucleotide similarity with a viral genome of the NCBI RefSeq genome database on more than 5% of its genome, indicating they all correspond to new viruses. According to current taxonomic metrics whereby phages within a genus share at least 40% of their proteins, and phages of the same species have more than 95% identity on the whole genome [24, 25], the 23 prophages correspond to 18 species and 8 genera (Additional file 2), for which we propose names (Table 1 and Fig. 1). Of note, Taranis and Toutatis belong to related genera since they share 26% of their proteome, with an overall 39% amino acid identity.

Given the ubiquity of *F. prausnitzii* in human gut microbiota, we investigated whether similar phage sequences had been found in other studies. To this end, we compared the prophage genomes against the NCBI nr database. Prophages Lagaffe, Mushu, and ToutatisSL3/3 have homology to four large contigs assembled from reads obtained by sequencing DNA from virus-like particles (VLPs) isolated from human fecal samples [26] (Fig. 1). In addition, ToutatisL2/6 is homologous to a 9-kb contig assembled from reads of the VLPs of a human fecal sample. This contig had been predicted to correspond to a *F. prausnitzii* prophage [10]. This provides further support for the proposition that these prophage regions are of viral nature. We then looked for CRISPR spacers matching the prophages in the 15 *F. prausnitzii* genomes, and found spacers for eight prophages (Table 1), which is indicative of recent infection of these strains by the respective phages [27]. Overall, indications of activity were obtained for at least one prophage of each proposed genus, as summarized in Table 1.

For six of the eight prophage groups, the host range seemed to be restricted to *F. prausnitzii*, since no similar

Table 1 Main characteristics of the eight proposed *F. prausnitzii* phage genera

Proposed genus	Classification (Virfam)	Bacterial hosts	Prophage name	Strains with complete prophage[a]	Common features				Evidence of activity	
					Genome size, kb	RT	Integrase type	Recombinase	CRISPR spacers[a]	Viral contig
"FPlagaffevirus"	Myoviridae type 1, cluster 6	*F. prausnitzii* *B. hansenii* CM62/1	Lagaffe	A2-165, *B. hansenii* 20583, CM62/1	48.3	Yes	IS30 TPase	ERF		Yes
"FPmushuvirus"	Myoviridae type1, cluster 8 "Saltoviridae"	*F. prausnitzii*	Mushu	A2-165, C_4543, C_4574	36.4	Yes	Mu TPase			Yes
		Ruminococcaceae bacterium D16	MushuRBD16	R. bacterium D16	36.2	Yes	Mu TPase			
"FPlughvirus"	Siphoviridae type 1, cluster 1	*F. prausnitzii*	Lugh	C_4574, C_4543	34.1		Tyr Int	Sak4	C_4544	
			Lugh4544	C_4544	35.9		Tyr Int	ERF		
			Lugh4542	C_4542	36.4		Tyr Int	ERF		
			LughKLE1255	KLE1255	36.6		Tyr Int	ERF		
			Lugh4540	C_4540	31.1		Tyr Int	Sak4		
			LughL2-6	L2-6	31.2		Tyr Int			
"FPtaranisvirus"	Myoviridae type 1, cluster 6	*F. prausnitzii*	Taranis	C_4574, C_4543	56.0	Yes	LSR	Sak4	C_4541(2), L2/6	
			TaranisL2-6	L2-6	41.9		LSR	Sak4	C_4541	
"FPeponavirus"	Myoviridae type 1, cluster 6	*F. prausnitzii*	Epona	C_4573, C_4575	49.5	Yes	LSR		C_4541, C_4575	
			EponaM21-2	M21/2	48.5	Yes	LSR			
"FPtoutatisvirus"	Myoviridae type 1, cluster 6	*F. prausnitzii*	Toutatis	C_4542	54.7	Yes	LSR		C_4541	
			ToutatisL2-6	L2-6	53.6	Yes	LSR		C_4541	
			ToutatisKLE1255	KLE1255	54.5	Yes	LSR		C_4541, L2/6	Yes
			ToutatisSL3/3	SL3/3	50.1		LSR			
"FPbrigitvirus"	Myoviridae type 1, cluster 7	*F. prausnitzii*	Brigit	C_4644	61.6	Yes	LSR	RecT	C_4541	
"FPoengusvirus"	Siphoviridae type 1, cluster 2	*F. prausnitzii*	Oengus	C_4542	58.4		Tyr Int			Yes

RT retron-type reverse transcriptase, *Tpase* transposase, *Tyr int* tyrosine recombinase superfamily, *LSR* large serine recombinase, *CM62/1* unclassified *clostridium* sp. CM62/1
[a]C_ stands for CNCM

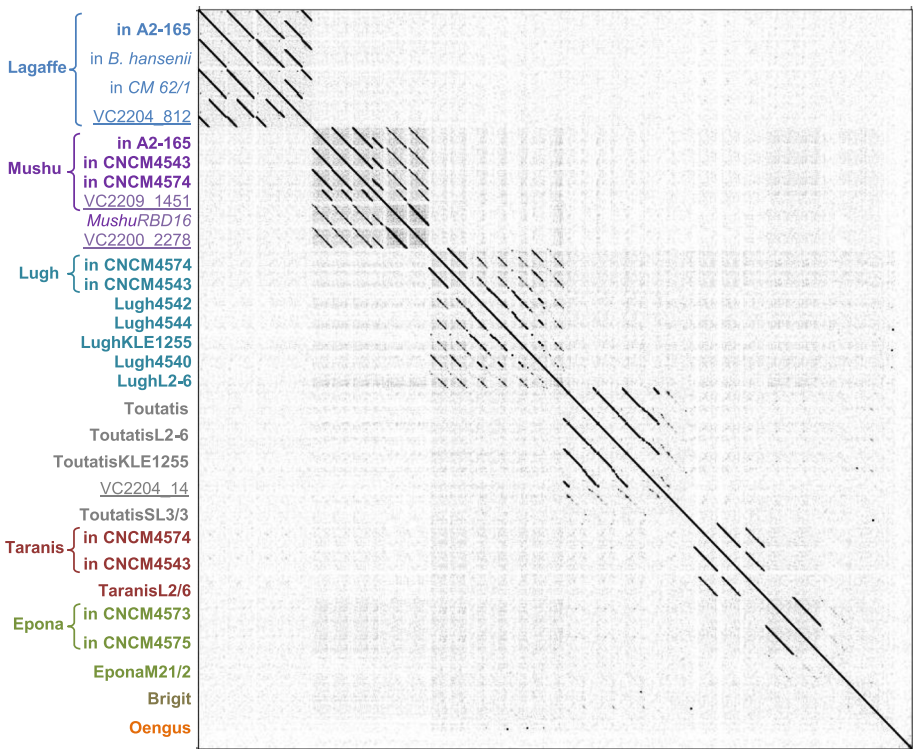

Fig. 1 Whole genome dot plot of the 23 *F. prausnitzii* prophages and 7 homologous sequences retrieved from the nr/nt database define six clades and two singletons (Brigit and Oengus, last lanes). Braces group similar prophages, which correspond to the same phage species found in different bacterial genomes. Names beginning "VC" (underlined) correspond to metagenomic viral contigs, and names in italic correspond to prophages in non-*F. prausnitzii* bacterial species

prophage could be detected by BLASTn in other genomes of the NCBI RefSeq database. On the other hand, prophages with 65% overall sequence identity to Mushu prophage were found in the unclassified *Ruminococcaceae* bacterium D16 (Fig. 1 and Additional file 3) and in the unclassified bacterium SS3/4 (Genbank sequence FP929062.1). Even more striking was the detection of a prophage almost identical to Lagaffe (98% nucleotide identity) in the genome of *Blautia hansenii* DSM 20583 (Fig. 1 and Additional file 3), a bacterium only related to *F. prausnitzii* A2-165 at the order level (they share only 29% of their proteins).

To classify these proposed new phage genera, since they greatly differ from known viruses, we used the Virfam webserver (biodev.cea.fr/virfam). Virfam uses virion protein remote homology and gene synteny to classify phages [28]. All the phages were classified among the Neck type 1, the largest clade of the Caudoviridae order (Additional file 4). Four genera grouped within the cluster 6 of this neck type, which branches deeply in the overall tree and comprises phages that could not be attributed to better-resolved clusters. Phage Mushu belongs to cluster 8, which corresponds to the "Saltoviridae" proposed family, composed of Mu-like transposable phages [29]. Accordingly, Mushu and its relative from the *Ruminococcaceae* bacterium D16 genome have been assigned to this proposed family [29, 30]. Phages from the proposed "FPlughvirus" genus and phage Oengus group in clusters 1 and 2 respectively, with other unclassified Siphoviridae phages infecting Firmicutes. Surprisingly, Brigit phage groups in cluster 7, which typically comprises phages infecting *Enterobacteriaceae*.

Functional annotation of prophage genes reveals a high prevalence of diversity-generating retroelements

Within each proposed genus, we chose one representative prophage, which we subsequently refer to as the "type phage" for the group. Each type phage was annotated by using the Phagonaute webserver (genome.jouy.inra.fr/phagonaute), which uses remote homology searches and gene synteny information [28, 31]. We improved the functional predictions as compared to the previous automatic annotation using RAST server by 2.5- to 5-fold (Fig. 2). All phage functional modules were detected in each prophage, providing further indication that these prophages correspond to complete phage genomes.

A remarkable feature of these genomes is the high prevalence of diversity-generating retroelements (DGRs). DGRs generate variability in target genes through a reverse transcriptase (RT)-mediated mechanism that

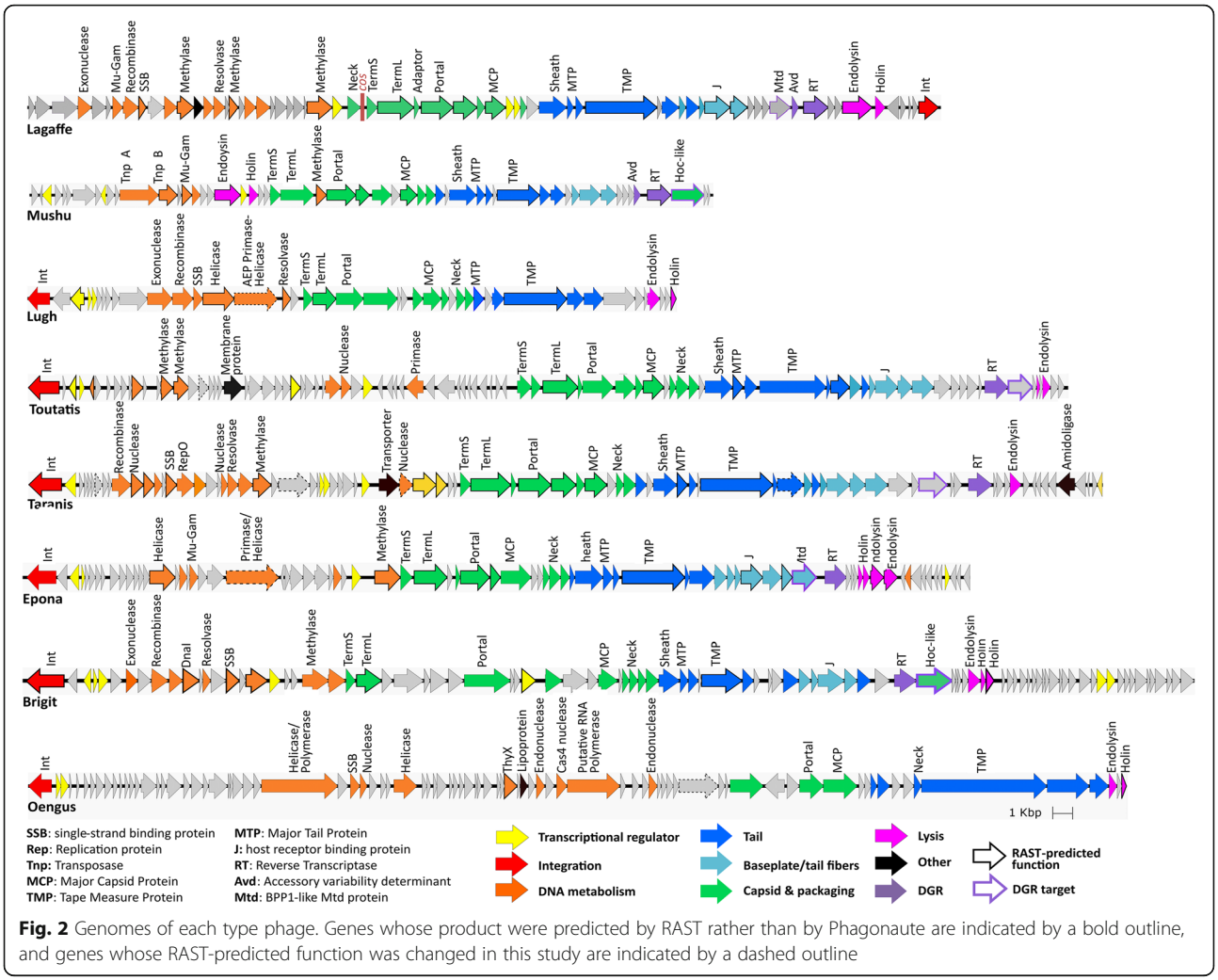

Fig. 2 Genomes of each type phage. Genes whose product were predicted by RAST rather than by Phagonaute are indicated by a bold outline, and genes whose RAST-predicted function was changed in this study are indicated by a dashed outline

introduces nucleotide substitutions at defined locations in specific genes (reviewed in [32]). In this process, adenine residues are converted to random nucleotides in a unidirectional process from a donor template repeat (TR) to a recipient variable repeat (VR) present in the target gene. TR and VR sequences were identified in each prophage type coding for an RT (Additional file 5). As with other DGRs, TR/VR sequences were between 100 and 135 bp long, they differed almost exclusively at sites corresponding to TR adenines, and the VRs were located at the 3′ end of protein-coding genes. We then searched for analogues of Avd, a low-molecular-weight protein required for mutagenic homing in the prototype DGR of the *Bordetella pertussis* phage BPP1 [33]. Homologs of Avd were detected in Mushu and Lagaffe genomes only (Fig. 2). In other phages, DGRs might function with another small protein, as has previously been suggested for other DGRs [34]. In Lagaffe and Epona, the putative target gene encodes a protein belonging to the formylglycine-generating enzyme (FGE) superfamily, which includes the BPP1 tail protein MTD, implicated in bacterial recognition [35]. Mushu and Brigit DGR target genes code for a Hoc-like capsid decoration protein, which in T4 improves the binding to intestinal mucus [36]. In Toutatis and Taranis, no putative function could be attributed to the DGR target.

Lagaffe and Mushu produce virions

We next evaluated the activity of Lagaffe and Mushu in laboratory conditions by isolating virions from the supernatant of a culture of strain A2-165 and sequencing the encapsidated DNA. We obtained 19,702 and 30,274 reads mapping to the Lagaffe and Mushu prophages, which correspond respectively to an 8- and 16-fold higher coverage than the rest of A2-165 genome (Additional file 6). This demonstrated that the two prophages produce virions and that there is no other active prophage in the A2-165 strain. Analysis of Mushu reads confirmed that it replicates by transposition: at the extremity of the Mushu genome, reads were flanked by variable host DNA sequences on both sides (Fig. 3a).

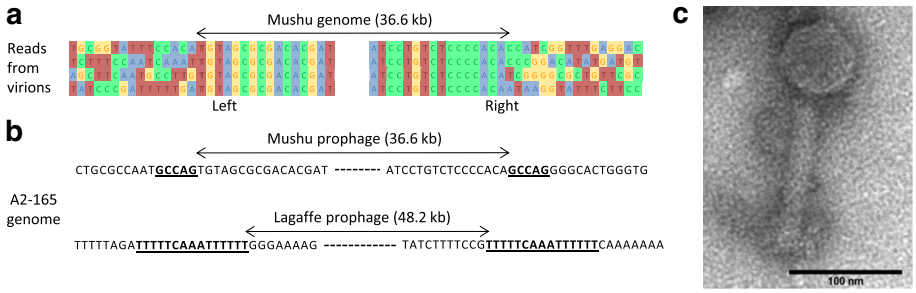

Fig. 3 Mushu and Lagaffe are active phages. **a** Fragments of eight arbitrarily chosen reads of the two extremities of Mushu encapsidated DNA, showing the variable bacterial sequences on each side, typical of transpositional replication. **b** Direct repeats (underlined) on both sides of the prophages are present in the A2-165 genome. **c** Photograph of Mushu virion obtained by transmission electronic microcopy

This particular configuration is similar to that of *E. coli* Mu phage encapsidated DNA, and results from packaging by a full head mechanism of integrated viral copies spread randomly through the host genome [37]. Analysis of reads also revealed the Mushu and Lagaffe precise integration sites in the A2-165 chromosome (Fig. 3b). Mushu integration site presents the typical 5 bp direct repeat of some Mu-like phages [37]. More surprisingly, Lagaffe is also flanked by direct repeats, whereas a sequence composed of inverted repeats is expected to form the integration site of IS30 transposases [23]. Analysis of Lagaffe reads with the PhageTerm tool [38], dedicated to the detection of the strategy used by phages to package their genome into capsids, identified the *cos* region of Lagaffe, upstream of the terminase gene *termS*.

For both phages, reads spanning the VR region of the DGR are more variable than the rest of the genome, an evidence of DGR activity (Additional file 5C). The level of variability observed in Mushu is however remarkable, since the phage genomes sequenced result in theory from a single cycle of multiplication (they are produced following induction in lysogenic cells), and suggests a very high efficacy of TR-to-VR transfer in this DGR.

Lagaffe virions could also be detected in *Blautia hansenii* DSM 20583 cultures by quantitative PCR (Table 2), which demonstrates the remarkable polyvalent nature of this phage. Indeed, usually, phages are restricted to strains of the same species or sometimes the same genus [39]. Only few cases have been reported of a phage infecting bacteria from different families, similar to Lagaffe [40, 41]. Finally, observation by electron microscopy of A2-165 culture supernatant revealed Mushu virions, with a head diameter of 61 ± 1 nm and a tail of length of 134 ± 2 nm and width of 22 ± 1 nm (Fig. 3c). Due to its very low concentration in A2-165 culture (3×10^5 ml^{-1}, Table 2), intact Lagaffe virions could not be observed.

Lagaffe and Mushu prophages are not inducible by DNA damage

To increase virion titers and search for inducing signals, A2-165 *F. prausnitzii* cultures were treated with different concentrations of three potent inducers of phage lytic cycle: mitomycin C, hydrogen peroxide, and bile salts (cholate/deoxycholate [1:1]) [13, 42]. Mitomycin C in particular causes DNA damage, which is a general signal for prophage derepression. However, none of the treatments significantly increased virion concentrations (Additional file 7). For each compound tested, the highest dose tested almost abolished bacterial growth, indicating that they are toxic for *F. prausnitzii* and that the absence of prophage induction is not due to too low concentrations of potent inducers. Lagaffe and Mushu are thus probably not inducible by DNA damage, similar to most Mu-like phages.

Mushu virions are produced in the mouse gastro-intestinal tract but are not increased by colitis

We next investigated the production of A2-165 prophages in the mouse gastro-intestinal tract (GIT). Indeed, recent studies have shown increased prophage induction in the murine GIT, as compared to in vitro conditions [43, 44]. In order to obtain a colonization level of *F. prausnitzii* sufficient to detect virions in feces, we used dixenic mice colonized with *F. prausnitzii* A2-165 and *Escherichia coli*. In

Table 2 Lagaffe and Mushu virion concentration in culture supernatants and in mouse feces

	Virions/ml or virions/g with *F. prausnitzii* A2165			Virions/ml in *B. hansenii* cultures
	Cultures	Healthy mouse gut	Inflamed mouse gut	
Lagaffe	$3 \pm 0.3 \times 10^5$	$< 1 \times 10^6$	$< 1 \times 10^6$	$6 \pm 3 \times 10^5$
Mushu	$8 \pm 2 \times 10^6$	$9 \pm 2 \times 10^6$	$7 \pm 2 \times 10^6$	

Virions were quantified by qPCR on encapsidated genomes. Intestinal inflammation in mouse was induced by adding DSS to drinking water. Cultures correspond to saturated 24 h cultures grown in LY-BHI

these mouse feces, the Lagaffe virion concentration was too low to be detected, but Mushu virions were present in similar concentrations to those observed in vitro (Table 2 and Additional file 8). Of note, however, the potential binding of virions to bacterial cells and intestinal mucus prevents the direct deduction of prophage induction rates from the concentrations of free virions. We then monitored the effect of moderate inflammation on the prophage induction rate. To mimic moderate inflammatory conditions, dextran sodium sulfate (DSS) was added at 2% to drinking water for 7 days. This treatment resulted in mild inflammation but did not increase virion concentrations in mouse feces (Additional file 8).

Some *F. prausnitzii* phages are more abundant in the gut viromes of IBD subjects

To evaluate the activity of the identified prophages in the human GIT, we investigated the sequences obtained from the viral fraction of 171 fecal samples from a large cohort of 52 IBD patients and 21 healthy controls [18]. Alignment by BLASTn of reads from this dataset to 17 of the identified prophages retrieved a large number of hits (ToutatisL2/6 being very close to Toutatis, 96% identical over 80% of its length; it was not included in this analysis). Altogether, virome reads mapping to at least one of the eight type phages represented 0.02% of all reads in healthy samples and 0.30% in IBD samples (mean of the proportion of reads matching on *F. prausnitzii* type phages over total number of reads). Remarkably, reads mapping to phage Oengus represented 22% of reads in one sample (Fig. 4, sample ERR843984). In that sample, we could assemble the matching reads into a contig covering 93% of Oengus genome with 99% identity. Yet matching read coverage was not always even on the prophages (Fig. 4), suggesting the presence of phages only sharing genome fragments with the phages of this study. Concerning phages of the "FPlughvirus" proposed genus, this may be due to the high level of genomic mosaicism in this genus, i.e., the alternation of regions of high identity with regions of low identity between phage genomes (Additional file 9). Of note, various genes coding for homologous recombination proteins, involved in phage mosaicism, were found in this genus (Table 1) [45, 46].

We then wished to evaluate whether *F. prausnitzii* phages are significantly more prevalent and/or abundant in fecal samples of healthy individuals as compared to those of IBD patients. To this end, for each phage, we first determined phage-positive viral samples, i.e., the samples that unambiguously contain that phage or related ones. A flowchart of the procedure is represented in Fig. 5a. Since Lagaffe and Mushu relatives infect species other than *F. prausnitzii*, whose link with IBD is not known, they were removed from the analysis. Eleven of the 15 investigated phages were present in more than 3% of samples, and one of them, Oengus, was found in up to 40% of samples (Fig. 5b). Two "FPlughvirus" phages were statistically more prevalent in IBD samples than in healthy controls (Fig. 5, two tailed two-proportion z test, $p < 0.05$). We then compared the proportion of reads homologous to each phage in phage-positive samples by fitting a logistic regression with status (healthy/IBD) as fixed effect and patient as mixed effect. Overall, all phages except those from the "FPtaranisvirus" proposed genus are more abundant in the IBD samples than in the healthy ones, although the difference reaches significance only for LughL2/6 and Epona (Fig. 5c, p value < 0.05). Since *F. prausnitzii* populations are generally less abundant in IBD patients than in healthy controls [7], the higher occurrence and proportion of some *F. prausnitzii* temperate phages suggest that they are more active in the microbiota of IBD patients and that this activity is related to the depletion of *F. prausnitzii*.

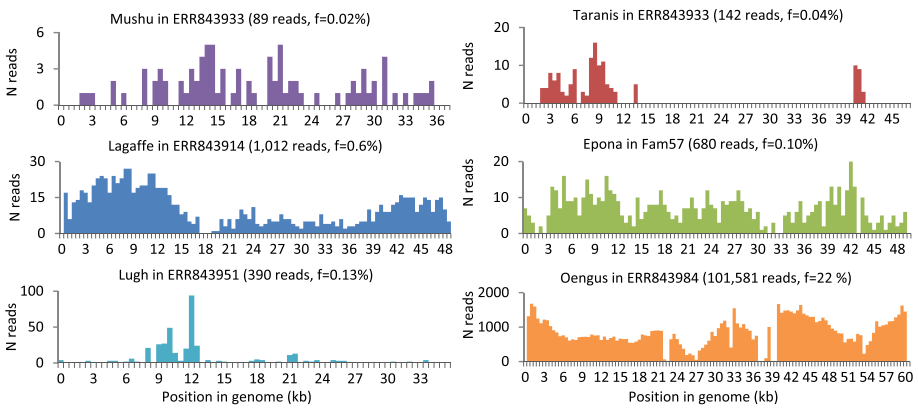

Fig. 4 *F. prausnitzii* phages are present in intestinal viromes. Read coverage for six type phages in samples with the best abundance and coverage from the datasets of Norman et al. [18] and Reyes et al. [52]. Number of aligned reads (ordinate) is represented as a function of position in the phage genome (abscissa). Phage name, virome name, total number of aligned reads, and ratio of aligned to total reads are shown for each panel

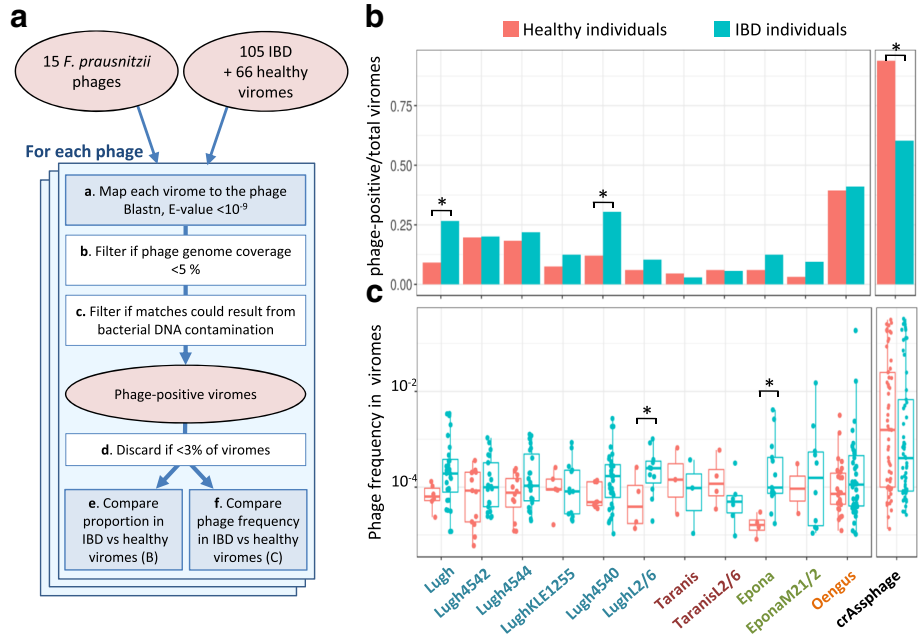

Fig. 5 *F. prausnitzii* phages in healthy versus IBD-diseased individuals from the dataset of Norman et al. [18]. **a** Workflow of the procedure. a To infer the presence of a phage in a given sample, virome reads were aligned to the phage genome, and hits showing greater than 75% identity were retained. b If hits of a given sample were concentrated on less than 5% of the phage genome, the sample was not considered further for this phage. c The number of hits was compared to the maximal theoretical number of hits that could result from the measured bacterial DNA contamination. If the number of hits was significantly higher, the sample was classified as phage-positive. d If phage-positive samples represent less than 3% of total samples, this phage was not further studied. e The proportion of phage-positive viromes is compared to IBD versus healthy viromes. f In phage-positive viromes, phage frequency is compared to IBD versus healthy samples. **b** Fraction of phage-positive samples according to the health status of individuals. Three phages are significantly more prevalent in IBD-diseased than in healthy individuals. The highly prevalent crAssphage was included as a positive control in the analysis. **c** Phage abundance (read frequency) in positive samples according to the health status of individuals. Two phages, Epona and LughL2/6, are significantly more abundant in IBD patients than in healthy individuals

Discussion

F. prausnitzii genomes are particularly rich in prophages: in this study, we found that 85% of them harbor at least one complete prophage, as opposed to only 46% of all completely sequenced bacterial genomes [47]. The 23 identified complete prophages in this study correspond to 18 different phages, which can be grouped into eight new viral genera. Thanks to experimental and in silico investigations, we found evidence of activity for at least one member of each proposed genus. Their novelty and diversity are striking and greatly enrich the viral databases, which are at present highly skewed towards phages infecting *proteobacteria* or other pathogens, which together account for more than 75% of phages in the NCBI viral genome database (as of December 2017). The novelty of the viral proteins coded by *F. prausnitzii* phages should not only improve the analysis of viral metagenomic sequences but might also lead to biotechnology applications, such as genetic engineering of *F. prausnitzii*.

A very high proportion *F. prausnitzii* phages possesses DGRs: these diversification systems are present in 10 out of the 18 phages. By comparison, only 0.05% of the 876 phages present in the Phagonaute database (comprising all fully sequenced phage genomes available at the NCBI as of May 2014) possess DGRs. This high prevalence in *F. prausnitzii* phages might explain the previously observed high frequency of DGRs in intestinal viromes [26]. By generating variants of target genes, DGRs presumably confer selective advantages to the phages, such as the ability to infect bacteria with different variants of the surface component used as receptor by the phage, as has been shown in the case of *B. pertussis* phage BPP1. In this phage, the DGR-mediated diversification of the MTD protein generates at low-frequency phage variants able to infect bacteria with receptors different from those of the original host [33]. In phages Lagaffe and Epona, the proteins diversified by the DGR are homologous to BPP1 MTD, so their diversification probably also facilitates phage adaptation to different bacterial receptors. This property would provide a selective advantage to the phage either in the case of a preexisting high bacterial surface receptor diversity or in conditions of continually evolving bacteria in the context of an ongoing arms race between phage and their hosts. In phages Brigit and Mushu, by targeting Hoc-like capsid

decoration proteins, the DGR system could improve the binding of virions to mucus, a strategy that has been suggested to increase the probability for phages to encounter their bacterial host [36]. Whatever the role of these DGRs, they contribute to the generation of microdiversity regions in the genomes of intestinal phages, which notoriously complicates the assembly of reads, and hence the analysis of GIT viral metagenomes [48].

We found that the *F. prausnitzii* temperate phages described here are present in 5 to 40% of the human fecal samples from the study of Norman and colleagues [18]. Despite their relatively low abundance in most samples of this dataset, they can represent up to 22% of total viral reads (Additional file 10). Most importantly, some of them are either more prevalent or more abundant in samples from IBD patients compared to that of healthy controls. This result is striking since IBD patients have generally a lower abundance of *F. prausnitzii* in their microbiota [6, 7]. This suggests that this higher prevalence or abundance is not just the consequence of a higher abundance of their bacterial host but rather that these *F. prausnitzii* prophages are more active in the microbiota of IBD patients as compared to healthy controls. This phage activity could aggravate or even trigger *F. prausnitzii* depletion. Since this depletion is highly suspected to participate in intestinal inflammation, our results suggest for the first time a role of *F. prausnitzii* phages in IBD.

Conclusion

Here, we describe for the first time complete *F. prausnitzii* temperate phage genomes. These phages belong to the new viral genera, and most of their proteins have only remote homology with previously described viral proteins. The novelty of *F. prausnitzii* phages warrants great enrichment of databases, which should in turn significantly improve annotation of viral gut metagenomic sequences. In addition, we show that some of these *F. prausnitzii* phages are either more prevalent or more abundant in the fecal samples of IBD patients compared to that of healthy controls, suggesting that these phages might play a role in the disease. Future mechanistic exploration of the nature of phage-bacteria interactions is now necessary to elucidate the role of temperate phages in dysbiosis.

Methods

Prophage detection and comparison

F. prausnitzii genomes listed in Additional file 1 were analyzed with PHASTER (http://phaster.ca). "Intact" and "questionable" regions were then validated by manual inspection. We retained regions with genes in each of the following modules: replication, capsid, lysis, and lysogeny, and no genes specific of integrative plasmids or insertion sequences. Of note, several prophage regions were on borders of contigs, suggesting that some complete prophages could have been missed because of splitting between two contigs and that some of the retained prophages might lack some genes. Prophage borders were refined by searching for phage integrase and phage-specific genes by BLASTp against the nr NCBI database. In several cases, the ORF corresponding to the integrase had not been detected automatically and was added manually. For type prophages, few other ORFs were also added. In addition, repeats on both sides (attL and attR sites) were searched for. They could be identified only in prophages encoding a tyrosine integrase, which is not surprising since tyrosine integrases generally use longer identical repeat than serine integrases. A multi-fasta file containing all prophage nucleotide sequences is available in the Additional file 11. The type phage genomes were deposited in the NCBI viral database (accession numbers in Additional file 1) and are consultable on the Phagonaute webserver (genome.jouy.inra.fr/phagonaute).

Genomic regions homologous to the 18 prophages were searched for in the nr/nt nucleotide collection of the NCBI by BLASTn, with the discontinuous megaBLAST default parameters. Sequences producing alignments covering more than 5% of the query were retained. Whole genome dotplots were realized with Gepard [49], using a length of 11 nt for minimal exact matches. CRISPR spacers were searched for with CRISPRDetect [50], using default parameters.

Shared proteins between prophage pairs

For each prophage pair to be compared, all proteins of the first phage were compared to the proteins of the second one (BLASTp, E value $< 10^{-3}$), and best hits were considered shared proteins. The proportion of shared proteins was then calculated by dividing the number of shared proteins by the total protein number (minimal value of the pair), and the average identity percent of the shared proteins was computed. To select an outgroup phage for each proposed genus, all proteins of the type phage were compared to the viral protein database subset of the NCBI. The most frequent subject phage among the top 10 results for each protein was selected as the outgroup phage.

Gene product prediction

Gene function prediction for type prophages was achieved by combining Virfam [28] and Phagonaute [31] web tools. Concerning Virfam, default homology thresholds were used for recombinase and connector proteins. In Phagonaute, the proteins having a HHsearch hit against Pfam ($P > 95\%$) were annotated with the Pfam information. Otherwise, the protein was searched against

all other phage proteins (probability threshold of 95% and two iterations of HHsearch) and annotated using the most prevalent annotation. If no homolog with experimentally validated function was retrieved by the second iteration, the word "putative" was indicated in the Genbank file.

Virome read treatments

Reads from 171 viromes extracted from feces of healthy subjects and patients with inflammatory bowel diseases [18] were downloaded from the EBI (fastq files, accession numbers ESR698653-ESR699023). Paired-end reads were treated with trimmomatic [51] to remove low-quality reads and unclip primer sequences present at the 3′ ends of reads, using the following command ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:15 MINLEN:36. Reads were then dereplicated with fastx uniques from the usearch9 suite. Since ~ 30% of reads were no longer paired after the trimmomatic treatment, only one read from each pair was used for mapping analyses, to remain quantitative and avoid over-representation of some phage sequences compared to others. Average read length was 250 bp. To complement this analysis, we also analyzed the viral contigs generated in the Reyes et al. study [52]. In this study, fecal viruses from 20 Malawian families were extracted at different time points and sequenced by 454 shotgun pyrosequencing. A cross-assembly strategy yielded 17,696 contigs, among which 780 had a size above 10 kb. These contigs were downloaded from ENA (PRJEB 9818) and compared with the prophages by megaBLAST alignments. Four of these contigs (> 10 kb) shared 93–97% identity to Epona, and together, the similarity regions covered 73% of Epona, but each of these contigs also contained large segments dissimilar to Epona. All four contigs originated from reads of family 57. These reads were therefore downloaded from ENA (PRJEB 9818, samples ERS795750 to 66), quality treated with trimmomatic, and dereplicated with fastx_uniques. This generated 694,650 reads, among which 680 mapped on Epona, whereas only 71 mapped to bacterial *F. prausnitzii* DNA (see the mapping method below and Fig. 4).

Quantification of *F. prausnitzii* phages in viromes

The proportion of reads homologous to each prophage was determined with BLASTn (E value < 10^{-9}, > 75% identity). To account for different genome lengths, read counts were normalized by dividing them by genome length and then multiplying the result by the average genome length (50 kb). This allows comparing phage read proportions as if all phages had a genome of length 50 kb. Since temperate phages can also be present in the bacterial DNA as prophages, we verified the level of bacterial DNA contamination of samples. An analysis of reads mapping to 16S rRNA genes with Bowtie2 [46] indicated that the vast majority of the 171 viromes had a low proportion of bacterial DNA: 166 out of 171 had less than 10^{-4} reads matching to 16S rRNA sequences. Two additional specific controls of *F. prausnitzii* DNA presence were included for the mapping analysis: a 50-kb segment corresponding to the beginning of the *F. prausnitzii* A2-165 Genbank file and a 50-kb segment encoding mainly metabolic genes (coordinates 1,451,689–1,501,688). Samples were classified as phage-positive for a given phage when the proportion of reads homologous to that phage was significantly higher than the proportion of reads homologous to these two bacterial regions (one tailed two-proportion z test, $p < 0.05$). As a control of the method, we included the highly prevalent crAssphage in the analysis and found that it is more prevalent in healthy controls than in IBD patients (Fig. 5a, two-tailed two-proportion z test, $p < 0.01$), in agreement with previously published results [17]. Inspection of prophage coverage within each virome (number of reads per 500 bp interval) revealed three situations: (i) the prophage was largely covered (> 50% length, see Fig. 4), (ii) coverage was partial (Taranis and Lugh in Fig. 4); and (iii) matches were concentrated into 1–2 contiguous prophage intervals. We excluded these last cases (less than three non-overlapping windows of 500 bp, corresponding to at most 5% of the genome) that correspond to spurious matches from various phages, mapping to a conserved region, such as an anti-repressor or an integrase gene. Additionally, to minimize the impact of these spurious matches, the median and average values of matches among all 500 bp intervals with a non-null amount of matches were computed for each phage/sample pair, and in cases were average was 10-fold higher to median, intervals displaying a number of matches above 10 times the median were considered spurious and replaced by an arbitrary value of 10 times the median. The final abundance matrix and coverage value for each phage/virome pair is reported in Additional file 10.

Viral reads assembly for phage Oengus

Given the elevated abundance of Oengus in virome ERR843984, all reads from this virome mapping on Oengus were retrieved (n = 193,414) and assembled with SPAdes (version 3.9) [53], using the following options: -k 21,33,55,77,99,127 –careful –only-assembler. This generated a 53.7-kb contig 99% identical to Oengus, covering 93% of the prophage, with a mean coverage of 322.

Sequencing of the DNA present in virions produced by the A2-165 strain

Five hundred milliliters of A2-165 culture grown overnight in LY-BHI (brain-heart infusion medium supplemented with 0.5% yeast extract [Difco], cellobiose (1 mg/ml [Sigma]), maltose (1 mg/ml [Sigma]), and cysteine (0.5 mg/ml [Sigma])), in an anaerobic chamber filled with

90% N2, 5% CO2, and 5% H2, was centrifuged at 5,200*g* for 15 min at 4 °C. Supernatant was recovered and centrifuged at 5,200*g* for 30 min at 4 °C. This step was repeated with 1 h of centrifugation. Polyethylene glycol (PEG 8000) and NaCl (Sigma) were then added to the resulting supernatant to a final concentration of 10% and 1 M respectively. The resulting mixture was kept at 4 °C overnight and then centrifuged at 5200*g* for 1 h at 4 °C. Supernatant was removed, and the phage pellet was resuspended in 2 ml of SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris pH 7.5). Sample was then treated with 0.25 μg of RNAse A (Sigma) and Dnase I (Sigma) at 37 °C for 1 h. Phage DNA was extracted using Promega kit Wizard® DNA Clean-up system. Sequencing was performed with the Ion proton sequencing technology. Alignment of obtained reads was performed using bowtie2 (-N 1 –L 32) and then visualized with Tablet, using default parameters [54]. We obtained 204,837 reads mapping to the A2-165 genome, of which 19,702 and 30,274 reads aligned to the Lagaffe and Mushu prophages respectively.

Prophage induction assay and quantitative PCR

Quantification was made as previously described [44]. Briefly, phage particles from mouse feces (see below) or from 10 ml of overnight culture in LY-BHI were recovered by PEG precipitation, as described above. For induction assays, overnight cultures of *F. prausnitzii* A2-165 were diluted in LY-BHI 25-fold. When culture absorbance reached 0.2, potent inducers were added at final concentrations of 1 or 5 μg/ml for mitomycin C (Sigma, Ref: M4287), 0.1 and 0.05% for bile salts (Sigma 48305), and 100 μM and 10 μm for hydrogen peroxide (Sigma H1009). Five hours after induction, phages were recovered from culture supernatants by PEG precipitation as described above. Samples were then treated with 10 U of Turbo DNase (Ambion) for 1 h at 37 °C and incubated at 95 °C for 30 min in 0.2 ml PCR tubes. qPCR on 100-fold diluted samples in water was performed using the Takyon ROX SYBR Mastermix blue dTTP kit (Eurogentec) and the StepOnePlus real-time PCR system (Applied Biosystem). Phage and bacterial concentrations were determined using specific primer pairs (Additional file 12). Copy numbers were calculated with a standard curve built with bacterial DNA containing the prophages, whose concentration was estimated with Qubit (Thermo Fisher). The reaction mix was the following: 7.5 μl Takyon mix, 0.9 μl dH2O, 0.3 μl of each primer (200 nM final concentration), and 6 μl of DNA diluted in H_2O. The PCR conditions were the following: 95 °C 15 s, 58 °C 45 s, and 72 °C 30s for 45 cycles, 72 °C 5 min, followed by melting curves. Results were analyzed using the StepOne Software 2.3.

Mouse colonization experiments

Germfree 7- to 8-week-old C3H/HeN mice (female) were obtained from the germfree rodent breeding facility of Anaxem-Micalis (INRA, Jouy-en-Josas, France). Animals were kept in flexible-film isolators (Getinge-La Calhène) in standard cages (two to five mice/cage) with sterile wood shavings as animal bedding. Mice were given free access to autoclaved tap water and to standard diet R03-40 (Scientific Animal food and Engineering, Augy, France), sterilized by gamma irradiation at 45 kGy (IBA Mediris). Isolators were maintained under controlled conditions of light (12 h), temperature (20 to 23 °C), and humidity (45 to 55%). To obtain dixenic *E. coli*/*F. prausnitzii*-diassociated mice, germfree mice were orally inoculated with a fresh culture of *E. coli* LF82 (10^8 to 10^9 CFU/ml). Forty-eight and 72 h later, 10^9 CFU of *F. prausnitzii* A2-165 in 0.2 ml were administered by intragastric gavage to mice pretreated with sodium bicarbonate (0.2 M, 0.1 ml by intragastric gavage, given 10 min before inoculation of bacteria). To induce colitis, 21 days after *F. prausnitzii* colonization, mice were given drinking water supplemented with 2% (*w*/*v*) dextran sulfate sodium (DSS; MP Biomedicals) for 7 days. Body weight, blood in stool, and stool consistency were monitored daily. Phage particles were recovered from one or two frozen fecal pellets unfrozen on ice for 10 min and diluted 40-fold in cold PBS. Re-suspended feces were kept on ice for 5 min with regular agitation prior to centrifugation for 10 min at 5251*g* at 4 °C. Supernatants were recovered and filtered through a 0.2-μm filter (PALL Corporation Acrodisc PF syringe filter). Samples were supplemented with 10% PEG 8000 (Sigma) and 0.5 M NaCl (Sigma). After incubation overnight at 4 °C, phage particles were harvested by centrifugation (5250*g* for 1 h at 4 °C with a swinging rotor). Supernatant was carefully removed, and pellets were re-suspended in 100 μl of SM buffer (100 mM NaCl, 8 mM MgSO4, 50 mM Tris pH 7.5) for 1 h at 4 °C.

Transmission electronic microscopy

Phages were concentrated by ultracentrifugation (20,000*g*, 1 h, 4 °C), starting from 500 ml of *F. prausnitzii* A2-165 culture. The phage pellet was resuspended in 100 μl of $MgSO_4$ 10 mM. Droplets of this preparation were directly placed on Formvar carbon-coated grids for 5 min. The grids were stained with 1% uranyl acetate and then viewed for TEM using a HITACHI HT 7700 (Elexience, France) at 80 kV. Microphotographies were acquired with a charge-coupled device camera AMT.

Additional files

Additional file 1: Accession numbers of the bacterial and type phage genomes and bacterial coordinates of all prophages. (DOCX 22 kb)

Additional file 2: Similarity matrices of phage proteins within each proposed genus. For each pair of phages, the percentage of shared proteins as well as their mean amino acid identity is indicated. For each proposed genus, the closest known phage was included in the analysis, and its host is indicated. Numbers in brackets indicate the number of ORF per genome, and most of the variability within a group is due to annotation errors. Numbers in bold indicate highly homologous prophages, which were considered to correspond to the same species. (DOCX 28 kb)

Additional file 3: Alignment of Mushu and Lagaffe with close relatives. A) Lagaffe prophages. Remarkably, Lagaffe prophage is present in the genome of *B. hansenii* DSM20583, only distantly related to *F. prausnitzii*. Alignment with a viral contig obtained from metagenomic reads suggests that the packaging site is just before the terminase gene. B) Mushu prophage is present in three *F. prausnitzii* isolates. Its synteny is perfectly conserved with a distantly related prophage in the genome of the *Ruminococcaceae* bacterium D16 (RBD16). (PPTX 466 kb)

Additional file 4: Classification of *F. prausnitzii* phage genera by Virfam. Classification of *F. prausnitzii* phage genera by Virfam (*biodev.cea.fr/virfam/*). Four genera fall into the less resolved cluster, the cluster 6. (PPTX 913 kb)

Additional file 5: DGRs in prophages. A) The position of TR and VR regions in phage genomes and the percentage of nucleotide identity between the sequences (% id), as well as the repeat length, are indicated for each type phage. In all cases, the differences are restricted to adenine residues, and the target genes containing the VR are typical of DGR. B) Genetic organization of DGRs of phages Lagaffe and Mushu, and alignment of their template repeat (TR) and variable region (VR). Red bases highlight point mutation between the two regions. Positions are given relative to the beginning of the prophage, except for Lagaffe (beginning of the encapsidated form of the genome). As in other VR/TR repeats, mismatches concern only adenine residues. C) Mapping of sequence reads of encapsidated phage DNA reveals the high variability in the VR region of both phages. Illustrations were obtained using Tablet [54]. Note that in the case of Lagaffe, the sequence corresponds to the complement of the VR sequence in B). (PPTX 194 kb)

Additional file 6: Read coverage of Lagaffe and Mushu prophages. Coverage of the bacterial DNA is represented by surrounding regions of the two prophages. Numbers correspond to positions in the A2-165 genome. Coverage pictures were obtained using Tablet [54]. (PPTX 141 kb)

Additional file 7: Lagaffe and Mushu virion concentrations in culture supernatants. Quantification of Mushu and Lagaffe virions in different conditions. Numbers represent means of genomes/ml ± s.e.m. of two to five independent experiments. Absorbance of the culture ($O.D_{600nm}$) reflects the toxicity of the compound on bacterial growth. (DOCX 13 kb)

Additional file 8: Lagaffe and Mushu virion concentration in the feces of mice. A) *F. prausnitzii* concentration (dark blue and dark red) and Mushu virion concentration (light blue and light red) in the feces of mice, as determined by quantitative PCR. DSS treatment does not modify phage and bacteria population levels compared to untreated mice. B) Variation of mice weight with time. Dots represent the average of five mice; bars represent S.E.M. Mice given DSS in drinking water loosed weight, a sign of inflammation. C) Concentration of lipocalin, a marker of intestinal inflammation [55], in mice feces. Concentration was estimated on feces after 7 days of DSS ingestion, using R&D ELISA kit (DY1857) and following manufacturer protocol. The DNA of fecal bacteria was extracted from feces resuspended in 500 μl of PBS. Suspension was centrifuged 1 min at 500*g*. Supernatant containing bacteria was recovered and mixed with 250 μl of lysis buffer (200 mM NaCl, 20 mM EDTA, 5% SDS), 250 μl of phenol/choloform/isoamyl alcohol (25:24:1;[pH 8.0], Sigma-Aldrich) and half of a tube of silica beads (100 μM MP Bio, Lysing matrix B). Bacteria were lysed using Fast-prep MPBio (5.5, 4 times 30 s). Samples were then centrifuged (13,000*g*, 3 min, 20 °C) and the aqueous phase was recovered. Four hundred microliters of chloroform/isoamyl alcohol(24:1) was added, mixed vigorously and centrifuged (13,000*g*, 3 min, 20 °C). DNA was precipitated with two volumes of ethanol and sodium acetate at 0.3 M final. (PPTX 42 kb)

Additional file 9: Whole genome dotplot of "FPlughvirus" phages. Whole genome dotplot of the six genomes of the proposed "FPlughvirus" genus reveals mosaicism. Phages share regions of high identity interspaced by regions without homology. The dotplot was realized with Gepard. (PPTX 123 kb)

Additional file 10: Number of virome reads homologous to each prophage for all 171 viromes. Number of reads homologous to each prophage (N reads), as determined with BLASTn (*E* value $< 10^{-9}$, > 75% identity). For each phage/virome pair, the number of 500 bp interval covered by reads (N int) is also reported. (XLSX 81 kb)

Additional file 11: Genome sequences of the 18 *F. prausnitzii* prophages (multifasta format). multifasta file of the 18 *F. prausnitzii* prophages. (FA 976 kb)

Additional file 12: Sequences of oligonucleotides used for quantitative PCR. (DOCX 15 kb)

Acknowledgements
We are very grateful to Chantal Bridonneau for her expertise in anaerobic culture; to Anne Foussier, Fatima Joly, and Aurélie Balvay from the Anaxem facility for their help with the animal handlings; to Christine Longin from the MIMA2 facilities (UMR 1313 GABI, INRA) for the TEM observations; to the Migale platform (INRA) for the bio-informatics environment; and to Colin Tinsley for his careful reading of the manuscript.

Funding
This work was supported by a grant from Association François Aupetit.

Availability of data and materials
The genomes of bacteria used are freely available on the NCBI website; the accession numbers are given in Additional file 1.

Authors' contributions
JC and MDP performed the prophage detection. LB provided the *F. prausnitzii* genomes. JC and EM performed the experiments. MAP performed the viral metagenomic analyses. MM performed the statistical analyses. JC, MAP, PL, HS, and MDP analyzed the results. MDP wrote the manuscript with the help of MAP and MM, and all authors revised and approved the manuscript.

Ethics approval and consent to participate
All animal procedures were carried out according to the European Community Rules of Animal Care and with authorization 1234-2015101315238694 from French services.

Consent for publication
The manuscript does not contain any individual personal data in any form.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. [2]MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France. [3]Department of General Biology, Federal University of Minas Gerais, Belo Horizonte, Brazil. [4]Sorbonne Universités, UPMC Univ. Paris 06, École normale supérieure, CNRS, INSERM, APHP Laboratoire des Biomolécules (LBM), 27 rue de Chaligny, 75012 Paris, France. [5]Department of Gastroenterology, Saint Antoine Hospital, Assistance Publique – Hopitaux de Paris, UPMC, Paris, France.

Received: 24 October 2017 Accepted: 21 March 2018
Published online: 03 April 2018

References

1. Rios-Covian D, Ruas-Madiedo P, Margolles A, Gueimonde M, de Los Reyes-Gavilan CG, Salazar N. Intestinal short chain fatty acids and their link with diet and human health. Front Microbiol. 2016;7:185.
2. Canani RB, Costanzo MD, Leone L, Pedata M, Meli R, Calignano A. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. World J Gastroenterol. 2011;17(12):1519–28.
3. Quevrain E, Maubert MA, Michon C, Chain F, Marquant R, Tailhades J, Miquel S, Carlier L, Bermudez-Humaran LG, Pigneur B, et al. Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii, a commensal bacterium deficient in Crohn's disease. Gut. 2016;65(3):415–25.
4. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux JJ, Blugeon S, Bridonneau C, Furet JP, Corthier G, et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A. 2008;105(43):16731–6.
5. Miquel S, Martin R, Lashermes A, Gillet M, Meleine M, Gelot A, Eschalier A, Ardid D, Bermudez-Humaran LG, Sokol H, et al. Anti-nociceptive effect of Faecalibacterium prausnitzii in non-inflammatory IBS-like models. Sci Rep. 2016;6:19399.
6. Rossi O, van Berkel LA, Chain F, Tanweer Khan M, Taverne N, Sokol H, Duncan SH, Flint HJ, Harmsen HJ, Langella P, et al. Faecalibacterium prausnitzii A2-165 has a high capacity to induce IL-10 in human and murine dendritic cells and modulates T cell responses. Sci Rep. 2016;6:18507.
7. Cao Y, Shen J, Ran ZH. Association between Faecalibacterium prausnitzii reduction and inflammatory bowel disease: a meta-analysis and systematic review of the literature. Gastroenterol Res Pract. 2014;2014:872725.
8. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Dore J. Low counts of Faecalibacterium prausnitzii in colitis microbiota. Inflamm Bowel Dis. 2009;15(8):1183–9.
9. Mills S, Shanahan F, Stanton C, Hill C, Coffey A, Ross RP. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut Microbes. 2013;4(1):4–16.
10. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. 2011;21(10):1616–25.
11. Bondy-Denomy J, Davidson AR. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. J Microbiol. 2014; 52(3):235–42.
12. Brussow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev. 2004;68(3):560–602. table of contents
13. Nanda AM, Thormann K, Frunzke J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. J Bacteriol. 2015;197(3):410–9.
14. De Paepe M, Leclerc M, Tinsley CR, Petit MA. Bacteriophages: an underestimated role in human and animal health? Front Cell Infect Microbiol. 2014;4:39.
15. Manrique P, Dills M, Young MJ. The human gut phage community and its implications for health and disease. Viruses. 2017;9(6)
16. Kim MS, Park EJ, Roh SW, Bae JW. Diversity and abundance of single-stranded DNA viruses in human feces. Appl Environ Microbiol. 2011;77(22): 8062–70.
17. Manrique P, Bolduc B, Walk ST, van der Oost J, de Vos WM, Young MJ. Healthy human gut phageome. Proc Natl Acad Sci U S A. 2016;113(37):10400–5.
18. Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell. 2015;160(3):447–60.
19. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome. 2017;5(1):69.
20. Benevides L, Burman S, Martin R, Robert V, Thomas M, Miquel S, Chain F, Sokol H, Bermudez-Humaran LG, Morrison M, et al. New insights into the diversity of the genus Faecalibacterium. Front Microbiol. 2017;8:1790.
21. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39(Web Server issue):W347–52.
22. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44(W1):W16–21.
23. Kiss J, Szabo M, Olasz F. Site-specific recombination by the DDE family member mobile element IS30 transposase. Proc Natl Acad Sci U S A. 2003; 100(25):15000–5.
24. Lavigne R, Seto D, Mahadevan P, Ackermann HW, Kropinski AM. Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. Res Microbiol. 2008;159(5):406–14.
25. Adriaenssens E, Brister JR. How to name and classify your phage: an informal guide. Viruses. 2017;9(4)
26. Minot S, Grunberg S, Wu GD, Lewis JD, Bushman FD. Hypervariable loci in the human gut virome. Proc Natl Acad Sci U S A. 2012;109(10):3962–6.
27. Heler R, Marraffini LA, Bikard D. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Mol Microbiol. 2014;93(1):1–9.
28. Lopes A, Tavares P, Petit MA, Guerois R, Zinn-Justin S. Automated classification of tailed bacteriophages according to their neck organization. BMC Genomics. 2014;15:1027.
29. Hulo C, Masson P, Le Mercier P, Toussaint A. A structured annotation frame for the transposable phages: a new proposed family "Saltoviridae" within the Caudovirales. Virology. 2015;477:155–63.
30. Toussaint A. Transposable Mu-like phages in Firmicutes: new instances of divergence generating retroelements. Res Microbiol. 2013;164(4):281–7.
31. Delattre H, Souiai O, Fagoonee K, Guerois R, Petit MA. Phagonaute: a web-based interface for phage synteny browsing and protein function prediction. Virology. 2016;496:42–50.
32. Guo H, Arambula D, Ghosh P, Miller JF. Diversity-generating retroelements in phage and bacterial genomes. Microbiol Spectr. 2014;2(6)
33. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, Maskell DJ, Simons RW, Cotter PA, Parkhill J, et al. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science. 2002;295(5562):2091–4.
34. Paul BG, Burstein D, Castelle CJ, Handa S, Arambula D, Czornyj E, Thomas BC, Ghosh P, Miller JF, Banfield JF, et al. Retroelement-guided protein diversification abounds in vast lineages of bacteria and archaea. Nat Microbiol. 2017;2:17045.
35. Schillinger T, Zingler N. The low incidence of diversity-generating retroelements in sequenced genomes. Mob Genet Elements. 2012;2(6):287–91.
36. Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, Stotland A, Wolkowicz R, Cutting AS, Doran KS, et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A. 2013; 110(26):10771–6.
37. Allet B. Mu insertion duplicates a 5 base pair sequence at the host inserted site. Cell. 1979;16(1):123–9.
38. Garneau JR, Depardieu F, Fortier LC, Bikard D, Monot M. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep. 2017;7(1):8292.
39. Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, Hochberg ME. Phage-bacteria infection networks. Trends Microbiol. 2013;21(2):82–91.
40. de Leeuw M, Baron M, Brenner A, Kushmaro A. Genome analysis of a novel broad host range Proteobacteria phage isolated from a bioreactor treating industrial wastewater. Genes (Basel). 2017;8(1)
41. Malki K, Kula A, Bruder K, Sible E, Hatzopoulos T, Steidel S, Watkins SC, Putonti C. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. Virol J. 2015;12:164.
42. Kim S, Ryu K, Biswas D, Ahn J. Survival, prophage induction, and invasive properties of lysogenic Salmonella typhimurium exposed to simulated gastrointestinal conditions. Arch Microbiol. 2014;196(9):655–9.
43. De Paepe M, Tournier L, Moncaut E, Son O, Langella P, Petit MA. Carriage of lambda latent virus is costly for its bacterial host due to frequent reactivation in Monoxenic mouse intestine. PLoS Genet. 2016;12(2):e1005861.
44. Diard M, Bakkeren E, Cornuault JK, Moor K, Hausmann A, Sellin ME, Loverdo C, Aertsen A, Ackermann M, De Paepe M, et al. Inflammation boosts bacteriophage transfer between Salmonella spp. Science. 2017;355(6330):1211–5.
45. Bobay LM, Touchon M, Rocha EP. Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. PLoS Genet. 2013;9(9):e1003825.
46. De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, Petit MA. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. PLoS Genet. 2014;10(3):e1004181.

47. Touchon M, Bernheim A, Rocha EP. Genetic and life-history traits associated with the distribution of prophages in bacteria. ISME J. 2016;10(11):2744–54.
48. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Pena MJ, Martinez JM, Anton J, Gasol JM, Rosselli R, Rodriguez-Valera F, et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat Commun. 2017;8:15892.
49. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 2007;23(8):1026–8.
50. Biswas A, Staals RH, Morales SE, Fineran PC, Brown CM. CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC Genomics. 2016;17:356.
51. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
52. Reyes A, Blanton LV, Cao S, Zhao G, Manary M, Trehan I, Smith MI, Wang D, Virgin HW, Rohwer F, et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc Natl Acad Sci U S A. 2015;112(38):11941–6.
53. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
54. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 2013;14(2):193–202.
55. Chassaing B, Srinivasan G, Delgado MA, Young AN, Gewirtz AT, Vijay-Kumar M. Fecal lipocalin 2, a sensitive and broadly dynamic non-invasive biomarker for intestinal inflammation. PLoS One. 2012;7(9):e44328.

ÉCOLE DOCTORALE
Agriculture, alimentation, biologie, environnement, santé (ABIES)

Title: COMPARATIVE GENOMICS OF *Faecalibacterium spp.*

Keywords: *Faecalibacterium*, phylogenomic analysis, new species, safety aspects, Probiotics.

Abstract Within the human colon, the genus *Faecalibacterium* is the main member of the *Clostridium leptum* cluster and comprises the second-most common representative genus in fecal samples, after *Clostridium coccoides*. It has been recognized as an important bacterium promoting the intestinal health and today is considered as a potential next generation probiotic. Until recently, it was believed that there was only one species in this genus, but since 2012, some studies have begun to suggest the existence of two phylogroups into the genus. This new proposition of reclassification into this genus increases the importance of new studies, with all strains, to better understand the diversity, the interactions with the host and the safety aspects in its use as probiotic. Briefly, in this work we introduce the comparative genomics analyzes to the genus *Faecalibacterium* performing a deep phylogenetic study and evaluating the safety aspects for its use as a probiotic. The phylogenetic analyzes included not only the classical use of 16S rRNA gene, but also the utilization of 17 complete genomes and techniques like whole genome Multi-Locus Sequence Typing (wgMLST), Average Nucleotide Identity (ANI), gene synteny, and pangenome. Also, this is the first work to combine an analysis of pangenome development with ANI analysis in order to corroborate the assignment of strains to new species. The phylogenetic analyzes confirmed the existence of more than one species into the genus *Faecalibacterium*. Moreover, the safety assessment involved the (1) prediction of horizontally acquired regions (Antibiotic resistance islands, Metabolic islands and phage regions), (2) prediction of metabolic pathways, (3) search of genes related to antibiotic resistance and (4) search of bacteriocins. These analyzes identified genomic islands in all genomes, but none of than are exclusive to one strain or genospecies. Also, were identified 8

genes related to antibiotic resistance mechanisms distributed among the genomes. 126 metabolic pathways were predicted and among than some were highlighted: Bisphenol A degradation, Butanoate metabolism and Streptomycin biosynthesis. In addition, we studied the genomic context of one protein (Microbial Anti-inflammatory Molecule - MAM) first described by our group. This investigation shows that MAM appears close to genes related to sporulation process and, in some strains, close to an ABC-transporter.

Titre: GÉNOMIQUE COMPARATIVE de *Faecalibacterium* spp.

Mots clés: *Faecalibacterium*, analyse phylogénomique, nouvelles espèces, Probiotiques.

Résumé: Le microbiote intestinal humain a fait l'objet de recherches approfondies au cours des dernières années et notre connaissance des espèces résidentes et de leur capacité fonctionnelle potentielle augmente rapidement. Notre intestin abrite une communauté complexe de plus de 100 milliards de cellules microbiennes qui influencent la physiologie humaine, le métabolisme, la nutrition et la fonction immunitaire, tandis que la perturbation du microbiote intestinal a été associée à des troubles gastro-intestinaux tels que les maladies intestinales inflammatoires et l'obésité. La composition du microbiote dans le Tractus Gastro-Intestinal (TGI) varie beaucoup entre les espèces et à l'intérieur de chaque espèce. La majorité des bactéries intestinales humaines appartiennent à deux phyla, les *Bacteroidetes* et les *Firmicutes*. Le phylum des *Firmicutes* comprend la bactérie *Faecalibacterium prausnitzii*, qui peut représenter 3,5% du microbiote du TGI. Dans le côlon humain, le genre *Faecalibacterium* est le principal membre du groupe *Clostridium leptum* et constitue le deuxième genre le plus commun dans les échantillons fécaux, après

Clostridium coccoides. Cette bactérie a suscité un grand intérêt en raison de ses propriétés immunomodulatrices et a été reconnue comme étant une bactérie importante, favorisant la santé intestinale et est aujourd'hui considérée comme un potentiel probiotique de prochaine génération. En outre, l'abondance et la présence de *F. prausnitzii* chez divers mammifères, tels que les porcs, les souris, les veaux et les poulets, suggèrent que cette bactérie est un membre important du microbiote et peut avoir un impact sur la physiologie et la santé de l'hôte. Des modifications de l'abondance de *F. prausnitzii* ont été largement décrites dans diverses maladies intestinales et métaboliques chez l'homme telles que le cancer colorectal (CCR), la maladie de Crohn (MC) et la recto-colite hémorragique (RCH). Jusqu'à récemment, on pensait qu'il n'y avait qu'une seule espèce dans ce genre, mais depuis 2012, certaines études ont commencé à suggérer l'existence de deux phylogroupes. Cette nouvelle proposition de reclassification de l'espèce augmente l'importance de nouvelles études pour mieux comprendre la diversité et les interactions avec l'hôte. Avec l'arrivée des nouvelles technologies de séquençage, il y a eu un dépôt massif de génomes entiers dans les bases de données publiques ce qui a amélioré le développement de la génomique comparative. Brièvement, dans ce travail, nous introduisons les analyses de génomiques comparatives sur le genre *Faecalibacterium* en effectuant une étude phylogénétique étendue et en évaluant les aspects de sécurité pour son utilisation en tant que probiotique. Les analyses phylogénétiques comprennent non seulement l'utilisation classique du gène de l'ARNr 16S, mais aussi l'utilisation de 17 génomes et techniques complètes comme le typage de séquence multilocus (wgMLST), l'identité nucléotidique moyenne (ANI), la synténie de génomes et le pangenome. Il s'agit du premier travail combinant une analyse du développement des pangenomes à une analyse de l'ANI afin de corroborer l'affectation de souches à de nouvelles espèces. Les analyses phylogénétiques ont confirmé l'existence de plus d'une espèce dans le genre *Faecalibacterium*. Par

ailleurs, l'évaluation de la sécurité impliquait : (1) la prédiction des régions acquises horizontalement (îlots d'antibiotiques, îlots métaboliques et régions phagiques) ; (2) la prédiction des voies métaboliques ; (3) la recherche de gènes liés à la résistance aux antibiotiques et (4) la recherche des bactériocines. Ces analyses ont permis d'identifier des îlots génomiques dans tous les génomes, mais rien n'est exclusif à une ou plusieurs souches. En outre, ont été identifiés 8 gènes liés aux mécanismes de résistance aux antibiotiques répartis dans les génomes. Cent vingt-six voies métaboliques ont été prédites et parmi elles: la dégradation du bisphénol A, le métabolisme du butanoate et la biosynthèse de la streptomycine. Aussi, nous avons étudié le contexte génomique (synténie) d'une protéine, la Molécule Anti-inflammatoire Microbienne (MAM) décrite pour la première fois par notre groupe. Cette étude montre que la MAM apparaît proche des gènes liés au processus de sporulation et, chez certaines souches, proche d'un transporteur-ABC. En tout, ces travaux ont présenté des études pionnières sur les aspects génomiques du genre *Faecalibacterium* et ont permis de mieux comprendre la diversité du genre, son interaction avec l'hôte et ses importantes caractéristiques probiotiques. Les résultats présentés ici ouvrent de nouvelles perspectives pour les études futures de ces bactéries. D'autres études ont montré que dans certaines maladies, l'abondance de *Faecalibacterium* varie entre les deux phylogroupes proposés auparavant. Ensuite, l'existence d'espèces distinctes et les caractéristiques génomiques présentées ici, contribueront à améliorer nos connaissances sur la façon dont ces bactéries peuvent agir en tant que capteur et promoteur de la santé humaine et quel est leurs rôles dans les maladies intestinales. Les études des souche spécifiques pourraient non seulement expliquer la variation de l'abondance de *Faecalibacterium* spp. associée à certaines maladies (p. ex. dermatite atopique, maladie de Crohn et colite ulcéreuse), mais offrent également de nouvelles possibilités pour le diagnostic et le traitement spécifique de l'inflammation intestinale et des

maladies associées. Par conséquent, il est important d'effectuer davantage d'analyses métagénomiques en tenant compte de diverses maladies et de séquencer davantage de génomes, en particulier ceux isolés chez des volontaires malades, ce qui facilitera notre compréhension de la biologie de ces bactéries.