UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dottorato di ricerca in Scienze, Tecnologie e Biotecnologie Agro-alimentari

Ciclo XXXIV

Mucin degraders from human gut microbiota

Candidato: Musmeci Eliana

Relatore: Prof.ssa Rossi Maddalena

Correlatore: Dott. Raimondi Stefano

Coordinatore del Corso di Dottorato: Prof. Ulrici Alessandro

Abstract

Mucus is a viscoelastic gel barrier that covers wet epithelial surfaces throughout the body including gastrointestinal (GI), respiratory, reproductive, and urinary tracts (Alemao et al., 2020; Bansil & Turner, 2018). It is secreted by goblet and mucous cells and achieves a complete layer already several days after birth (Bunesova et al., 2018). Different roles and functions are carried out by mucus barrier such as gaseous exchange, nutrient and cofactor adsorption, lubrification, chemical sensing and an important relationship with the immune system (Anthony P. Corfield, 2015). The major building block of mucus are represented by mucins glycoproteins (Bansil & Turner, 2018). Mucins are characterised by an high level of glycosylation, oligosaccharides such as *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), fucose, galactose and sialic acidic, constitute up to 80% of their molecular mass (Bansil & Turner, 2006).

The aim of this study was to investigate gut mucin degraders of healthy subjects through a culture dependent and independent approach. The faeces of five healthy adults were subjected to three steps of enrichment in a medium with only mucins as carbon and nitrogen sources. The bacterial community was compared before and after the enrichment by 16S rRNA gene profiling. Only three species of strict anaerobes able to grow on mucin were isolated: *Clostridium celatum*, *Clostridium tertium*, and *Paraclostridium bifermentans*. Genome analysis of the strains was carried out and compared with other available genome present on NCBI GenBank database to better understand their metabolic and functional potential, and to determine their ability to use mucin. Genes coding for glycoside hydrolases (GHs) involved in mucin degradation were found in all the genomes, with a higher abundance in *C. celatum* that possess at least 25 GHs. The distribution of genes required for utilization of Gal, GlcNAc and GalNAc were widespread among all the strains, while only *C. celatum* degrades fucose.

The three strains were investigated also in terms of physiological characterization and functional properties. They all were able to grow in a pH range between 5.5 and 8, with an optimum of pH 6.5-7, producing ethanol, acetic, propionic, formic acids, and hydrogen. Biofilm-forming ability was observed in *C. celatum* and *P. bifermentans.* Like other members of Clostridia class, the three clostridia produced spore that resisted to stress conditions. To evaluate whether spore formation improve survival, tolerance to oxygen and high temperature exposure were checked. The three strains were able to resume vegetative growth after exposure to oxygen, albeit *C. celatum* resulted more susceptible to oxygen stress, whereas heat treatment up to 80°C caused a decrease of maximum one Log.

In vivo experiment on Balb/c mice are in progress, to test the ability of the three strains to colonize intestinal mucus and to exert some effect on immune system. To assess their impact, the expression of IL-1β and IL-10 will be measured after the treatment.

Abstract

Il muco è un colloide viscoso che costituisce una barriera fisica in tutte le superfici epiteliali del corpo umano, inclusi il tratto gastrointestinale, respiratorio, riproduttivo e urinario (Alemao et al., 2020; Bansil & Turner, 2018). É secreto dalle cellule mucipare che producono uno strato protettivo completo già dopo pochi giorni dalla nascita (Bunesova et al., 2018). La barriera mucosa è coinvolta in più funzioni come ad esempio l'assorbimento di cofattori e nutrienti, la lubrificazione ed inoltre svolge un ruolo fondamentale in correlazione con il sistema immunitario (Anthony P. Corfield, 2015). Le mucine sono glicoproteine e sono le maggiori costituenti del muco (Bansil & Turner, 2018). Queste ultime sono caratterizzate da un alto livello di glicosilazione, sono composte da oligosaccaridi quali N-acetilglucosammina (GlcNAc), N-acetilgalattosamina (GalNAc), fucosio, galattosio e acido sialico, i quali rappresentano fino all'80% della massa molecolare (Bansil & Turner, 2006).

Lo scopo di questo lavoro è stato quello di identificare, attraverso un approccio cultura dipendente e indipendente, la popolazione batterica intestinale isolata in soggetti sani che fosse in grado di degradare la mucina. Le feci di cinque adulti sani sono state sottoposte a tre step successivi di arricchimento in un terreno di crescita contenente la mucina come unica fonte di carbonio e azoto. La popolazione batterica è stata confrontata prima e dopo l'arricchimento mediante la metodica del 16S rRNA *gene profiling*. Sono state isolate solamente tre specie di anaerobi capaci di sviluppare su mucina: *Clostridium celatum, Clostridium tertium e Paraclostridium bifermentans*.

Di queste tre specie è stata effettuata l'analisi genomica e in parallelo un'analisi comparativa con altri genomi presenti nella banca dati NCBI GenBank, in modo da approfondire gli aspetti funzionali e metabolici, oltre che determinare la loro capacità di utilizzo della mucina. In tutti i genomi analizzati, sono stati identificati geni codificanti glicosidasi (GH) coinvolte nella degradazione della mucina, con un massimo di 25 GH riscontrate nel genoma di *C. celatum.* I geni coinvolti nei *pathway* di utilizzo dei singoli monosaccaridi quali Gal, GlcNAc and GalNAc, sono stati riscontrati in tutti i tre genomi, mentre solo *C. celatum* è in grado di utilizzare il fucosio. I tre ceppi oggetto di studio sono stati inoltre caratterizzati dal punto di vista fisiologico. In tutti i ceppi è stata registrata crescita batterica in un intervallo di pH fra 5.5 e 8, con un optimum di crescita di pH 6.5-7, producendo etanolo, acido acetico, propionico, formico e idrogeno. È stata osservata una propensione a formare biofilm solamente in *C. celatum* e *P. bifermentans*. Come altri membri della classe Clostridia, i tre clostridi sono in grado di produrre spore, attraverso le quali resistere a condizioni di stress. Per testare se la formazione di spore aumentasse la sopravvivenza, sono stati effettuati dei

trattamenti di esposizione all'ossigeno e alle alte temperature. In seguito all'esposizione all'ossigeno tutti i ceppi sono stati in grado di ripristinare lo sviluppo delle cellule vegetative, sebbene in *C. celatum* sia stata osservata una maggiore sensibilità allo stress indotto da ossigeno. Diversamente, il trattamento con temperature fino a 80°C ha causato una riduzione massima di circa un ordine logaritmico in tutte le specie analizzate.

Infine, sono in corso test *in vivo* su modelli murini Balb/c, per testare la capacità dei tre isolati di colonizzare la mucosa intestinale e di produrre effetti sul sistema immunitario. Per determinare il loro impatto sul modello murino a seguito del trattamento, verrà valutata l'espressione di interleuchine come IL-1β e IL-10.

Table of contents

1. Introduction

1.1 The human gut microbiota

1.1.1 The no longer “forgotten organ”

Human beings, like all the other organisms, live a symbiotic relationship with their coevolved microbiota (Rajilić-Stojanović & de Vos, 2014). Almost 40 trillion of microbes inhabit the human body (Sender et al., 2016), organized in different microbial communities widespread on external surfaces covered by skin and on internal surfaces lined with mucus (Wang et al., 2021). The most concentrated district is the human gastrointestinal tract, colonized by a complex ecosystem of microorganisms, including bacteria, fungi, archea and viruses, collectively referred to as the gut microbiota (Puértolas-Balint & Schroeder, 2020).

It has been estimated that an amount of almost 10^{11} cells/g microorganisms reside in human colon (Rossi et al., 2016), number that is of the same order as the number of human cells, while in small intestine reaches 10^{6}-10^{7} cells/g due to the low pH and the rapid flow of content of these regions (Sender et al., 2016). The entire microbial genome called “microbiome” express 100 times more genes than the human genome (Gao et al., 2021), so the human intestinal microbiota can be conceived as an “organ” that exceeds the liver, which is the most versatile human organ, in terms of biochemical conversions (Egert et al., 2006).

Gut commensal bacteria play pivotal and essential roles: they digest substrate inaccessible to host enzymes, synthetize vitamins, are involved in maturation and modulation of immune system, avoid the proliferation of harmful microorganisms, maintain intestinal epithelial homeostasis and overall human health (Kostic et al., 2014; Ruan et al., 2020).

Albeit gut microbiota was considered for long time as the “forgotten organ”, in the last 10-15 years the bloom of research has allowed a better comprehension about its composition, functions, and interactions with the host (Marchesi et al., 2016; Seo & Holtzman, 2020).

1.1.2 Gut microbiota composition

Gut microbiota is mainly composed by bacteria and its composition (in terms of species and relative abundance) differs not only between individuals but also between different regions of the gut (Kudelka et al., 2020). Although humans harbour more than 200 different bacterial species, almost 30% of this population is shared by different subjects, defining the core microbiota (Baldelli et al., 2021). Most of the identified gut bacteria species are strictly anaerobic and are located in mucosal regions of the gastrointestinal (GI) tract.

Bacterial species detected in the human gut are classified in seven different phyla, with two dominant phyla, Firmicutes and Bacteroidetes, and subdominant ones like Actinobacteria, Proteobacteria, Verrucomicrobia, Fusobacteria, and Tenericutes (Meng et al., 2018; Pierro, 2021).

Along the GI tract, stomach, small intestine, and colon are characterized by the colonization of different bacterial communities. Acidic environment of the stomach, due to gastric acids and proteolytic enzymes, limits the presence of bacteria. Typically, taxa such as Bacillales *incertae sedis*, Streptococcaceae, Enterobacteriaceae, Leptotrichiaceae, Veillonellaceae, and Pseudomonadaceae are found in corpus and antrum of the stomach (Vasapolli et al., 2019). Duodenum, jejunum, and ileum compose the small intestine, and are the districts where most of the nutrient digestion and absorption take place. Jejunum is dominated by facultative anaerobes like *Streptococcus, Prevotella, Veillonella, Fusobacterium, Escherichia, Klebsiella,* and *Citrobacter* (Sundin et al., 2017). Duodenum exhibits similar genera composition as the stomach, while the ileum encompasses Clostridiaceae, Lachnospiraceae, Peptostreptococcaceae, Ruminococcaceae, Enterobacteriaceae, and Bacteroidaceae that form a microbiota more similar to the microbial communities of colon (Vasapolli et al., 2019).

Colon is the last section of the GI tract, and in this district complex carbohydrates fermentation occurs by the action of bacteria ascribing to the five main bacterial phyla Bacteroidetes, Firmicutes, Verrucomicrobia, Proteobacteria, and Actinobacteria. A different spatial organization occurs in the colon, where microorganisms establish an axis from the lumen to the mucosa coating the epithelium (Ruan et al., 2020).

Bacteroidetes and Firmicutes represent most of the bacteria inhabiting the colon. The phylum Firmicutes is highly diverse and encompasses the orders Lactobacillales, Clostridiales, Erysipelotrichia, and Negativicutes. The order Lactobacillales includes some of the most popular health promoting bacteria, and species commonly used as probiotics (Gareau et al., 2010). The order Clostridiales includes obligate anaerobic bacteria with wide phylogenetic and metabolic diversity. Clostridiales obtain energy by a number of fermentative pathways and contribute to feed other microbes and colonocytes with a plethora of metabolites, being involved in complex cross-feeding

interactions. The main butyrate producing bacteria, remarkably important to promote a health status, are members of Clostridiales (Sorbara & Pamer, 2022). On the other side, some Firmicutes members have been associated with obesity or to exacerbation of inflammatory bowel disease (IBD), such as *Ruminococcus gnavus* (Hall et al., 2017).

Bacteroidetes can represent the major taxa of intestinal bacteria, most ascribed to the families Bacteroidaceae, Prevotellaceae, Rikenellaceae, and Porphyromonadaceae. Bacteroides express a strong hydrolytic activity, that allows degradation and fermentation of ingested complex polysaccharides (Wexler & Goodman, 2017). Some species negatively affect the human health, such as *Prevotella copri* (Scher et al., 2013) associated with rheumatoid arthritis and *Bacteroides fragilis* with colon cancer (Wu et al., 2009), and other are recognized as main proteolytic fermenters.

The phylum Verrucomicrobia includes *Akkermansia muciniphila*, known to decompose mucin in the human intestine (Derrien et al., 2004). This abundant species regulates glucose homeostasis, inflammation, intestinal barrier function, adipose tissue metabolism, fat mass accumulation, and other metabolic functions of the host. The proximity of *A. muciniphila* to the human intestinal favours the interaction between this bacterium and the host, contributing to control the physiological and homeostatic functions that hinder the onset of obesity and type 2 diabetes (Everard et al., 2013).

Actinobacteria encompasses the genus *Bifidobacterium*, that includes major bacteria associated with early life (Sorbara & Pamer, 2022). The passage from infancy to adolescence and then to adulthood is related to a decline of the relative abundance of bifidobacteria, that still remain a main genus of health promoting bacteria. For this reason, bifidobacteria are regarded as the most robust and validated probiotics, with evidence of effect on treatment or prevention of several diseases (Agans et al., 2011).

Proteobacteria is the phylum that includes Gram-negative bacteria. The order Enterobacterales is the most abundant among Proteobacteria, is composed by permanent colonizers of the human gut that (Sorbara & Pamer, 2022). In healthy conditions, these opportunistic bacteria are a minor component of the microbiota, and as along as the microbiota is balanced, their overgrowth is limited. A bloom of Enterobacterales may occurs as a result of microbiota dysbiosis, yielding pathogen-mediated infections and activating inflammatory host response (Amaretti et al., 2020).

1.1.3 Microbial colonization throughout life

1.1.3.1 The infant gut microbiota

When human first encounters microbial population remains a debated theme (Silverstein & Mysorekar, 2021). Over the years, several studies have supported the hypothesis that microbial colonization occurs in utero, with the detection of bacteria in human placenta, amniotic fluid, and fetal membranes. Many of these "first" bacteria were detected in placenta samples through a whole-genome shotgun metagenomic approach, and ascribed to the genera *Lactobacillus, Bifidobacterium, Propionibacterium, Rhodococcus, Streptomyces, Bacteroides, Prevotella, Neisseria, Enterobacteria* and *Micrococcus,* with *E. coli* as the most abundant species. The microbiota composition described had a match with the taxa found in the maternal oral microbiome (Aagaard et al., 2014; Rackaityte et al., 2020). Recently, Mishra and colleagues (Mishra et al., 2021) demonstrated the presence of low microbial concentration in fetal tissues, during the 2nd trimester of pregnancy, with an enrichment of seven bacterial genera. Among the isolates, mostly facultative or obligate anaerobes, they identified *Staphylococcus* and *Lactobacillus* like the most abundant. However, other researchers defend the sterile womb hypothesis ascribing the detected microbial particles to contamination (Senn et al., 2020).

After birth GI tract of newborns is rapidly colonised and its microbiota composition is initially shaped by three main factors: gestational age at birth, delivery mode, and infant feed mode (Milani et al., 2017). The first cause influencing the microbiota composition is the gestational age of birth. Indeed microbiota diversity changes between full-term and pre-term neonates, being characterized in the latter group by the presence of higher abundances of pathogenic facultative anaerobes like *Enterococcus, Enterobacter, Klebsiella*, and *Staphylococcus* (Grech et al., 2021).

In addition, gut bacteria composition of infants is influenced by the mode of delivery. After natural childbirth, infants harbour bacteria such as *Prevotella* and *Lactobacillus* resembling the mother's vaginal microbiota. Differently, in case of caesarean section gut microbiota differs not only in terms of composition but also in timing of colonization, and babies acquire microorganisms like *Staphylococcus, Corynebacterium,* and *Propionibacterium* that mimic "skin"-type bacteria (Gorkiewicz, 2018; O'Hara & Shanahan, 2006). Several studies have demonstrated that during the first week of life infants harbour complex and dynamic microbial communities (Favier et al., 2002; Palmer et al., 2007).

At birth, the gut is colonized by primarily aerobic bacteria, ascribed to the genera *Enterobacter, Staphylococcus,* and *Streptococcus,* many of which potentially harmful for the host.

These early colonizers begin to change the gut environment consuming oxygen e preparing an anaerobic environment, allowing the colonization by an increasingly number of anaerobic microorganisms (Pop, 2012). Overall, the composition of the gut community continuously changes in response to environmental factors, like type of feeding, weaning, and the successive introduction of different type of foods, reaching a sort of mature equilibrium around one to three years (Aagaard et al., 2014; Wopereis et al., 2014). In this time framework, microbial communities of infant are highly individuals and characterized by a lower diversity index compared to the childhood and adulthood microbiota diversity (Grech et al., 2021).

Another major cause that shape gut colonization development is represented by infant fed mode (Milani et al., 2017). Breast milk contains oligosaccharides (HMO-Human Milk Oligosaccharides) that constitute the third solid component after lactose and fat. HMOs are complex carbohydrates indigestible for the host, and provide a fermentative substrate for the growth of a limited number of gut bacteria. These oligosaccharides act as prebiotic stimulating in particular the growth of bacteria of the genus *Bifidobacterium*. The presence of this latter group of bacteria is related to a health status and to strengthening of gut mucosal protection against pathogens (Cresci et al., 2016). On the other side, microbiota of formula-fed infants' is characterized by lower abundance of bifidobacteria and lactobacilli, and higher diversity due to the presence of *Bacteroides*, *Staphylococci*, *Enterococci*, and *Clostridia* (Milani et al., 2017).

The infant gut microbiota undergoes a transition from being dominated by *Lactobacillus* and *Bifidobacterium* to being dominated by a more complex population of bacteria (Grech et al., 2021). Lactobacilli and bifidobacteria promote the development of infant-acquired immunity and innate immunity in early life, and their relative abundance is related to the risk of developing noncommunicable diseases, such as obesity and asthma (Björkstén et al., 2001). The introduction of solid foods and dietary fiber fermented by the gut microbiota around 6 months of age, establishes a microbial community with higher level of diversity. After weaning, it is generally observed a replacement of Proteobacteria (Enterobacteriaceae) and Actinobacteria (mainly *Bifidobacterium)*, and a concurrent bloom of anaerobes and fermenters like *Bacteroides*, *Clostridium*, *Lachnospiraceae* and *Ruminoccocaceae*. (Koenig et al., 2011; Milani et al., 2017). The beneficial commensal bacteria *A. muciniphila* and *F. prausnitzii* absent in the early infancy, increase in abundance in children at 24 and 12 months, respectively (Yassour et al., 2016).

1.1.3.2 The adult-like microbiota

Dynamic changes of human gastrointestinal microbiota occur over the whole life. Albeit the adulthood microbiota has been considered for long time quite stable, other evidence suggested its continuous development from infancy to adult state (Hollister et al., 2015).Children present a higher abundance of *Bifidobacterium*, and their gut microbiota composition is mainly characterized by a lower diversity compared to the adults. However, during childhood, the phyla of Firmicutes and Bacteroidetes start being the most dominant, like in adults (Moran-Ramos et al., 2020; Ringel-Kulka et al., 2013).

This trend of microbiota composition occurs also in the adolescent stage of life (11-18 years). All these observations are in line with the hypothesis that bifidobacteria slowly decrease in children from 2 to 18 years, reaching stable level in adult microbiota (Paliy et al., 2009).

However, adolescent and adult microbiota exhibit similar amounts of *Clostridia*, especially of the genus *Ruminococcus*, which is consistent with the role of these members as the primary carbohydrate fermenters (Agans et al., 2011).

In adult, gut microbiota is mainly dominated by the phyla Firmicutes (especially the order *Clostridiales*) and Bacteroidetes, which constitute up to 90% of total bacteria, followed by Proteobacteria (up to 10%) and Actinobacteria (up to 5%) (Paliy et al., 2009).

1.1.3.3 The aging of gut microbiota

Additional changes on the gut microbiota composition occur in elderly people, where the decline of diversity is due in part to a decline of the immune system functions (immunosenescence). This latter feature of aged people concurs to the establishment of a chronic low-grade inflammatory status, which together with changes in dietary life style is responsible of evolution towards an unbalanced microbiota (Fan & Pedersen, 2021; Toward et al., 2012). Gut microbiota composition of elderly is characterized by increased abundance of facultative anaerobes and by a shift in term of the main species that inhabit gastrointestinal tract (Mariat et al., 2009).

In elderly, the status of inflammation is exacerbated by a decrease in the anti-inflammatory bacterial species, including *F. prauznitzii* and bifidobacteria spp. (Thevaranjan et al., 2017). A reduction of Firmicutes and an increase of Bacteroidetes occurs from adulthood to old age, resulting in a decreased Firmicutes to Bacteroidetes ratio (Quercia et al., 2014; Vaiserman et al., 2020). As these phyla represent the majority of gut microbiome, the Firmicutes/Bacteroidetes ratio is critical for

the production of short chain fatty acids (SCFAs) such as butyrate and propionate. Accordingly, in a Japanese cohort of healthy subjects, following the changes of microbiota composition from 0 to 104 years, the phylum Firmicutes resulted the most abundant, and its relative abundance started to decrease at the age of 70. This trend was opposite to the increasing abundance of Proteobacteria which occurred in elderly people. Furthermore, in old age the phylum Bacteroidetes became more abundant in a stepwise manner, whereas Actinobacteria decreased progressively (Odamaki et al., 2016).

Similar results were previously obtained in a study that compared intestinal microbiota of young adults (≥30 years old), elderly (70 years old) and centenarians (100 years old) (Biagi et al., 2010): in the latter group a shift in gut bacteria composition involved a significant decrease of Firmicutes subgroup *Clostridium* cluster XIVa, and an increasing abundance of bacilli and Proteobacteria. Within the Firmicutes, the families of Ruminococcaceae, Lachnospiraceae, and Clostridiaceae encompass the main butyrate-producing bacteria of the human gut that positively influence host health (Louis & Flint, 2017). In the study of Biagi and coworkers (2010), the species *Ruminococcus obeum, Roseburia intestinalis*, *E. ventriosum, E. rectale, E. hallii, Papillibacter cinnamovorans*, and *F. prausnitzii* were instead less abundant in centenarians. On the other side, a bloom of the Proteobacteria *Alcaligenes faecalis, Leminorella, Proteus, E. coli, Haemophilus, Klebsiella pneumoniae, Pseudomonas, Serratia, and Yersinia*, positively correlated with the higher level of pro-inflammatory cytokines IL-6 and IL-8 (Biagi et al., 2010).

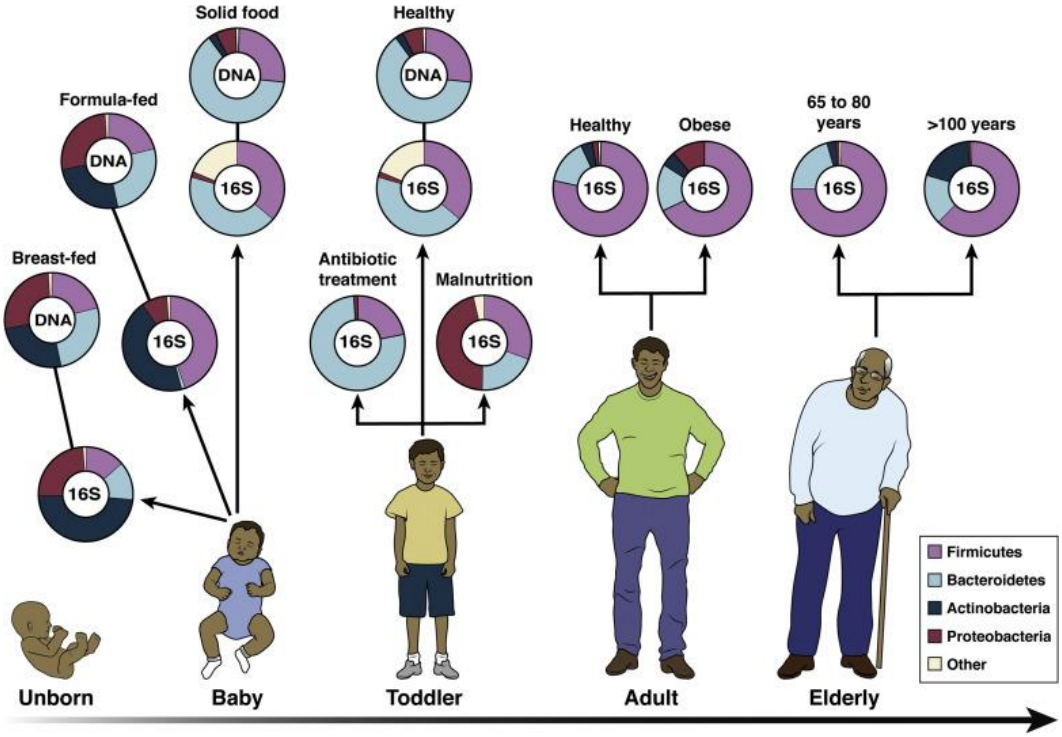


Figure 1 The structure of the human intestinal microbiota across the life cycle. The composition of gut microbiota changes throughout life. Dynamic changes of the bacterial community take place into infant gut, being shaped by different factors such as, mode of delivery, gestational age at birth, feeding mode, and weaning. From 2 and 18 years, microbiota composition begins to resemble the "adulthood" structure, differing from the latter by the higher abundance of *Bifidobacterium*. In adult, the phyla Firmicutes and Bacteroidetes are the most abundant, followed by Actinobacteria and Proteobacteria. The elderly's microbiota is characterized by increasing abundance of Bacteroidetes and Proteobacteria. This trend continues in centenarians, in those is also present a concomitant decrease of the phylum Firmicutes. (Haran & McCormick, 2020)

1.1.4 Saccharolytic and proteolytic metabolism

The intestinal microbiota gains energy and nourishment from polysaccharides and peptides which are indigestible from human's enzymes (Ottman et al., 2012).

It is estimated that around 20-30 g of diet-derived carbohydrates reach the colon every day escaping digestion by the enzymes of the small intestine, and providing the carbon sources for the saccharolytic bacteria that are the vast majority of the intestinal microbiota. Most of the carbohydrates reaching the colon are resistant starches (RS), non-starch polysaccharides (NSP) including cell wall polysaccharides, oligosaccharides, and some mono- and di-saccharides (G. T. Macfarlane & Macfarlane, 2011).

In the large intestine fermentative metabolism represents the major energy source for intestinal bacteria, with the polymers broken down by various bacterial enzymes such as polysaccharides, glycosidases, proteases, and peptidases to smaller oligomers and their component,

mostly sugars and amino acids. The metabolism of carbohydrates yields the short chain fatty acids (SCFA) acetate, propionate, and butyrate, and lowers the pH in the large intestine. These three main metabolites, produced in a ratio of 3:1:1, differ in their utilization pathway and in tissue distribution, but together participate to maintain gut functions and host immune homeostasis (Scott et al., 2013). Others minor metabolites such as ethanol, lactate, and succinate do not accumulate in the colon because they are metabolized by other cross-feeding species (Louis & Flint, 2017).

The production of oxidized substances as acetate is directly associated with ATP generation, for this reason acetate is the most abundant SCFA released in the colon. It is produced by anaerobic bacteria as a fermentation product and by acetogenic bacteria, as *B. hydrogenotrophica,* from H_2, CO_2, and formate (Anand et al., 2016). Acetate is is released in peripheral tissues, reach the highest concentration in blood, and is involved in lipogenesis (Louis & Flint, 2017). Propionate is mostly absorbed by the liver where contributes to gluconeogenesis, and is produced through the succinate and propanediol pathway. The amount of propionate released depends on the relative abundance of Bacteroidetes and some Firmicutes. In particular, some bacteria from the *Bacteroides* genus, and also some members of Ruminococcaceae like *R. flavefaciens*, are able to form propionate predominantly by succinate pathway. The propanediol pathway produces propionate from deoxy fucose and rhamnose, and is utilized by members of the Lachnospiraceae family, such as *Roseburia inulivorans* and *Blautia spp.* (Reichardt et al., 2014; Thursby & Juge, 2017).

Butyrate is mainly released by the families Ruminococcaceae, Lachnospiraceae, Erysipelotrichaceae, and Clostridiaceae, from carbohydrate fermentation via glycolysis and acetoacetyl-CoA pathways. The acetate CoA transferase pathway is performed by the most abundant species that colonize a healthy gut microbiota, including *Fecalibacterium prausnitzii, Roseburia* spp., *Eubacterium rectale, Eubacterium halii* and *Anaerostipes* spp., while several members of the genus *Clostridium* produce butyrate via phosphotransbutyrylase and butyrate kinase (Louis & Flint, 2009; Thursby & Juge, 2017).

Butyrate, succinate and propionate are rapidly adsorbed from the gut lumen and serve as nutrients for the colonocytes. Moreover, they improve solubility and absorption of minerals such as calcium. SCFA also modulate colonic regulatory T cells and influence the down regulation of pro inflammatory cytokines, such as IL-6 ad IL-12, in colonic macrophages (Chung et al., 2012; Smith et al., 2013). The metabolism of dietary complex carbohydrates counteracts the proteolytic metabolism and reduces the concentrations of potentially harmful metabolites derived from proteolytic activity in the colon (Hamer et al., 2012).

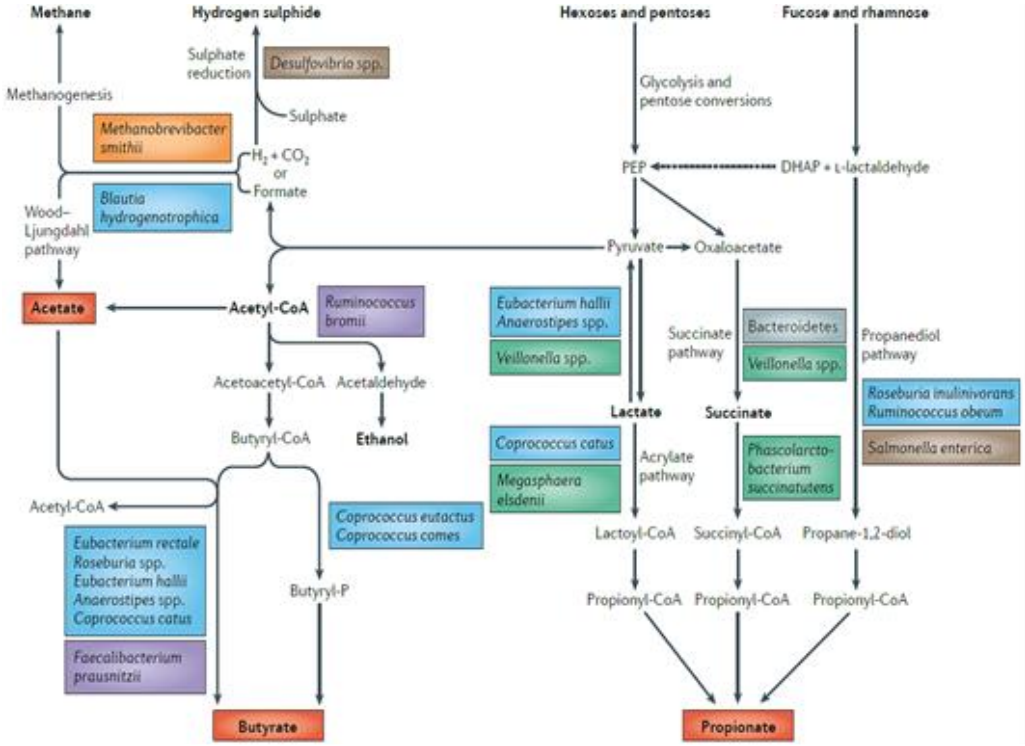


Figure 2 Overview of saccharolytic catabolic pathway carried out by intestinal bacteria (Adapted from Louis et al., 2014)

Another source of carbon and energy for colonic bacteria is represented by proteins. It was estimated that every day an amount of 13 g of protein and peptides reach the human colon. Most of dietary proteins are digested and absorbed in the small intestine. Undigested proteins and peptides that reach the colon are fermented with production of carbon dioxide, ammonia, linear and branched organic acids (acetate, butyrate, propionate, valerate, isobutyrate, 2-methylbutyrate, and isolavalerate), phenols, indole derivatives, and other nitrogen or sulphur-containing compounds (Amaretti et al., 2019; Louis & Flint, 2017).

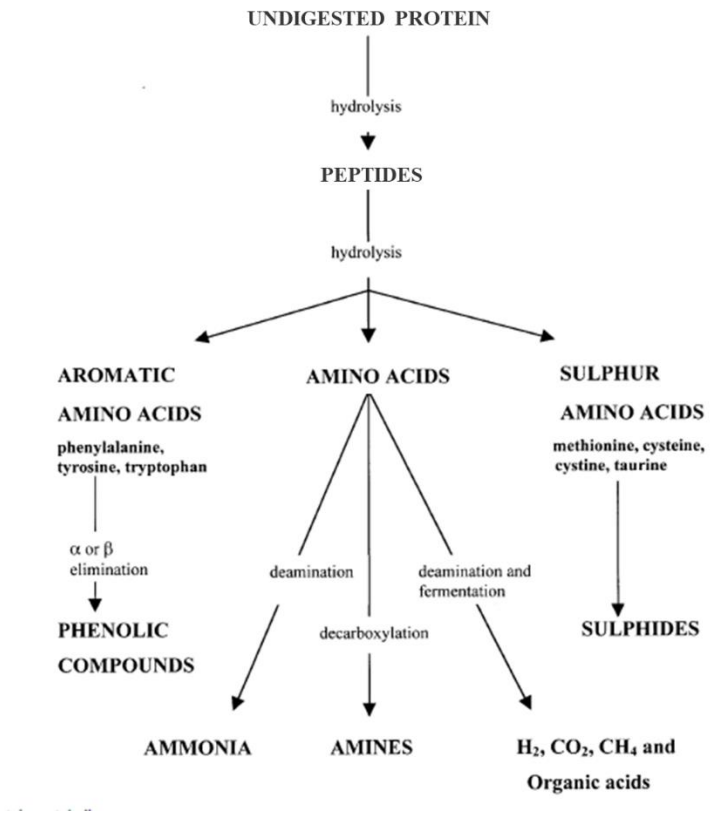


Figure 3 Overview of protein catabolic pathways carried out by intestinal bacteria

Many of these products are detrimental to human health (in particular ammonia, phenols, indoles, amines, sulfides and *N*-nitroso compounds), being responsible for a variety of harmful effects, including systemic toxicity, nephrotoxicity, and carcinogenesis (Guo et al., 2020; Kobayashi, 2018).

In the colon, the first steps in protein breakdown are carried out by a variety of hosts endopeptidases and bacterial proteases (Cummings & Macfarlane, 1991), that release peptides and/or amino acids fermented by true amino acid-fermenting through the Stickland reactions or other fermentation pathways (Kim et al., 2004; Smith & Macfarlane, 1998).

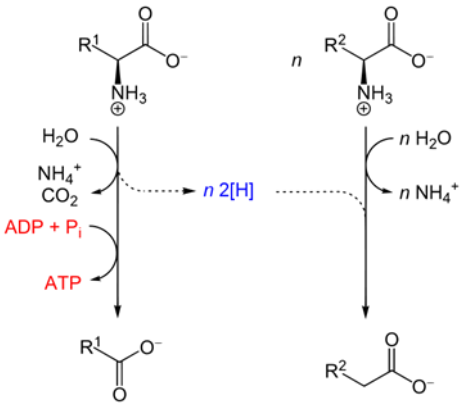


Figure 4 General scheme of Stickland fermentation of amino acids

According to MacFarlane and Cummings (1991), intestinal bacteria able to ferment proteins are ascribed to the genera *Bacteroides*, *Propionibacterium*, *Clostridium*, *Fusobacterium*, *Streptococcus*, *Eubacterium* and *Lactobacillus*. In particular, the species *Bacteroides thetaiotaomicron*, *B. eggerthii*, *B. ovatus*, *B. fragilis*, *Parabacteroides distasonis*, *Clostridium bartlettii*, and *Eubacterium hallii*.

A recent paper (Amaretti et al., 2019) aimed to identify intestinal proteolytic bacteria of healthy subjects by a metagenomic approach on protein enriched fecal cultures, revealed that the families Enterobacteriaceae, Burkholderiaceae, and Desulfovibrionaceae, and the genera *Escherichia-Shigella*, *Sutterella*, *Parasutterella*, and *Bilophila* proliferated in the protein based medium. Furthermore, many taxa that get enriched were ascribed to the families Lachnospiraceae and Ruminococcaceae and showed the strongest correlation with products of amino acid fermentation such as ammonium, indole, and p-cresol. In particular, the genera *Anaerotruncus*, *Dorea*, *Oscillibacter*, *Eubacterium oxidoreducens*, *Lachnoclostridium*, *Paeniclostridium*, *Rombutsia* took advantage in the protein based medium.

1.1.5 Gut dysbiosis

The intestinal microbiota impacts the host by releasing metabolites, cellular components, and secreted proteins that can activate a wide range of host receptors mediating effects that range from inflammation to immunosuppression (Rooks & Garrett, 2016). The rich microbial community play crucial role in maintaining host physiology and homeostasis (Agus et al., 2021). However, different factors can contribute to a shift in the microbiota composition, disrupting microbial homeostasis and leading to dysbiosis (Ventura et al., 2018). Generally, a disease status is associated to a reduced microbiota diversity that can derive, for instance, from fibre-deficient diets and antibiotic treatments (Egert et al., 2006). Gut dysbiosis leads to altered metabolism and immune regulation of the host, moreover the lost of equilibrium between commensal gut bacteria and opportunistic pathogens causes a major vulnerability to damage and loss of intestinal homeostasis (Di Tommaso et al., 2021).

The healthy gut microbiota handles an active crosstalk with the host cells. A disturbed ecology of the microbial community has been associated with various human diseases, such as obesity and metabolic syndromes, non-alcoholic fatty liver disease, coronary heart disease, irritable bowel syndrome, inflammatory bowel disorders, allergy, and asthma (de Vos & De Vos, 2012). Extensive research highlighted that those alterations of microbial composition affect the secondary metabolites produced by bacterial, impairing signalling in distal tissues, such as liver, brain, muscle and adipose tissues. As a whole, a normal microbiota participates in the activities that maintain a healthy gastrointestinal status and controls various central nervous systemic activities, while altered gut microbiota activates neuropsychiatric disorders, such as anxiety, depression, schizophrenia, and autism (Yarandi et al., 2016).

Genetical features of the host, intestinal microbiota, and the mucosal immune system participate in the pathogenesis of Crohn's disease (CD) and ulcerative colitis (UC), the two major forms of inflammatory bowel diseases (IBD) (Gorkiewicz, 2018). In ulcerative colitis, inflammatory state is continuous, uniform, and restricted to the superficial layers of the large intestine. On the other side, CD inflammation involves any region of the intestine in a discontinuous pattern and occurs in all of the layers of GI tract (Guan, 2019). Microbiome signatures of these diseases are the decreased abundance of the butyrate-producer *F. prausnizii,* and the beneficial commensal *A. muciniphila,* and the increase of Proteobacteria, associated to the loss of integrity of the mucosal barrier covering the epithelium surface of the gastrointestinal tract (Boltin et al., 2013; Kostic et al., 2014; Png et al., 2010; Matijaši et al., 2016).

Dysbiosis characterized by decreased diversity and abundance of bacteria and fungi is involved in cardiovascular, neuronal, immune, and metabolic disorders. Microbiota-host interactions

are affected by the availability of short chain fatty acids, the host inflammatory response, several signaling pathways, the integrity of the intestinal barrier, and the interaction of the gut-brain axis. In particular, gut microbiota perturbations induce obesity and type 2 dyabetes through the production of SCFAs and metabolites that directly influence glucose and insulin regulation, determining immune dysregulation, altered energy regulation, altered gut hormone regulation, and pro-inflammatory mechanisms (Agus et al., 2021; Cunningham et al., 2021).

1.2 The mucosal barrier

1.2.1 Mucus

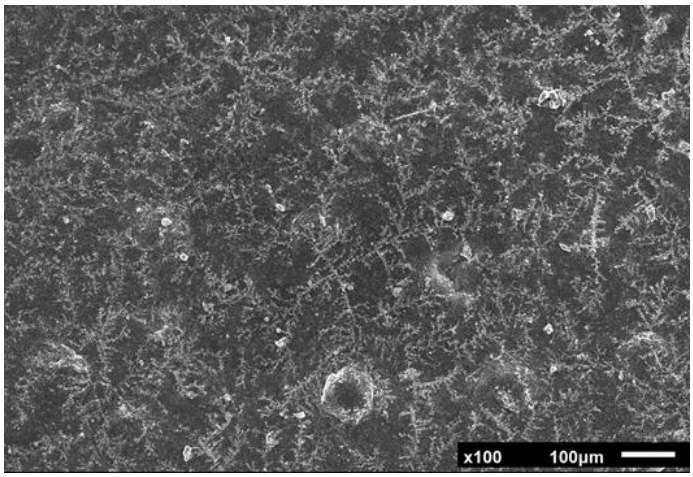
Mucus is a viscoelastic gel barrier that covers wet epithelial surfaces throughout the body including gastrointestinal, respiratory, reproductive, urinary tracts, and ocular surface (Alemao et al., 2020; Bansil & Turner, 2018). It is secreted by goblet and mucous cells and achieves a complete layer already several days after birth (Bunesova et al., 2018). In the gut, different goblet cell populations regulate the production of mucin in response to variable stimuli, playing distinct roles to maintain intestinal homeostasis. In healthy condition, in the colon, mucins are continuously secreted at a rate of ~2–4 μm/ min, creating a flow that physically repels microorganisms and turns over the mucus barrier every hour (Lee et al., 2021). Otherwise goblet cells can be stimulated to quickly release mucin in response to bioactive factors like neuropeptides, cytokines, and lipids (S. Yang, 2021). The immune system establishes an important regulation on goblet cell activity, indicating that these cells represent the gatekeepers of the mucosal immune system (Birchenough et al., 2015). The amount of goblet cells increases along the GI tract following the distribution of microbial organisms from duodenum to colon (Kim & Ho, 2010).

Gastrointestinal tract is the largest mucosal surface in human body and is continuously in contact with dietary antigens and different microorganisms (Turner, 2009). Almost 10 L of mucus is secreted into the GI tract, most of which is digested and recycled (Cone, 2009). The thickness of the inner mucus layer changes increasing along the intestine, being thicker in the colon than in the small intestine, correlating with increasing bacterial loads. Along the GI tract, mucus layer is progressively converted into a looser and expanded coating through the lytic action of proteases and glycosidases of both the host and the commensal bacteria. The outer mucus layer is the colonization site of the commensal microbiota (Johansson et al., 2011). Interaction between the host and gut bacteria

represents a balanced and symbiotic relationship which benefits both and influences mucus structure (Belzer & De Vos, 2012). Mucins select bacteria colonizing the mucus, providing carbon and nitrogen substrates and exposing *O*-glycan chains that serve as attachment sites for colonization (Johansson et al., 2011; Tailford et al., 2015). On the other hand, commensal bacteria using oligosaccharides for their metabolism limit the penetration of harmful microorganisms in the outer layer of the mucus (Linden et al., 2008). Bacteria tightly interact with the host and modulate mucin gene expression, glycosylation, and secretion, thus affecting mucus and epithelial homeostasis (Cornick et al., 2015; Wagner et al., 2018).

Different roles and functions are carried out by mucus barrier such as gaseous exchange, nutrients and cofactor adsorption, lubrification, chemical sensing and an important relationship with the immune system. To maintain the defence character of the barrier, it is crucial a continuous turnover, with a dynamic balance between biosynthesis, secretion, and degradation of the structural components (Anthony P. Corfield, 2015).

Mucus is composed by water (90-95%), mineral salts, lipids, and several proteins including antimicrobial immune factors. Electrolyte's composition changes in different mucous tissues but generally consists of sodium and potassium chloride, phosphate, magnesium, calcium, and bicarbonate. This last electrolyte is very important due to its capability to buffer the inner layer, close to the epithelium, to pH 7. This is crucial in the stomach where bicarbonate exerts a protective effect against HCl acid secretion (Allen & Flemström, 2005). Lipids are approximately 1-2% of mucus. They interact with mucus proteins and influence wettability, hydrophobicity, and barrier functions of the mucus layer. The major building blocks of mucus are mucin glycoproteins (Bansil & Turner, 2018).



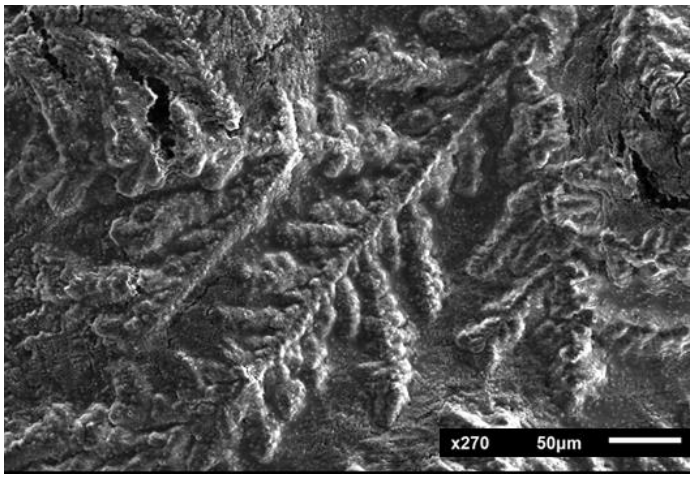


Figure 5 SEM images of mucin suspension at different magnifications, showing the "bottle brush" like conformation.

1.2.2 Mucins

The main structural component of mucus are mucins, high weight glycoprotein (Godl et al., 2002). To date, 21 genes encoding mucins have been identified in human genome, classified according to their monomeric peptide domain (S. Yang, 2021). Mucins can be secreted into the lumen or expressed on the cell surface. Secretory mucins are divided in two subgroups, non-gel forming and gel-forming mucins, the latter being the main component of mucus gel and providing hydration, transport, protection and lubrication (Theodoropoulos & Carraway, 2007). Synthesis and secretion of mucins take place in goblet cells of the small intestine and colon, or on the surface mucous cells of the stomach (Moncada et al., 2003). Membrane-bound mucins act like membrane anchored glycoproteins, harbouring a monomeric structure that contains different transmembrane domains involved in signal transduction (Bankole et al., 2021). The expression profile of mucins changes in the mucosal barrier and is tissue-dependent, with the GI tract being the body district characterized by the highest and most different level of mucin expression (Linden et al., 2008). Typically, mucins of the human gut include gel-forming mucins (MUC2, MUC5AC, MUC5B, and MUC6), non-gel forming mucin (MUC7), and membrane-associated mucins (MUC1, MUC3A/B, MUC4, MUC12, MUC13, MUC15, MUC17, MUC20, and MUC21) (Anthony P. Corfield, 2018).

MUC2 was the first mucin identified and characterized, and represents the major component of the mucus layer in the small intestine and colon, where weakly expression of MUC5B is also present. Gastric mucus in the stomach is mainly composed of MUC5AC and MUC6, expressed in

mucous cell neck at the surface of gastric epithelium and in the antrum, respectively (Ringot-destrez et al., 2018). All types of mucins share the high level of glycosylation. Oligosaccharides such as *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), fucose, galactose, *N*-acetylneuraminic acid, and smaller amounts of mannose, constitute up to 80% of the final mucin mass (Bansil & Turner, 2006).

The first step of mucin glycosylation takes place in the Golgi apparatus and is catalyzed by GalNAc-transferase (GalNAc-Ts) that links GalNAc residue by *O*-glycosidic bond to the hydroxyl side chains of threonine (Thr) or serine (Ser), forming GalNAcα1-O-S/T, also called Tn antigen (Jensen et al., 2010). The initial step of glycosylation is crucial, according to the especially high number of enzymes monitoring the process compared to other protein glycosylation, and to the abundance of different GalNAc-Ts, each one having preference towards Ser or Thr acceptors (Bennett et al., 2012; Bergstrom & Xia, 2013; Daniel et al., 2020). These two amino-acids, together with proline, account for most of the protein core composed by PTS-rich domains (Proline-Threonine-Serine domains), often repeated in tandem. While Thr and Ser are widely *O*-glycosylated and confer a "bottle brush" like conformation (Fig 5), proline ensures that the mucin structure remains unfolded (Paone & Cani, 2020). Mucin N- and C- terminal regions present reduced *O*-glycosylation and few *N*-glycosylation sites. Furthermore, these regions are characterized by rich cysteine portions containing von Willebrand factor (vWF) C and D domains, and C-terminal cysteine-knot domains involved in dimerization via disulphide bond with subsequent polymerization of the dimers to form multimers (Bansil & Turner, 2006).

The extension of branched oligosaccharides starts on position C3 or C6 of GalNAc, leading to the formation of eight different mucin core structures (Fig. 6) (Corfield, 2018). The attachment of galactose or GlcNAc residue to the Tn antigen forms core 1 and core 3, respectively. The further extension of these cores by the addition of GlcNAc at the C6 hydroxyl group of GalNAc shapes the core 2 and 4 structures (Zhang et al., 2021). Core 1 to 4 are the main core structures commonly found in human intestinal mucins, with core 3 widespread along the colon and representing the major *O*-glycan of MUC2, with more than 100 different *O*-linked glycans (Holmén-Larsson et al., 2009; Robbe et al., 2004). Core structures are further elongated by up to 20 residues of galactose, GalNAc, or GlcNAc leading to linear or branched chains that end with fucose, sialic acid, sulphate groups, or present ABO and Lewis histo-blood group antigens (Holmén Larsson et al., 2013). These terminal epitopes are the major responsible for the variety between oligosaccharides chain, presenting a different distribution among regions of human gastrointestinal tract (Tailford et al., 2015). Using NMR and mass spectrometry, Robbe and colleagues (2003) have demonstrated the presence of a

gradient from ileum to colon with a decreasing expression of fucose and ABO blood group, and an increasing ratio of sialic acid.

The mucin glycosylation patterns differs between GI tract regions and among individual, and depends on several factor such as the expression of glycosyltransferases, the health status, and the microbiota composition (Hansson, 2020). However, this glycan variability is almost absent in human colon, where a uniform glycosylation profile has been shown. As a whole, healthy individuals present a qualitatively and quantitatively uniform MUC2 glycosylation profile (Johansson et al., 2011).

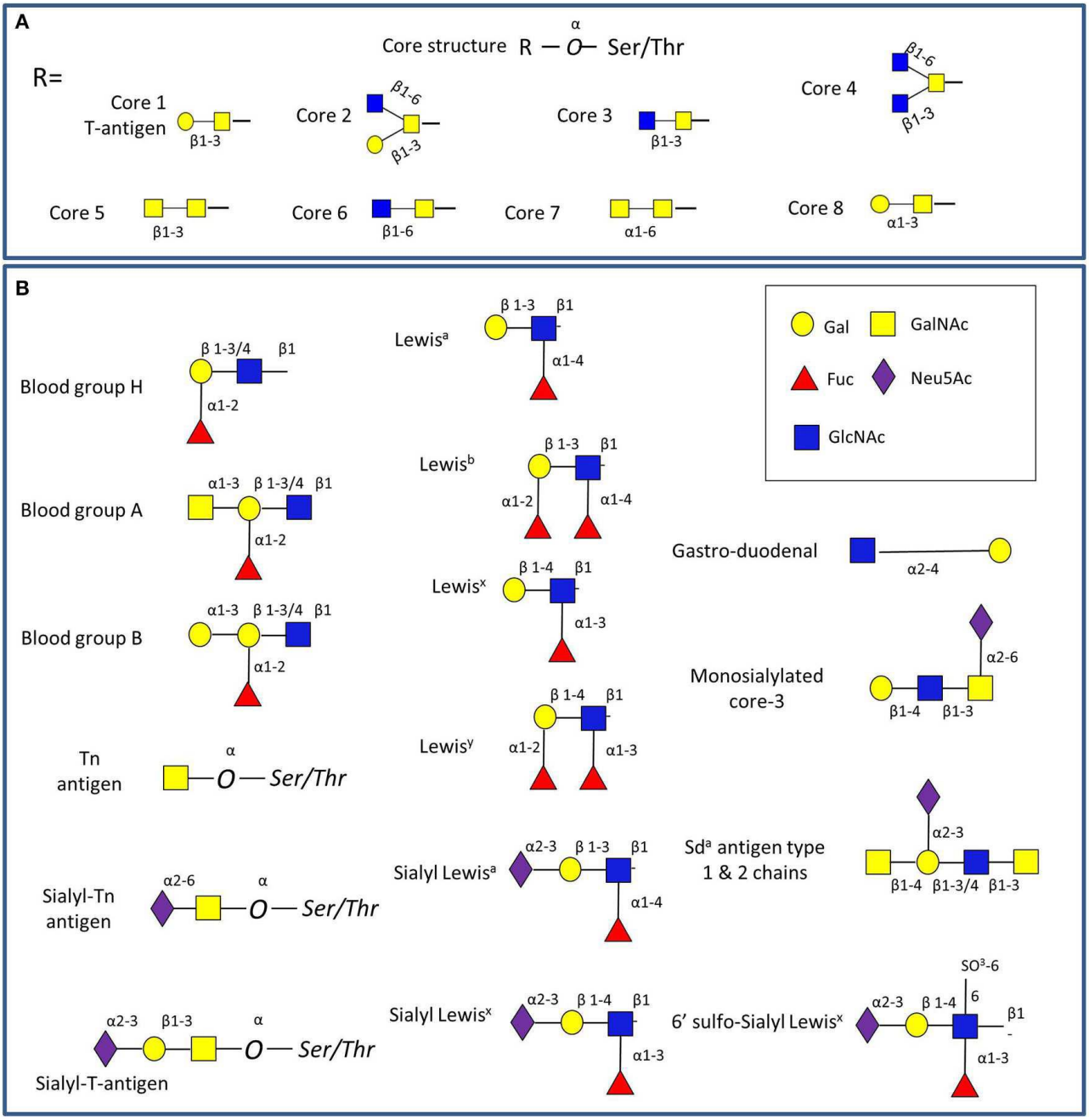


Figure 6: Representation of GI mucin glycan structure. (A) Core 1-8 found in GI tract. (B) Main glycan epitopes in GI mucins. (Tailford et al., 2015).

1.2.3 Enzymes involved in mucus degradation

Although microbial populations are widespread all-over human body, the colon hosts the most concentrated microbiota. It is estimated that bacterial community of the colon is larger than any other body districts by two orders of magnitude (Sender et al., 2016). Fluorescence in situ hybridization (FISH) analysis has shown that the outer mucus layer of the colon is the colonization site of resident commensal microbiota, which differs from that of the digesta-associated and fecal content in terms of relative abundance of the different taxa (Donaldson et al., 2015).

The ability of human microbiota to degrade mucus in the intestinal mucosa depends on the expression of proteases, sulfatases, peptidase, and of a number of glycosidases that catalyse hydrolytic cleavage of the oligosaccharides chain, releasing monosaccharides which can be assimilated as energy source by the intestinal microorganisms (Corfield et al., 1992). *In silico* analysis showed that 86% of human gut bacteria encode genes for the cleavage of mucin glycans, and 89% encode genes for the metabolism of the mucin chain monosaccharides (Ravcheev & Thiele, 2017). Different CAZyme classes are expressed by gut bacteria, and are involved in mucin degradation. The first class of enzymes involved in mucin degradation encompasses carbohydrate esterases (CEs), which remove ester substituents from mucin chains, enabling the access to other CAZymes (Berkhout et al., 2021). Then the degradation of mucin glycans occurs thanks to the sequential action of different glycosyl hydrolases (GHs) (Lombard et al., 2014). Every GH acts only on a specific *O*-glycan bond, underlying the importance of a step-by-step hydrolysis. As a consequence, the absence of certain GH activities leads to an incomplete catabolism pattern that may influence mucosal surface interactions and function (Ravcheev & Thiele, 2017). In addition, modification such as sulfation and acetylation not only affect chemical properties and charge state of the mucus, but also require enzymatic removal prior to the action of GHs (Marcobal et al., 2013). Most of the GHs are located on the cell surface or secreted into the environment, and act in the extracellular compartment due to the large size and degree of polymerisation of the mucin structure (Berkhout et al., 2021).

The first step of mucin degradation is the release of peripheral residues such as sialic acid and fucose by exo-acting GHs. Sialic acids, derivatives of neuraminic acid, are a class of 9-carbon amino sugars, among which *N*-acetylneuraminic acid (Neu5Ac) is the most common in humans. Neu5Ac is linked to Gal by α(2-3) or α(2-6)-linkages, and to GalNAc by α(2-6)-linkage (Kim et al., 2011). The removal of sialic acid from non-reducing ends of glycoproteins, glycolipids, and polysaccharides is carried out by the family GH33 of sialidases (also known as neuraminidases) (Ravcheev & Thiele, 2017). The GH33 family encompasses hydrolytic α-sialidases, trans sialidases, and intramolecular trans-sialidases (IT-sialidases, anhydrosialidases) (Berkhout et al., 2021). In bacteria, the genes

involved in the release of sialic acids, such as *nanA*, *nanK* and *nanE* generally cluster together in the so-called Nan cluster. The *nanA/K/E* cluster was firstly described in *E. coli*, and to date it has been reported only in gut commensals that use it to gain energy, nitrogen, and carbon source, and to obtain amino sugars for cell wall synthesis. The Nan cluster has been identified in *Anaerotruncus colihominis, Dorea formicigenerans, D. longicatena, F. prausnitzii, Fusobacterium nucleatum, R. gnavus, L. sakei, L. plantarum, L. salivarius,* and *B. breve*. On the other hand, also pathogens such as *Vibrio cholerae, Clostridium perfringens, C. botulinum*, enterohemorrhagic *E. coli*, *Shigella*, and *Salmonella enterica* encode the cluster and use sialidases to disclose adhesin or toxin-binding sites (Almagro-moreno & Boyd, 2009; Bell et al., 2020; Juge et al., 2016).

The desoxyhexose L-fucose is connected to galactose residues by α(1,2)-linkage and to GlcNac by α(1,2), α(1,3), or α(1,4)-linkage. The removal of fucose involves *α*-L-fucosidases, that are classified into GH29 and GH95 families. Recently, two new families of fucosidases were added: GH139 and GH141, other than *β*-fucosidases belonging to the GH1 and GH30 families (Grootaert et al., 2020).

Fucosidases characterized until now belong to *Bifidobacterium bifidum, B. longum, A. muciniphila, Bacteroides thetaiotamicron, C. perfringens*, and *Ruminococcus gnavus* (H. Wu et al., 2021). For instance, *B. bifidum* releases fucose, that enables the growth of *E. halii*, an early commensal species that produce butyrate and propionate from fermentation, which is not able to degrade complex oligosaccharides (Bunesova et al., 2018). Further removal of terminal sugars involves α- and β-galactosidases, that belong to the families GH4, GH27, GH36, GH43, and to the families GH2, GH20, GH35, GH42, respectively, and endo-acting β-galactosidases (GH98).

Galactose utilization occurs through two different pathways. In Leloir pathway galactose is utilized through transformation in UDP-galactose. In another pathway occurring only in Actinobacteria and Firmicutes, galactose is converted into tagatose-6-phosphate (Ravcheev & Thiele, 2017).

The α-*N*-acetylglucosaminidases promote the hydrolysis of α(1,4)-linkages between GlcNAc and galactose (GH89 family), whereas β-*N*-hexosaminidases hydrolyse β(1,3)- or β(1,6)-linkages of GlcNAc with Gal o GalNAc (GH84, GH85, GH3, and GH20 families), resulting in the release of *N*-acetylglucosamine (Bell et al., 2020; Fujita et al., 2011).

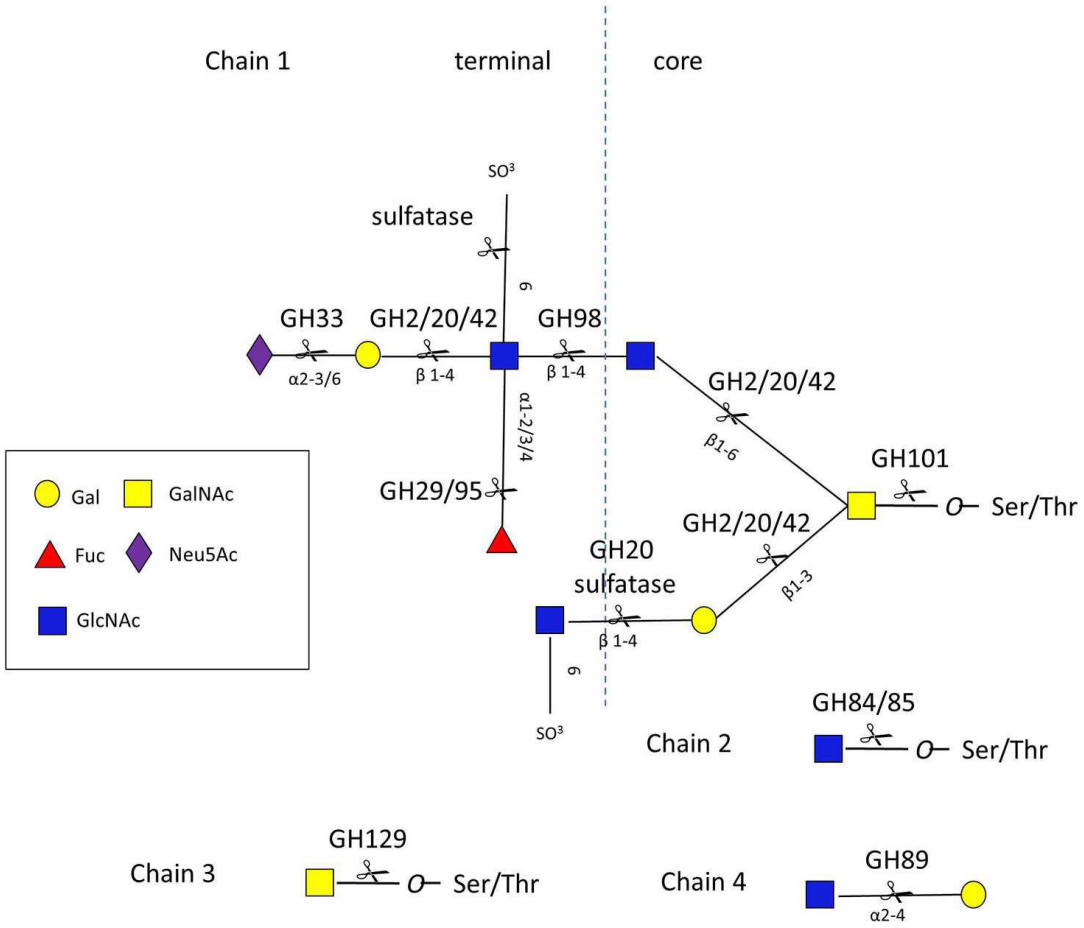


Figure 7 Schematic representation of GHs involved in mucin breakdown. Sialidases (GH33), galactosidases (GH2/20/42), fucosidases (GH29/95), *N*-acetylglucosaminidases (GH84/85/89) and endo-α-*N*-acetylgalactosaminidases (GH101/129). Chain 1 is a hypothetical mucin glycan chain, chain 2 is O-GlcNAc often found on other glycoproteins, chain3 (Tn antigen) and chain 4 are found in gastro-duodenal mucin (Tailford et al., 2015).

The removal of *N*-acetylgalactosamine (GalNAc) residues from mucin chain is mediated by the action of exo-α-*N*-acetylgalactosaminidases (GH27 family), and the above mentioned β-*N*-hexosaminidases (Ravcheev & Thiele, 2017). When the terminal sugars and blood group antigens are removed, the mucin core glycan are exposed to the action of endo-α-*N*-acetylgalactosaminidases (GH101, and GH129) that cleave the linkage with the protein backbone (Ashida et al., 2008). The GH101 family includes enzymes that release the core-1 *O*-linked glycans, while GH129 family has as substrate the GalNAc-α1-Ser/Thr Tn antigen present in mucin glycoproteins (Kiyohara et al., 2012; Koutsioulis et al., 2008).

The monosaccharides that are released can be used by commensal microbiota as energy and carbon source, other than complex carbohydrates deriving from diet, being fermented into short-chain fatty acids (SCFAs) that positively influence host' health (Hansson, 2012). Gases such as hydrogen

(H_2), methane (CH_4), hydrogen sulfide (H_2S), and carbon dioxide (CO_2) are also produced from the fermentation of mucin derived mono- and oligosaccharides (Kaoutari et al., 2013).

Several studies have been shown that in mucus layer bacteria belonging to the phylum Firmicutes are generally more abundant than Bacteroidetes, both in humans and rodents. Moreover, based on 16S rRNA gene sequences, mucosa-associated bacteria are mainly Lachnospiraceae, Ruminococcaceae, *Bifidobacterium bifidum*, *Bifidobacterium longum* and bacteria belonging to the phylum Verrucomicrobia such as *Akkermansia muciniphila*. In particular, FISH-studies detected in human mucus Lachnospiraceae, Enterobacteriaceae, *Eubacterium rectale*, *F. prausnitzii*, *Eubacterium cylindroides*, *Clostridium histolyticum*, *Clostridium lituseburense* and *A. muciniphila* (Ouwerkerk et al., 2013). A recent work that combines FISH, laser capture microdissection (LCM), and 16s rRNA gene sequencing of spatially distinct regions of the mouse gut lumen, found out that intestinal bacteria form a continuous biofilm-like structure coating the mucosa surface of the colon. This biofilm structure is rich of anaerobic bacteria such as bacteroides and clostridia, the latter having a role in maintaining the correct balance of the community and the integrity structure, and preventing the bloom of Proteobacteria (Duncan et al., 2021).

To date mucus degrading bacteria that have been identified belong to *Bacteroides* spp. that possess a high number of polysaccharide-utilization loci (PULs), *Ruminococcus* spp., *Bifidobacterium spp.*, *A. muciniphila*, and *Allobaculum mucolyticum* (Derrien et al., 2004; Martens et al., 2009; Png et al., 2010; van Muijlwijk et al., 2021). *A. muciniphila* has been isolated from human intestine using only mucins as carbon and nitrogen source by Derrien and colleagues (2004), it has a small genome encoding almost 60 proteins involved in mucus degradation (Ottman et al., 2017; Passel et al., 2011).

Mucin glycan degradation requires the cooperation among mucosa-associated bacteria expressing different type of GHs. Given the extracellular location of most of the enzymes, the products derived from bacterial enzymatic hydrolysis are released into the environment and could be available to other microorganisms (Berkhout et al., 2021). For example, *B. bifidum* possess two extracellular sialidases that release sialic acid that can be consumed by *B. breve* (Nishiyama et al., 2018), or *Desulfovibrio piger* and *Desulfovibrio desulfuricans* that benefit from the release of sulphate by the mucin degraders *B. thetaiotamicron* and *B. fragilis*, respectively (Rey et al., 2013).

The ability of mucin-degraders to use different glycan moieties derived from GI mucus may has a key role in early colonization, providing some bacteria with endogenous nutrient sources before the introduction of dietary glycans (Koropatkin et al., 2012). Structural similarity between mucin glycans and HMOs (Human Milk Oligosaccharides) highlights the role of these bacteria in establishing and maintaining microbial community in the gut (Yu et al., 2013).

Over the life of the host, mucosa-associated bacteria, because of the proximity to host tissues, play an important role in human health and disease (Tailford et al., 2015).

1.2.4 Compromised mucus, dysbiosis, and diseases

The intestinal barrier is composed of multiple layers. The outer one is formed by the association of the mucus layer, the commensal gut microbiota, and defence proteins such as antimicrobial proteins and secretory immunoglobulin A. Mucus prevents antigens, toxins, and bacteria from directly contacting the underlying epithelium. Intestinal epithelial cells (IECs) form the middle layer of the barrier, while the inner one is composed of innate and adaptive immunity cells (Vancamelbeke & Vermeire, 2017). This barrier is a dynamic entity interacting with and responding to various stimuli.

The role of the mucus layer is to protect intestinal cells from external agents and to facilitate nutrient absorption. On the other side, the intestinal microbiota regulates the gut environment and influence the integrity and function of the outer mucus layer. For instance, bacterial products such as lipopolysaccharide (LPS) and peptidoglycan stimulate mucus secretion and restore mucus properties (Petersson et al., 2011). Albeit mucous degradation is fundamental for turnover of the layers, in some cases, the perturbation of the equilibrium between degradation and synthesis of mucins causes a shift in microbiota composition, or *vice versa*. In this situation, mucins' degradation may be related with the primary stage of mucosal diseases, like in Inflammatory Bowel Diseases (IBDs) (Bankole et al., 2021).

IBDs, such as ulcerative colitis (UC) and Chron's disease (CD), are chronic inflammatory conditions of GI tract that affect almost 6.8 million people worldwide. Although the pathogenesis of IBDs remains unclear, it is evident that several factors contribute to the onset of these diseases, including host genes, environmental factors, immune dysregulation, and alteration of gut microbiota composition (Kudelka et al., 2020). Intestinal leakiness in patients with IBDs is related to dysbiosis, inflammatory response, and tight junctions' modifications. Compared to those of healthy subjects, microbiota of IBDs patients show enriched population of Enterobacteriaceae (Chen et al., 2014; Seksik et al., 2003), *Fusobacterium* and *Enterococcus faecalis* (Zhou et al., 2016), and reduced bifidobacteria and butyrate-producing bacteria such as *Faecalibacterium, Eubacterium, Roseburia,* Lachnospiraceae and Ruminococcaceae (Kostic et al., 2014). Furthermore, population of mucolytic bacteria are different in CD and UC, when compared to healthy subjects, with an increased abundance

of species like *R. gnavus* and *R. torques*, and a 10-fold reduction in CD patients and 100-fold in UC patients of *A. muciniphila*, which exerts protective and anti-inflammatory role (Png et al., 2010).

Even in absence of acute inflammation, IBD patients present an increase of total mucosa-associated bacteria (Boltin et al., 2013). In IBD, acute intestinal inflammation leads to a reduction of goblet cells that results in lower mucin synthesis, and thinner mucus layer. Loss of integrity of mucosal barrier increases the interactions between epithelium and bacteria, which promotes further inflammation (Matijaši et al., 2016). Moreover, in active UC inflammation subjects occur an alteration of glycan profile, which presents a subset of shorter glycans (Larsson et al., 2011). As a whole, several studies demonstrated that the alteration of mucin glycosylation is linked to increased inflammation in humans, underling the importance of glycan profile in maintenance of intestinal homeostasis (Antonini et al., 2019).

Enteric pathogens have developed different strategies to evade the mucosal barrier and start the infection. These include mechanism to allow efficient penetration of the mucus, through enzyme activities that degrade mucus layer and/or by the use of carbohydrates released from other commensal bacteria (Mcguckin et al., 2011). Accordingly, *C. difficile* does not harbours GHs but uses carbohydrate moieties released from mucin to colonize the host (Engevik et al., 2020). Mucin oligosaccharides act as attachment site, establishing an important colonization factor for both commensal and pathogenic bacteria. For example, *H. pylori* possesses adhesin which bind to glycan of MUC5AC in the stomach. *Campylobacter jejuni* binds sialylated glycan in MUC2 and stimulates the expression of virulence factors. The human pathogens *Streptococcus gordonii*, *S. sanguinis* and *S. mitis* adhere to sialic acids via serine rich region proteins (SRR proteins) or Hss adhesin (Bovenkamp et al., 2003; Engevik et al., 2020; Mcguckin et al., 2011; Ringot-destrez et al., 2018.; Tu et al., 2010). Terminal residues of sialic acids represent one of the most frequently targeted carbohydrate receptors for human pathogen adhesion, with a high number of sialic acid binding agglutinins, adhesins, and lectins. Among the pathogens that exploits this strategy, there are also *V. cholerae*, *H. pylori*, and *E. coli* (Lewis & Lewis, 2012).

1.2.5 The mucosal-microbiome-diet axis

Changes in host diet rapidly affect the mucosal barrier through alteration of microbial population (Beaumont et al., 2017). Many immune disorders including IBDs, diabetes, rheumatoid arthritis, psoriasis are associated with alteration of microbiota and are related to the diet (Alemao et al., 2020). Depending on nutrient availability, gut bacteria are able to switch from using dietary fiber to mucin glycoproteins as their main energy source, creating a link between host diet, mucin degradation, and efficiency of mucosal barrier (Mcguckin et al., 2011).

Dietary compounds are divided in macronutrients (lipids, proteins, and carbohydrates) and micronutrients (vitamins and minerals). A central role in mucus homeostasis is carried out by fiber, that enriches the gut providing a fermentable source for microbial growth (Fernández-Tomé et al., 2021). Fiber are usually defined as edible carbohydrate polymers, derived from plants and fungi, that cannot be digested by human enzymes, and include inulin, dextrin, pectin, cellulose, arabinoxylans, resistant starch, and chitin (Coker et al., 2021). Fibers reach undigested the colon, where are fermented by gut bacteria to produce SCFAs, mainly acetate, propionate, and butyrate. These products are accessible for both the host and commensal bacteria, the latter taking advantage of cross-feeding relationship. SCFAs participate in maintaining the mucosal integrity, improve glucose and lipid metabolism, control energy expenditure, and regulate inflammatory responses (Agus et al., 2021). Using fermenting dietary fiber, microorganisms can release other metabolites, including tryptophan metabolites and secondary bile acids, which can modulate intestinal immunity. In absence or deficiency of dietary fiber, commensal bacteria degrade mucus barrier to gain energy, exposing the underlying intestinal epithelium to luminal bacteria (Beukema et al., 2020; Desai et al., 2016).

Low intake of fiber is one of the features of a typical “Western-style” diet, together with the high assumption of sugars and fats, causes a higher permeability of colonic mucus, reduces the amount of commensal bacteria, including health promoting ones, such as *Bifidobacterium spp*. and *A. muciniphila,* and increases abundance of Proteobacteria, a phylum associated with IBD. Conversely, administration of accessible fiber favours a higher abundance of *Faecalibacterium* other than *Bifidobacterium spp.,* and increases serum SCFAs concentrations (H. Shi et al., 2020). Vitamins and minerals are substrates and cofactors in several microbial pathways, in this context it is easy to understand that also micronutrients deficiencies affect human gut microbiota composition. Moreover, additives of processed food, i.e. dietary emulsifiers, antimicrobial additives, and artificial sweeteners, improve intestinal permeability and inflammation through an increase of mucin degraders bacteria and endotoxins (Bolte et al., 2021).

1.2.6 The intestinal mucosal immune system

A close relationship links luminal bacteria, mucus, and immune system. The intestinal mucosal barrier is composed by epithelial cells, mucus layer, and immune cells (Perez-Lopez et al., 2016). Numerous immune cells are widespread in the gut, they can be found between epithelial cells and deep in the lamina propria, or can be organized in structure such as the lymphoid tissue known as Payer's patches and the mesenteric lymph nodes (Rescigno et al., 2008).

Most of the interactions between microbiota and host occur on the mucosa surface. Here, gut commensal bacteria exert several effects on different cell populations of the mucosal immune system, which in turn, induces the release of cytokines (Flaherty et al., 2010). To maintain homeostasis and discriminate between commensal and pathogen bacteria, immune cells activate two pattern recognition receptor (PRR) systems, the Toll-like receptors (TLRs) located on the surface of immune cells and the cytoplasmatic NOD-like receptors (NLRs). Different stimuli activate in different way these two systems, for example lipoteichoic acids and peptidoglycan, components of Gram-positive cell wall, activate TLR2, while lipopolysaccharide and flagellin activate TLR4 and TLR5, respectively (Flaherty et al., 2010; O'Hara & Shanahan, 2006).

The first line of defence of the mucosal immune system is represented by the epithelium. The epithelial cells, especially enterocytes, together with Paneth cells, release antimicrobial peptides (AMPs) in the mucus, with a decreasing gradient of concentration from the surface that keeps bacteria away from the epithelium (Birchenough et al., 2015). Within AMPs, defensins represent one of the main classes. These are divided in α-defensins produced by Paneth cells, and β-defensins ubiquitously expressed throughout the gastrointestinal tract (Sansonetti, 2004). Small intestinal goblet cells transfer these AMPs from intestinal lumen to the dendritic cells (DCs) (McDole et al., 2012), playing a key role in the intestinal immune system. DCs stimulated by the expression of CX_3C-chemokine receptor 1 (CX_3CR1) can sample the luminal environment extending dendrites between enterocytes without disrupting tight junctions, allowing transport of bacteria into mesenteric lymph nodes. Here bacteria induce B cells to produce IgA, other than the differentiation of T-cells (Perez-Lopez et al., 2016).

Different population of DCs reside in the lamina propria. In particular, $CD11c^+$ and $CD103^+$ subpopulations receive luminal antigens and interact with regulatory T-cells, inducing tolerance mechanism. To maintain the protective function of mucosal barrier against harmful pathogens, remaining tolerant to commensal bacteria and dietary antigens, mucosal $CD103^+$ DCs cells regulate a pool of naïve T-cells, and induce the expression of $CD4^+$ $CD25^+$ $FoxP3^+$ regulatory T-cell (Coombes et al., 2007). Regulatory T-cells are mostly located in the lamina propria where they act as gatekeepers

of immune homeostasis. These cells differentiate in response to the release of cytokines by the innate immune cells (Littman & Rudensky, 2010).

At the same time, the immune system directly regulates the production of gel-forming mucins by goblet cells, releasing different cytokines. The release of anti-inflammatory IL-22 and IL-10 by regulatory T cells has positive effect on goblet cell function, prevents protein misfolding, and maintains mucus production in order to preserve the integrity of the mucus barrier. In particular, IL-22 is produced by $CD4^{+}$T cells, which differentiation depends on the composition of mucosa-associated bacteria and by immune cells that amplify the T-helper cells 17 (Th17) response. Conversely, proinflammatory cytokines IL-1β, IL-4, IL-6, IL-9, and IL-13 are associated with goblet cells hyperplasia and mucus hypersecretion, that lead to barrier disruption (Hasnain et al., 2013; Linden et al., 2008). IL-1β cause increasing intestinal tight junction permeability, which increase intestinal bacteria penetration from the lumen to lamina propria, resulting in colitis in mouse models (Wu et al., 2022).

Increased evidences in the last few years demonstrate that a dysregulated immune response is a key cause of human inflammatory disorders (such as IBD), like in the case of aberrant signalling of regulatory cytokines IL-10 and IL-22 (Fernández-Tomé et al., 2021; Kidess et al., 2021). As a whole, considerable part of the immune system resides in the GI tract, lies in the subepithelial compartment, and is ready to react when mucosal layers and epithelial cells lining fail (Pelaseyed et al., 2014).

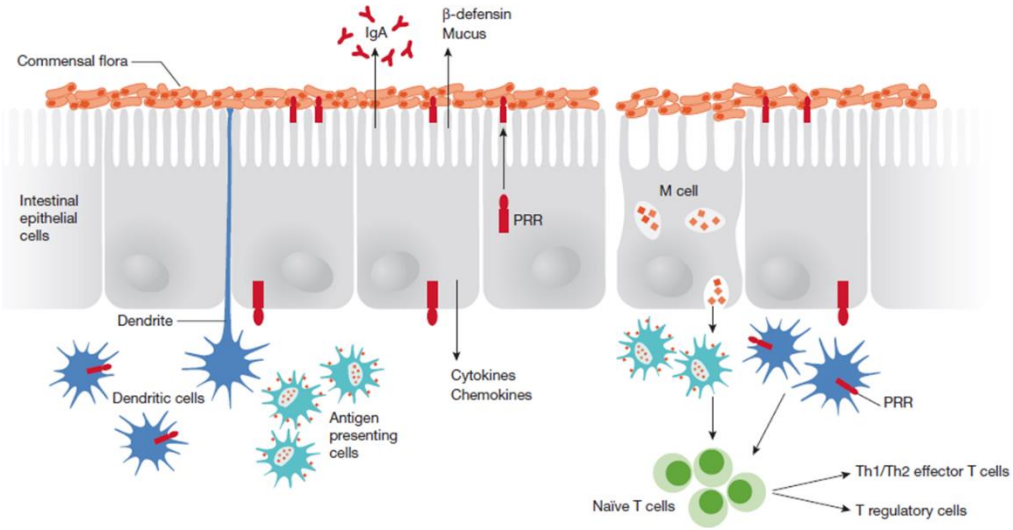


Figure 8 Association between mucus, gut microbiota, and immune system. Enterocytes secrete many immune factors (like AMP, IgA, and chemokines). These compounds are transported to other cells, which are then presented to naïve T cells. DCs cell reside in the lamina propria, sample the mucosal environment and modulate T regulatory cells. (O'Hara & Shanahan, 2006)

1.3 Microbial profiling

The complex ecosystem that populates human gastrointestinal tract is known to be crucial for host's health and functions. However, lacking of data on microbial diversity makes difficult to understand the role of single species (Almeida et al., 2021). Although the culturing studies have been improved, several microorganisms remain difficult to isolate under laboratory conditions and still lack of a sequenced genome even if they are prevalent in humans (Fodor et al., 2012). In this regard, it was estimated that about 40-50% of gut bacteria lack of a reference genome (Nayfach et al., 2019). Currently, culture-dependent approach allowed the identification of over 1500 microbial species but according to the recent work of the Unified Human Gastrointestinal Genome (UHGG) there are still more than 70% of gut bacterial species that remains uncultured (Almeida et al., 2021; Liu et al., 2021). In this prospective, culture- and single cell- based technologies are helping to fill the gap of these hard-to-cultured microbes (Fodor et al., 2012). In addition, metagenomic approaches are used to obtain genomes without requiring isolation or culturing process. They are based on the assembling of sequence reads into contigs, which are then binned into metagenome assembled genomes (MAGs) on the basis of nucleotide frequency, abundance and/or co-variation of abundance across a group of samples (Nayfach et al., 2019). In recent years, the advent of high-throughput technologies and new bioinformatic tools improved knowledge about human microbiota and are disclosing thus more and more complex ecosystem (Anantharaman et al., 2016; Parks et al., 2017).

In this context, the Human Microbiome Project (HMP) represents the first attempt to enrich knowledge of human-associated microbiota, and to discover linkages within the microbial communities of 18 different human body districts of over 200 healthy volunteers (Peterson et al., 2009). Thanks to HMP projects and other studies, nearly 5.000 bacterial species of the human body have been cultured and submitted to whole genome sequencing (Fodor et al., 2012).

1.3.1 Genomic approaches to study microbial diversity

The advent of high-throughput technologies allowed to overcome the limits of the culture-dependent approaches, starting a new way to study microbial diversity and communities (Bilal et al., 2018). Two different ways are possible to achieve this purpose, the first consisting in the sequencing of a single marker gene, whereas the whole-metagenomes shotgun (WGS) is based on the untargeted sequencing of all microbial genomes into a sample (Quince et al., 2017).

The 16S rRNA sequencing represents the current standard technique for profiling microbial communities, given its low cost and time efficiency. This methodology is based on the amplification of single or multiple hypervariable regions of the 16s rRNA gene using universal primers. Bacterial 16S rRNA has nine hypervariable regions (V1 to V9) each flanked by highly conserved DNA sequences. The 16S rRNA gene profiling defines a bacterial species as an organism sharing 97% of 16S rRNA sequence identity (Milani et al., 2017).

The introduction of NGS improved the first technology based on Sanger sequencing in terms of cost and depth of the analysis. The Illumina sequencing platform is the most used, thanks to its lower cost and deeper level of analysis. It produces reads of 100-150 bases, sequencing only one hypervariable region of the 16s rRNA gene (Weinstock, 2012). Reads obtained from the amplification process are processed using a number of bioinformatic pipelines.

The 16s rRNA gene profiling has need of three main steps (Fig 9), the first is the OTUs generation followed by their richness and abundance estimation. Among the diverse strategies to generate an OTU table, the algorithm DADA2 allows the identification of Amplicon Sequence Variant (ASVs) sharing 100% identity, that are ascribed to a species or to a higher taxonomic level, and that are aggregated to generate an abundance table (Rosen et al., 2012). The generated OTUs are used to define taxonomic profile of the microbial communities examined, quantifying richness (number of unique taxa), diversity and abundance. Among the diversity indices used in metagenomics analysis there are the Shannon index and Simpson's diversity index. The former measures the entropy and uncertainty of the sampling outcome, whereas the latter measures the probability that two reads will produce the same taxon (Li et al., 2012).

Whole-metagenomes shotgun (WGS) allows the analysis of whole bacterial genomes, providing more informations in term of functional aspects, diversity levels, and identified species (Li et al., 2021). This method is based on the production of several short sequences known as reads that have to be overlapped in order to obtain a *de novo* genome assembly. The shorts single reads (DNA fragments) are overlapped generating a continuous sequences known as contigs, while the paired reads derived from the end of DNA fragments that are too long to be sequenced, assemble contigs

into scaffolds. The quality of new genome assemblies is evaluated on the basis of number and length of scaffold and contigs, their length relative to the genome size, and the number of assembled reads (Baker, 2012).

Contigs are binned into metagenome assembled genomes (MAGs) on the basis of nucleotide frequency, abundance and/or co-variation of abundance across a group of samples. The taxonomy attribution of each MAG is achieved mapping is contig to a reference genomes database. Currently, several tools allow the analysis of metagenomic sequences without requiring the assembly step. For example, the tool MetaPhLan elaborate taxonomic attribution mapping the raw reads against its database of strain-specific marker genes generated through the analysis of publicly available genomes (Nayfach et al., 2019; Pasolli et al., 2019; Truong et al., 2015).

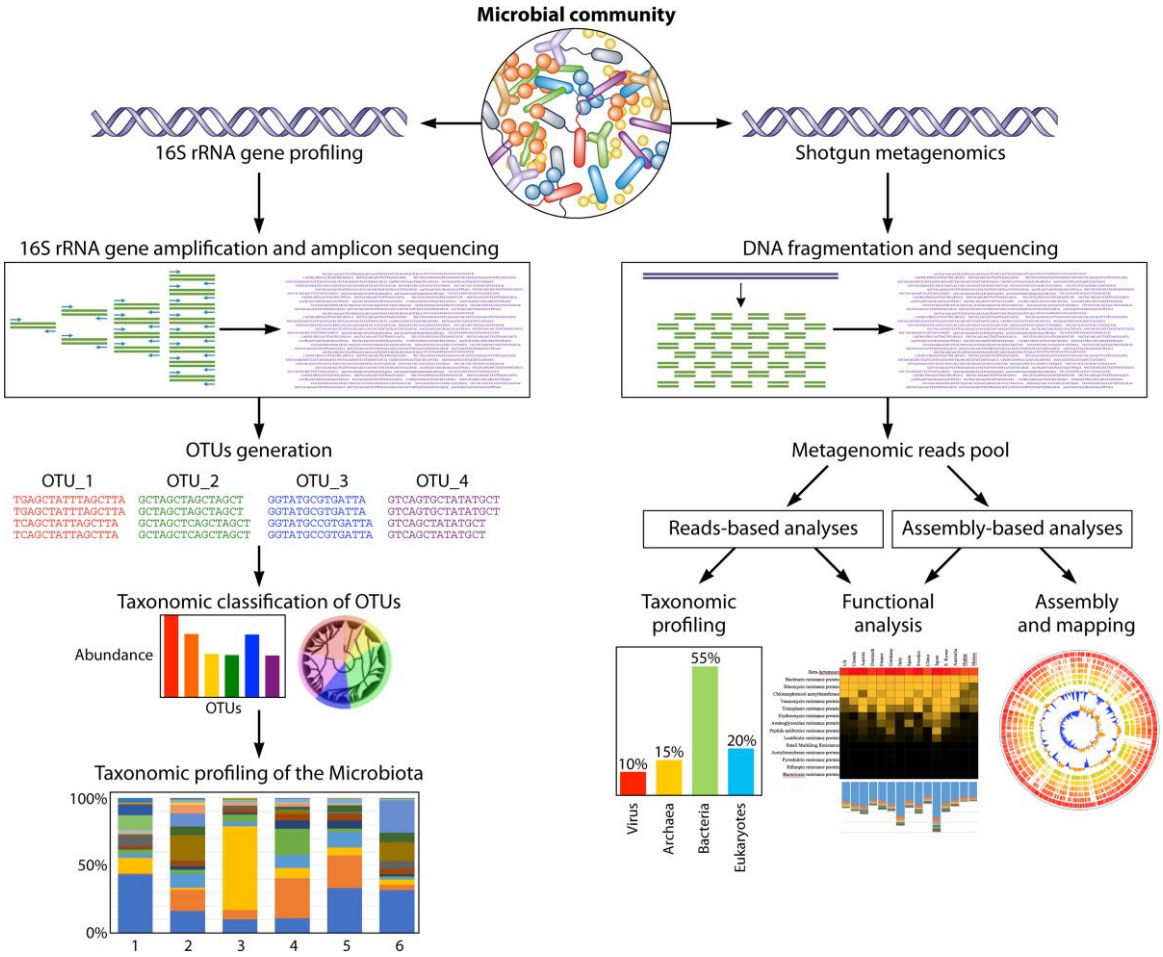


Figure 9 General overview on how to study microbial communities through 16S rRNA gene profiling and whole-metagenomes shotgun (WGS) (Milani et al., 2017)**.**

1.3.2 Microbial taxonomy

Microorganisms represent the largest undefined repertoire of biodiversity on Earth (Konstantinidis et al., 2006). It has been estimated that millions of prokaryotic species inhabit the Earth, but to date, only a small number of approximately 20.000 validly published names of prokaryotic species have been described, belonging to over 3.500 genera (Barco et al., 2020). SILVA databases encompass over 200.000 bacterial and archaeal species and 60.000 genera according to 16s rRNA taxonomy attribution (https://www.arb-silva.de/).

According to the small subunit ribosomal RNA (SSUrRNA)-based phylogeny, identified microorganisms are classified at least in 60 major phyla within the bacterial and archaeal domains, among which almost the 50% do not have cultivated representatives (so-called 'candidate' phyla). Moreover, more than 88% of all bacterial isolates belong to only four major phyla, the Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria (Rinke et al., 2013).

In the first place, phenotypic characters such as morphology, growth requirements, and pathogenic potential were used for bacterial taxonomy. Nowadays microbial taxonomy employs a polyphasic approach that encompasses genotypic, phenotypic, and chemotaxonomic data to characterize a microbial species (Ramasamy et al., 2014). This polyphasic approach takes into account the DNA-DNA hybridization (DDH), G+C content, 16s rRNA sequence, metabolites such as fatty acids, polar lipids, and exopolysaccharides, other than morphological and biochemical characterization (Varghese et al., 2015).

Since the late 1960s, the relatedness between two strains has been assessed by DDH, the "gold standard" method for species delineation with a settled threshold of 70% (Goris et al., 2007). Several methods for the measurement of DDH values have been described (Brenner et al., 1969; Crosa, J. H., Brenner, D. J. & Falkow, 1973; De Ley, 1970), but the experiments are anyhow time-consuming and labour-intensive (Hayashi Sant'Anna et al., 2019). Furthermore, the diverse protocols can yield different results in terms of reassociation values (Grimont et al., 1980).

The 16S rRNA gene sequencing became essential for bacterial taxonomy and nowadays is still among the first choice for a fast and preliminary identification of new microorganisms (Chun & Rainey, 2014). A first comparison between DDH and 16S rRNA proposed that strains with a score similarity of the 16s rRNA gene sequence < 97% could be ascribed to different species (Stackebrandt & Goebel, 1994). The cut-off was revised over the time, and currently it is comprised between 98.7 and 99% (Stackebrandt & Ebers, 2006). The utilization of 16S rRNA gene sequence for taxonomic attribution is based on its ubiquitous, stable, and highly conserved character (Hayashi Sant'Anna et al., 2019). However, 16s rRNA sequence similarity sometimes failed in the separation among two

closely related species (Chun & Rainey, 2014), with limitations related to the high level of conservation within some genera (e.g. the genus Brucella) and the possibility of being acquired by horizontal gene transfer (HGT) (Ramasamy et al., 2014).

Novel instruments for taxonomic delineation was provided by NGS that allowed the comparison of whole-genome data, leading to replacement of DDH with the digital DDH-like similarity indices (Kim et al., 2014). The approach of digital DDH (dDDH) was developed to perform an *in silico* DDH, its outcome presents good correlation to DDH's 70% threshold (Meier-Kolthoff et al., 2013), without presenting its drawbacks. Digital DDH can be accomplished with the software Genome-To-Genome Distance Calculator (GGDC) (https://ggdc.dsmz.de/ggdc.php).

Another genome based approach to infer taxonomy and phylogeny is Average Nucleotide Identity (ANI), firstly introduced by Konstantinidis & Tiedje (2005). Originally, ANI represented the arithmetic mean of the percent identity of all conserved protein coding genes between two strains identified through BLAST searches. ANI is calculated as the percentage of identity that can be found in the nucleotide sequences of orthologous genes common in the two genomes (Richter & Rossello, 2009). The robustness of the method is given by its simple use, the derivation from lineage-specific genes (in addition to the >1000 distributed genes), and the usage of more than a single gene like in the case of 16s RNA.

ANI value strongly correlates to DDH, in particular 94-95% of ANI is equivalent to 70% of DDH, and defines a new threshold for the species delineation in bacterial taxonomy (Goris et al., 2007; Richter & Rossello, 2009). Additional efforts were made to correlate ANI value to the 16s rRNA gene sequence, and it has been proposed that the sequence similarity threshold of 16S rRNA gene of 98.65% corresponds to 95-96% of ANI value (Kim et al., 2014). The use of ANI for bacterial classification is quickly grown, and already leads to re-classification of species within the genera *Pseudomonas* (Gomila et al., 2015), *Burkholderia* (Bach et al., 2017), *Acinetobacter* (Chan et al., 2012). However, ANI does not perform well in comparisons between strains belonging to different genera or higher taxonomic ranks (Qin et al., 2014).

Another method to evaluate relatedness among strains is the average amino acid identity (AAI) of all the common genes, that compares one-way and two-way reciprocal best hits between two sets of genomic protein data with BLASTX. In the AAI-based comparison a value of 95-96% corresponds to the 70% of DDH and defines the species (Konstantinidis & Tiedje, 2005).

Microbial Species Identifies (MiSI) method, takes into account both alignment fractions (AF) and ANI threshold for species, setting boundaries at 0.6 and 96.5%, respectively (Varghese et al., 2015). This approach was used by Barco et al. (2020) and showed several taxonomic issues within the genus *Clostridium.*

The definition for the existing taxonomic rank is far from being well delineated, especially at higher levels than the species. The definition of genera, families, and so on, are still mainly correlated to biochemical and morphological diversity among prokaryotes (Konstantinidis & Tiedje, 2005). For genera demarcation, mainly AAI and percentage of conserved proteins (PCOP) have been used, the former with value ranging from 65 and 72% and the latter with a threshold of 50% (Hayashi Sant'Anna et al., 2019). The PCOP expresses the ratio between the total conserved proteins of both the involved genomes and their total number of proteins (Hayashi Sant'Anna et al., 2019). This metric was already used to reassemble the genera *Chlamydia* and *Chlamydophila* to the single genus *Chlamydia* (Pannekoek et al., 2016), and to re-classify *Burkholderia andropogonis* to the novel genus *Robbsia* (Lopes-Santos et al., 2017).

For genus demarcation, 16s rRNA gene sequence has been proposed to a value of 94.5% (Yarza et al., 2014), even if in some cases this threshold underestimate genera diversity within some taxa (for example *Bacillaceae*) or overestimate it in others (*Enterobacterales*) as was demonstrated by Barco et al. (2020).

Sorting species into higher taxa still remains a fundamental problems in biology (Yarza et al., 2014). Nowadays, the illustrated genomic metrics have been recognized useful for taxonomic purpose even if their utilization is not yet mandatory. Over the past few years, more and more authors and taxonomists have pointed out the need to integrate genomic informatic on the purpose to better describe prokaryotes diversity (Barco et al., 2020).

1.4 Gut members of the order Clostridiales

1.4.1 Clostridiales

Human gut microbiota is composed of two main bacteria phyla, Firmicutes and Bacteroidetes, that account for more than 90% of total bacteria (Abbeele et al., 2013). Within Firmicutes, the order Clostridiales represents the major one, encompassing the highest microbial diversity, with the dominant genera *Clostridium, Eubacterium,* and *Ruminococcus* (Béchon & Ghigo, 2021).

Given the heterogeneity in terms of phenotypes and G+C content of the order of Clostridiales, a first attempt of revision for classification of clostridia was made in 1994 by Collins et al. (1994). 16S rRNA gene sequencing allowed the classification of clostridia in 19 clusters, with cluster I named as the *Clostridium sensu stricto* and containing the type strain of the genus, *C. butyricum*. Within the cluster I have been also identified species such as *C. acetobutylicum, C. beijerinckii,* and *C. tetani.* According to Collins et al. (1994) reclassification, also *C. celatum, C. intestinalis,* and *C. magnum* ascribed to the cluster I but exhibited loose relations with other species of the same group. The Cluster XI contained phylogenetically and phenotypically diverse assemblage of bacteria, and included for example *C. lituseburense, C. bifermentans, C. difficile, C. glycolicum, C. sordelli,* and *C. ghonii.*

Since that first classification, the taxonomy of clostridia is in continuous rearrangement, and leads to the identification of new genera within the family Clostridiaceae and of new families in the phylum Firmicutes (Chen et al., 2020; Lawson & Rainey, 2016).

According to NCBI Taxonomy, the order Clostridiales encompasses 26 families and a total of 284 genera, with the genus *Clostridium* that includes, to date, 153 rod-shaped, gram-positive and spore-forming anaerobes (Cruz-Morales et al., 2019) (http://www.bacterio.net).
In human and animal gut, *Clostridium* species represent one of the richest groups, especially the *Clostridium* cluster IV (Ruminococcaceae family) and XIVa (Lachnospiraceae family).
The first one is called *Clostridium leptum* group and contains 4 members, including *F. prausnitzii,* whereas the *Clostridium* cluster XIVa (known as *Clostridium coccoides* group) encompasses 21 species. This two clusters account for up to 40% of total bacteria in the gut (Guo et al., 2020). *Clostridium* species are mainly present in the large intestine, and in particular those belonging to the families Lachnospiraceae and Ruminococcaceae are concentrated in the mucosal layer of colon (Nagano & Itoh, 2012).

Commensal clostridia exert multiple benefits to host health. For instance, they are the major producers of butyrate as end-product fermentation, supplying the most important energy source for

colonocytes (Lopetuso et al., 2013). Thanks to this capability, *Clostridium* cluster XIV with the well-known butyrate producers *E. rectale* and *E. halii*, account for nearly 60% of the mucin-adhered microbiota (Lee et al., 2021). Considering their beneficial role for human health, they can represent a novel frontiers for the development of new probiotics (Guo et al., 2020).

On the other side, some intestinal Clostridiales are human pathogens, such as *C. perfringens, C. tetani,* and *C. botulinum* (members of *Clostridium* cluster I), with the main virulence factor being the production of toxins (Popoff & Bouvet, 2013).

A few other species not well-studied, including *Clostridium tertium*, *Clostridium cadaveris* and *Clostridium paraputrificum*, have been correlated with human infections and are likely opportunistic pathogens (Kiu et al., 2017).

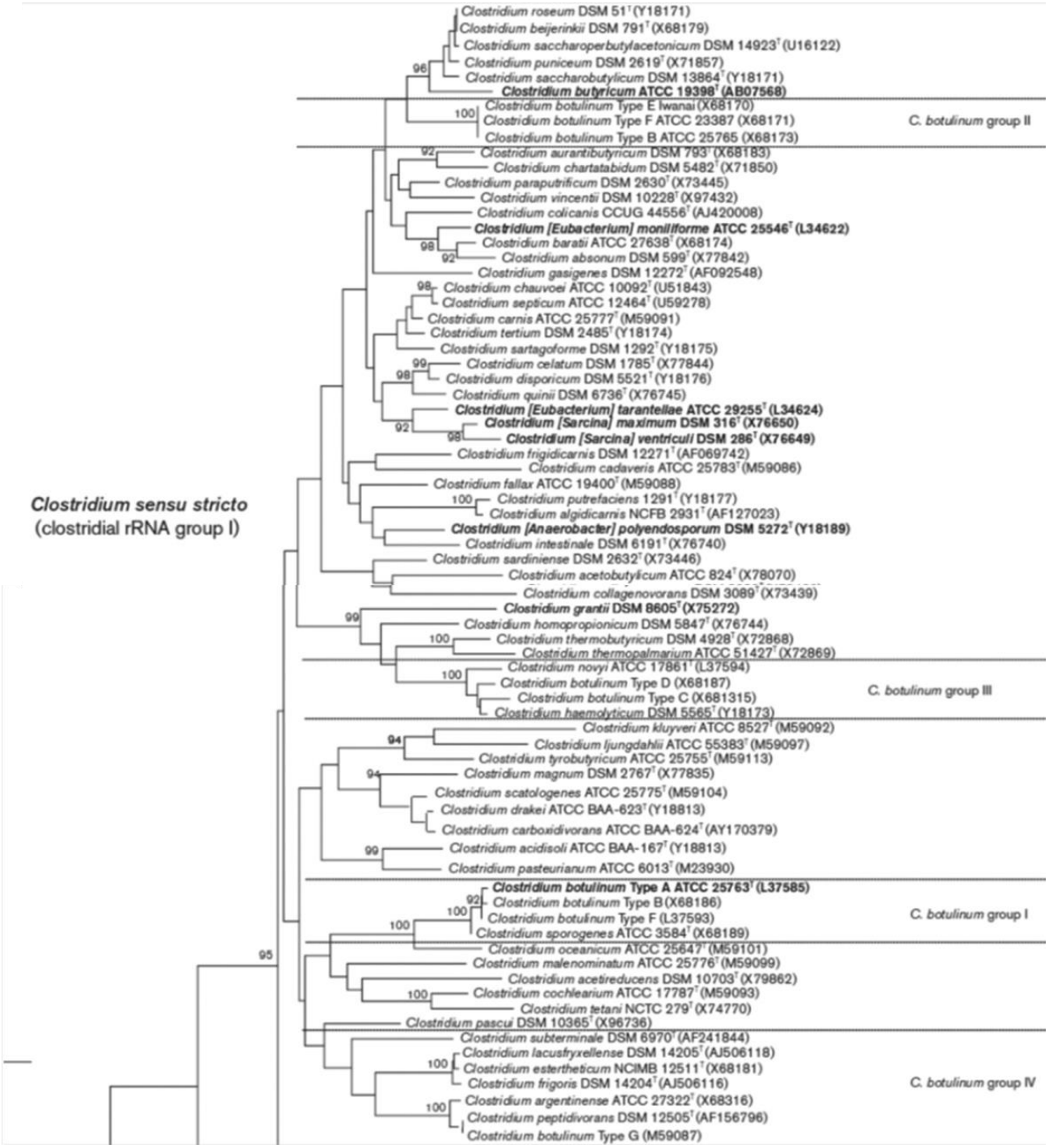


Figure 10 Phylogenetic tree of *Clostridium sensu stricto* (Guan & Goldfine, 2021)

1.4.1.1 *Clostridium tertium*

Clostridium tertium was firstly described by Herbert Henry in 1917, who isolated the bacterium from war wounds of soldiers. It was classified into the anaerobes saccharolytic group with the name of *Bacillus tertius.* Together with the two strains *Bacillus welchii* (today *C. perfringens*) and *Bacillus sporogens* (today *C. sporogens*), the strain *C. tertium* was among the most frequently isolates in wounds. According to Henry's description, the strain appeared like a thin bacillus, non-motile, spore-forming, capable to ferment mannitol and salicin (Henry, 1918). It was re-classified as *Clostridium tertium* in 1923 (Bergey et al., 1923).

C. tertium is a Gram-positive, slightly motile, anaerobic but aerotolerant bacterium that forms terminal and oval spores (Fig.11). It ferments glucose, fructose, galactose, lactose, mannose, maltose, sucrose, xylose, trehalose, esculin, mannitol, inositol and salicin while indole formation is negative (Bergey et al., 1923; Iseki & Okada, 1951). It is found in soil, animal, and human gut, and it is a component of the oral commensal microbiota. Unlike other clostridia, *C. tertium* doesn't produce toxins and forms spores only in anaerobic conditions (Vanderhofstadt et al., 2010). It produces enzymes directed against blood group A antigen in presence of glucosamine and GlcNAc (Howe et al., 1957). In 1963 was described the first case of bacteremia and septicemia caused by this strain, followed by other several clinical cases in literature (Shah et al., 2016; Vanderhofstadt et al., 2010). Most of these cases regarded neutropenic patients, with some exceptions (Barakat et al., 2018; Tappe et al., 2005; Wazir et al., 2019).

The pathogenic character of *C. tertium* remains still unclear and deserves further studies. *C. tertium* bacteremia is promoted by gut mucosal injury that allows bacterial translocation to the bloodstream, neutropenia, and therapy with β-lactam antibiotics, especially third generation cephalosporins (D. L. Miller et al., 2001; Wazir et al., 2019). Typically, patients who developed bacteremia are treated with a broad of antibiotics for 15 to 27 days (Shah et al., 2016). Despite the low pathogenic potential of this strain, it accounts for almost 5% of all clostridial bacteremia, being the most frequent after *C. perfringens* (Milano et al., 2019).

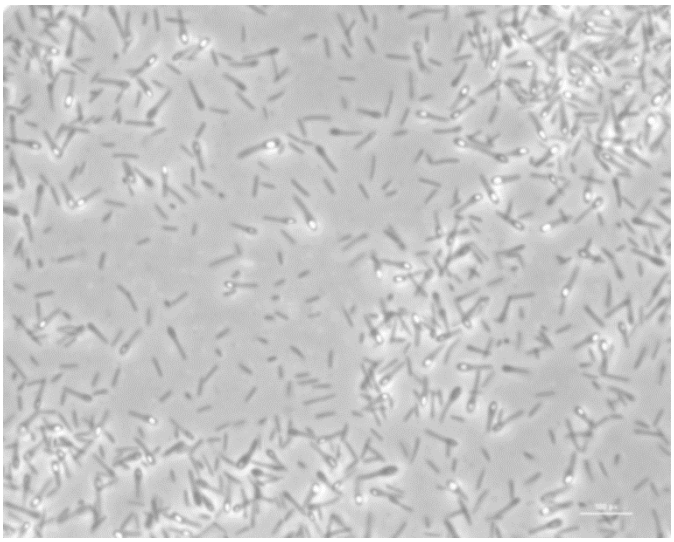


Figure 11 ***Clostridium tertium***

1.4.1.2 *Clostridium celatum*

Clostridium celatum is a Gram-positive, rod shaped, anaerobic, non-motile that forms large endospores (Fig.12). It was firstly isolated from stools of healthy adults, and it was as numerous as *C. perfringens*. It is a saccharolytic species that ferments cellobiose, fructose, galactose, glucose, lactose, maltose, mannose, ribose, sucrose, and trehalose. It produces acetic and formic acids other than ethanol from fermentation of glucose (Hauschild & Holdeman, 1974).

Limited information are available for *C. celatum* and in the last years two serious cases of human infection were reported for the first time (Agergaard et al., 2016).

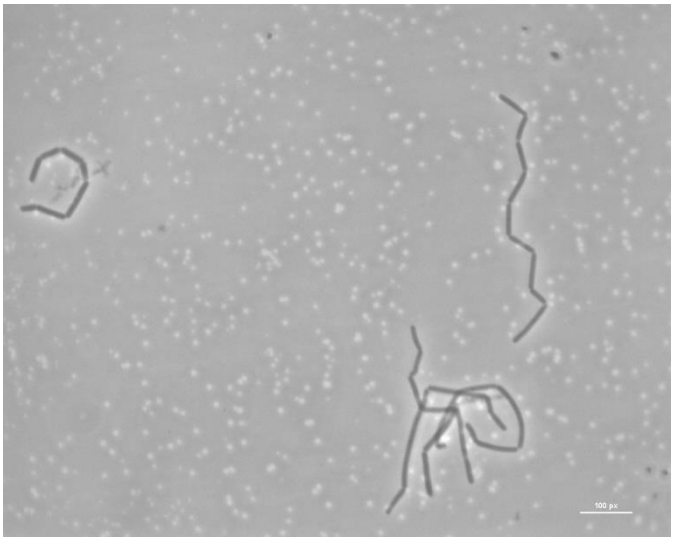


Figure 12 ***Clostridium celatum***

1.4.2 The genus *Paraclostridium*

The genus *Paraclostridium* belong to the family Peptostreptococcaceae of the Clostridia class. It was proposed in 2016 based on phenotypic features and G+C content, and includes Gram-positive, rod-shaped, producing endospore cells. Members of this genus are obligate anaerobes, motile, capable to grow on several substrates. The type species is *Paraclostridium bifermentans* (Sasi Jyothsna et al., 2016).

1.4.2.1 *Paraclostridium bifermentans*

P. bifermentans was firstly isolated from putrefied butcher's meat in 1902 and classified as *Clostridium* into the *Clostridium* cluster XI. In 2016 it was re-classified into the genus *Paraclostridium* (Kutsuna et al., 2019). It is commonly found in soil, sewage, and animal feces (Wong et al., 2014). This species is part of gut commensal microbiota, oral cavity, and female genital tract. It is a non-pathogenic bacterium, albeit few clinical infections have been reported in literature since 1980 (Barrett et al., 2020; Sankar et al., 2018). It can be involved in human diseases like endometriosis, endocarditis, empyema and metastatic osteomyelitis (Kutsuna et al., 2019).

P. bifermentans produces SCFAs such as butyric and acetic acid, lactate, and also formic acids, ethanol, butanol, carbon dioxide, and hydrogen (Leja et al., 2013). The process of spore germination in *P. bifermentans* is similar to that of *C. difficile,* involving the *csp* locus, and is stimulated by amino acids, in particular in response to L-alanine (Bhattacharjee & Sorg, 2018).

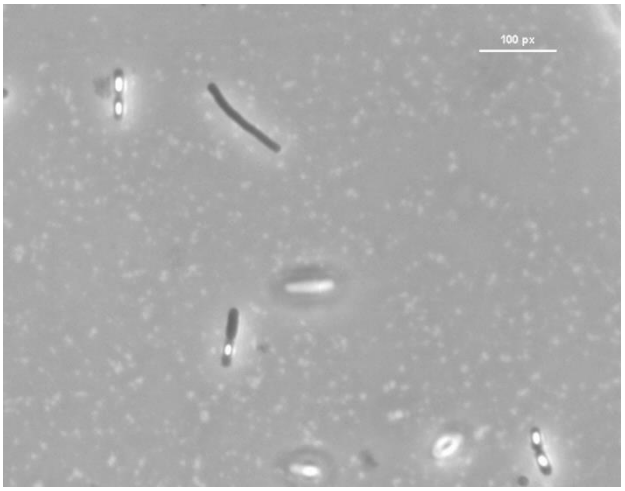


Figure 13 ***Paraclostridium bifermentans***

2. Aim

Gut microbiota, include more than 1000 heterogeneous species of microorganisms that constitute approximately 90% of all cells in the human body. Each subject harbours a different repertoire of hundreds of species, that feature his own microbiota, playing relevant roles in metabolism, pathogen resistance, immune system, and modulation of gastrointestinal development (Heintz-Buschart & Wilmes, 2018). A healthy microbiota remains relatively stable and presents high diversity (Wang et al., 2021), whereas changes in the composition of the gut microbiota are associated with a variety of diseases, including obesity, malnutrition, and inflammatory bowel disease, colorectal cancer, irritable bowel syndrome, allergies, type 1 and type 2 diabetes, celiac disease and obesity (de Vos & De Vos, 2012; Kelsen et al., 2012).

The so claimed mucosa-associated bacteria, are more properly mucus residing ones. In the large intestine of healthy subjects, the outer loose mucus layer provides a colonization and nutrition niche for the gut microbiota and contributes to the selection of various commensal microbes that make up our microbiota. On the other side, the inner layer of mucus is attached to the epithelium and forms an impenetrable barrier to the bacteria (Paone & Cani, 2020). The nutritive role of mucus largely stems from mucin glycoproteins, the major structural and functional units of the mucus barrier, used by mucus-associated gut bacteria as a source of carbon, nitrogen, and energy with production of short-chain fatty acids (SCFAs). Metabolism of mucin glycans requires the cooperation of several bacterial species, acting sequentially and concertedly to dismantle this complex structure (Wang et al., 2021). To date, there are little evidences concerning gut commensal bacteria involved in mucin degradation, with the species *Akkermansia muciniphila*, *Allobaculum mucolyticum*, *Bacteroides* spp., *Ruminococcus* spp., and *Bifidobacterium* spp. identified as mucinolytic (Berkhout et al., 2021; van Muijlwijk et al., 2021). Albeit mucus degradation is fundamental for human health, alteration rate of this process causes an increase permeability of the barrier, the disruption of microbiota composition and the beginning of inflammatory conditions. In this situation, mucin degradation may be related with the primary stage of mucosal diseases, like IBDs and cystic fibrosis (Bankole et al., 2021).

Considering the central role exerted by the intestinal mucus, it is of relevant interest to deeply investigate mucin degraders of the gut to unveil how mucins influence selection of commensal bacteria and how specific bacteria shape the mucus layer.

Major aims of this project were:

1. the isolation of mucin degrading bacteria;
2. the biochemical characterization of the mucinolytic species *Clostridium celatum*, *Clostidium tertium*, and *Paraclostridium bifermentans*;
3. the comparative genomic of *Clostridium celatum*, *Clostidium tertium*, and *Paraclostridium bifermentans*;
4. the study of the immunomodulatory properties of *Clostridium celatum*, *Clostridium tertium*, and *Paraclostridium bifermentans* in an *in vivo* model;
5. the inventory of Clostridiales encountered in gut microbiota of healthy subjects and the analysis of phylogenomic relationships among these species.

This research project aimed to identify gut bacteria of healthy subjects able to use mucins, through a culture dependent approach. The faeces of five healthy adults were subjected to three steps of strictly anaerobic enrichment in a medium with only mucins as carbon and nitrogen sources. Taxonomic characterization of the isolates resulted in the isolation of several different strains belonging to the species *Clostridium celatum*, *Clostridium tertium*, and *Paraclostridium bifermentans*, able to grow on mucin as sole carbon and nitrogen source. The same enrichment cultures were used to compare the bacterial community before and after the enrichment by 16S rRNA gene profiling (Raimondi et al., 2021).

The three mucinolytic species isolated in this study have been poorly characterized and only few information was available in scientific literatures. To better understand their functional role in gut, three strains, one for each species, were subjected to biochemical, physiological, and genomic characterization.

C. celatum, *C. tertium*, and *P. bifermentans* were investigated in terms of optimal growth medium, optimal pH and temperature, substrates fermented, and fermentation products. Phenotypical description of the isolates included the evaluation of their biofilm-forming ability. Antibiotic resistances tracked by genome analysis have been validated with broth microdilution method for MIC determination. Growth of anaerobic clostridia in the vegetative form is generally hampered in presence of oxygen, whereas in spore form, they can survive under harsh conditions and be exposed to air without damage (Egan et al., 2021). The tolerance to oxygen and the capability to produce resistance spores are major drivers of clostridia surviving in aerobic environments, contamination, and dissemination. As a matter of fact, pathogenesis of clostridia relies on the dormant spore morphotype (Morvan et al., 2021). Since preliminary evidence suggests a role of these species as

opportunistic pathogens (Hale et al., 2016; D. L. Miller et al., 2001), spore production and resistance to stress conditions was assessed for the three strains of *C. celatum*, *C. tertium*, and *P. bifermentans.* To evaluate whether spore formation improve survival, tolerance oxygen exposure and high temperature were checked.

Genome sequences were compared with other genomes present on NCBI GenBank database to better understand their metabolic and functional potential and to determine their ability to degrade mucin. Cleavage of mucin glycan is a complex process which involves a wide set of carbohydrate-active enzymes (CAZymes) (Berkhout et al., 2021). Genes coding for glycoside hydrolases (GHs) involved in mucin degradation were searched in all the genomes, followed by a comparative study on the metabolic pathway of single mucins' monosaccharides based on what was predicted by Ravcheev & Thiele (2017). To validate the predicted metabolic pathways, the three strains *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 have been grown in a basal medium supplemented with each of the mucins' monosaccharides, fucose, *N*-acetylneuraminic acid, *N*-acetylglucosamine, *N*-acetylgalactosamine, and galactose.

Gut mucus layers and epithelial cells are the primary gate keepers of bacterial interactions with the host immune system, playing a key role in host homeostasis and defence. However, little is known about how mucosa-associated bacteria contribute to maintain mucosal immunity and inhibit inflammatory condition (N. Shi et al., 2017). *In vivo* experiment on Balb/c mice have been performed, to test the ability of the three strains to colonize intestinal gut in the animal model. To assess their impact on the immune system, the phenotype of dendritic cells (DC) and lymphocytes $CD3^+$ in Peyer's patches was analyzed, and the expression level of different cytokines, like IFNγ, IL-4, IL-17, was measured.

The availability of the three genomes of *C. celatum*, *C. tertium*, and *P. bifermentans* encouraged a preliminary investigation of phylogenetic relationships among gut species ascribed to the class of Clostridiales. First of all, using two different datasets, all the detected species encompassed in this class were identified, and on the basis of the presence/absence and of relative abundances, the most frequently and abundant encountered were identified.
Phylogenetic relationships were reconstructed using ANI and AAI-RF, and the consistency of these outputs with similarities of genes encoding 16S rRNA was assessed.

3. Material and Methods

3.1 Mucin enrichment

3.1.1 Enrichment cultures

Fecal samples were collected from five healthy adults (3 men and 2 women, aged 25-50 year), who had not been treated with probiotics in the previous 2 weeks or with antibiotics for at least 3 months prior to sample collection. All the subjects enrolled in this study gave written informed consent, according to the experimental protocol approved with ref. no. 125-15 by the local research ethics committee (Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy). Fresh fecal samples were collected, sealed in anaerobic plastic bags (AnaeroGen, Oxoid, Basingstoke, UK), transferred in an anaerobic cabinet (Concept Plus, Ruskinn Technology, Ltd., Bridgend, UK) with an atmosphere of 85% N_2, 10% CO_2, and 5% H_2, and homogenized (10%, w/v) in sterile mucin-based medium (MM).

Mucin porcine stomach type II was provided by Sigma-Aldrich and was further purified according to the protocol by Miller and Hoskins (1981). From 25 g/L mucin suspension in 0.1 M NaCl containing 0.02 M phosphate buffer, pH 7.6, a clarified solution was obtained by centrifugation at 10,000 × g for 10 min at 4°C. Mucin was precipitated with ice-cold ethanol at 60% (v/v). The pellet was dissolved in 0.1 M NaCl, re-precipitated with ethanol (60% v/v) and dissolved in water. This mucin suspension was dialyzed (14 kDa cut-off) against water, and finally freeze dried. According to the producer, mucin had a molecular mass of 4,000–5,500 kDa, with a carbohydrate portion of 80–84%, and a protein portion of 16–20%. Proline, serine, and threonine residues in the peptide backbone comprised approx. 43% of the amino acid composition.

MM contained 3.0 g/L of purified mucin, 2.0 g/L KH_2PO_4, 4.5 g/L NaCl, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 0.045 g/L $CaCl_2 \cdot 2H_2O$, 0.005 g/L $FeSO_4 \cdot 7H_2O$, 0.01 g/L hemin, 0.05 g/L bile salts (Oxgall, BD Difco, Sparks, MD, United States), 0.6 mg/L resazurin, 2.0 mL/L minerals solution (0.5 g/L EDTA, 0.010 g/L $ZnSO_4 \cdot 7H_2O$, 0.003 g/L $MnCl_2 \cdot 7H_2O$, 0.03 g/L H_3BO_3, 0.02 g/L $CoCl_2 \cdot 6H_20$, 0.001 g/L $CuCl_2 \cdot 2H_2O$, 0.002 g/L $NiCl_2 \cdot 6H_2O$, and 0.003 g/L $NaMoO_4 \cdot 2H_2O$), 1.4 mL/L vitamins solution (1.0 g/L menadione, 2.0 g/L biotin, 2.0 g/L calcium pantothenate, 10 g/L nicotinamide, 0.5 g/L cyanocobalamin, 0.5 g/L folic acid, 4 g/L thiamine, and 5 g/L PABA), and 40 mL/L reducing solution (12.5 g/L L-cysteine • HCl and 80 g/L $NaHCO_3$). The basal medium was autoclaved at 121°C

for 20 minutes and supplemented with filter sterilized (0.22 µm) minerals, vitamins, and reducing solutions.

The fecal slurries were inoculated with a syringe (10% v/v) into butyl-rubber stoppered bottles containing 50 mL of sterile anaerobic MM. After an incubation of 72 h at 37 °C, these cultures were utilized to seed (10% v/v) the next passage in MM, repeated to accomplish three enrichment steps. Once grown, the third enrichment was utilized for isolation of mucinolytic bacteria.

The viable counts of anaerobic mucinolytic bacteria and Enterobacteriaceae in enriched samples were determined in triplicate onto agar plates of MM and MacConkey, respectively.

For both the media, one-way ANOVA followed by Tukey's post-hoc analysis was utilized to compare the counts of the different subjects. Paired samples t-test was utilized to compare MM and MacConkey counts.

3.1.2 Isolation of mucinolytic bacteria

All the steps of the isolation and purification procedure were carried out in the anaerobic cabinet. Soft-agar MM plates, containing 8 g/L of agar (BD Difco), were used to count and purify mucin degraders. While, MacConkey agar plates (BD Difco), were used to count *Escherichia coli* and Enterobacteriaceae and to discriminate them from strictly anaerobic isolates.

The EC was serially diluted (1:10) in anaerobic MM, then 1 mL of each dilution was poured into plates and covered with 20 mL of melted soft-agar MM. The plates were allowed to cool and incubated in anaerobiosis at 37° C for 72–96 h. Inclusion colonies were picked and checked for aerobic growth at 37°C in MacConkey agar plates. Colonies unable to grow in aerobiosis were further purified at least three times by serial dilution and isolation in soft-agar MM.

3.1.3 RAPD-PCR Fingerprinting

Approximately 10 pure clones per sample were fingerprinted by RAPD-PCR according to (Quartieri et al., 2016) and clustered into biotypes with a similarity level of 75% using the Pearson correlation coefficient. The RAPD-PCR mix consisted of Dream Taq Buffer (Thermo Fisher Scientific, Waltham, MA), 0.5 μM of M13-RAPD primer (5'-GAGGGTGGCGGTTCT-3'), 100 μM of each dNTP, 0.75 U Taq polymerase (DreamTaq, Thermo Fisher Scientific), and 50 ng of genomic DNA from the isolates. DNA amplification was carried out with the following protocol: 94°C for 4 min; 45 cycles at 94°C for 1 min, 34°C for 1 min, and 72°C for 2 min; 72°C for 7 min. PCR products were electrophoresed in a 2% agarose gel (25 × 25 cm) for 4 h at a constant voltage (160 V) in TAE buffer. After staining with ethidium bromide, RAPD-PCR products were digitally captured with Gel Doc XR+ (Bio-Rad, USA), and analysed with Gene Directory 2.0 software (Syngene, UK).

3.1.4 Taxonomy assignment

Single biotypes from each sample were taxonomically characterized by partial sequencing of the 16S rRNA gene (Eurofins Genomics, Ebersberg, Germany), utilizing primers targeting the V1-V4 regions (Raimondi, Luciani, et al., 2019). Amplification was performed using 16S 500f (5'-TGGAGAGTTTGATCCTGGCTCAG-3') and 16S 500R (5'-TACCGCGGCTGCTGGCAC-3') primers within 50 μL of PCR Master Mix (Thermo Fisher Scientific), containing 0,2 μM of each primer, and 50 ng of DNA. Amplification was carried out with the following parameters: 94°C for 4 min; 30 cycles at 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s; 72°C for 7 min.

Sequences were compared to the NCBI refseq_rna database with the alignment tools BLAST and MUSCLE (https://www.ncbi.nlm.nih.gov/; https://www.ebi.ac.uk/Tools/msa/muscle/).

3.2 Genomic and physiological characterization

3.2.1 Genomics

3.2.1.1 Genome extraction and sequencing

For whole-genome sequencing, each strain was cultivated in MM at 37°C for 48h under strictly anaerobic conditions. Biomass was collected by centrifugation for 30 min at 14,000 rpm.

The genomic DNA was extracted with DNeasy Blood & Tissue kit (Qiagen GmbH, Düsseldorf, Germany). Before DNA purification, the pre-treatment for Gram-positive bacteria was followed, making some modifications:

- incubation for 2 h at 37°C with twofold the volume of enzymatic lysis buffer;
- incubation for 1 h at 56°C with twofold the volume of proteinase K and Buffer AL.

The DNA concentration was quantified with a Qubit 3.0 fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA). The samples were sequenced with Illumina NovaSeq 6000 by Eurofins Genomics (Ebersberg Germany). For each sample, 150-bp paired-end reads were obtained.

3.2.1.2 Genome assembly

The raw reads were checked for quality with FastQC v0.11.8 (Andrews, 2010). To trim Illumina adapters and delete reads with a quality score lower than 20, the tool Cutadapt v1.16 was used with the following parameters: overlap=15, minimum length=30 and quality-cutoff=20 (Martin, 2011). Once obtained the trimmed reads, they were verified again with FastQC and then assembled with SPAdes v3.13.0 with parameters -careful, -cov-cutoff auto and -k auto (Bankevich et al., 2012). The quality of the assemblies was evaluated using QUAST 5.0.2 (Gurevich et al., 2013). Trimming, quality checking, and assembly were performed on Galaxy platform (usegalaxy.eu) (Afgan et al., 2018). The annotation was carried out with the tool Prokka v1.14.6 (Seemann, 2014). The taxonomy attribution was confirmed with tools SpeciesFinder (https://cge.cbs.dtu.dk/services/SpeciesFinder - (Larsen et al., 2014) and TypeMat of Microbial Genomes Atlas (MIGA-http://microbial-genomes.org) (L. M. Rodriguez-R et al., 2020).

The completeness of the assemblies was verified with CheckM 1.0.8 (Parks et al., 2015).

3.2.1.3 Genome analysis

CRISPRs and Cas genes were searched using CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) (Couvin et al., 2018), with default settings and subtype clustering of Cas genes. The presence of prophage sequences was analysed with PHASTER (https://phaster.ca/) (Arndt et al., 2016). The identification of antibiotic resistance genes was carried out using the web tool RGI (Resistance Gene Identifier) of CARD (Comprehensive Antibiotic Resistance Database), processing the contigs file for "Perfect, Strict and Loose hits" (https://card.mcmaster.ca/analyze/rgi) (Alcock et al., 2020). Insertion sequences were identified with ISFinder (https://isfinder.biotoul.fr/, Siguier et al., 2006). BAGEL 4 server was used to search bacteriocins genes (http://bagel4.molgenrug.nl/) (Van Heel et al., 2018).

The genomes of the other available *C. celatum*, *C. tertium* and *P. bifermentans* strains were downloaded from GenBank. The respective accession numbers are reported in Tab.1. As a whole, comparative genomics investigated 4 genomes of *C. celatum*, 15 of *P. bifermentans* and 9 of *C. tertium.*

3.2.1.4 Comparative genomics

Average Nucleotide Identity (ANI) and digital-DNA/DNA hybridization (dDDH) were calculated using the ANI Matrix web tool (http://enve-omics.ce.gatech.edu/g-matrix/) and Genome-to-Genome Distance Calculator GGDC 2.1 (https://www.dsmz.de/services/online-tools/genome-to-genome-distancecalculator-Ggdc), with the thresholds for species demarcation of 95% for ANI and 70% for dDDH (Richter & Rossello, 2009). Roary was utilized to calculate the pangenome and to define core and accessory genes, utilizing Prokka annotation files (Page et al., 2015).

KEGG tools, BlastKOALA and MAPPER, were used to predict functional properties of the genomes (Kanehisa et al., 2016; Kanehisa & Sato, 2020). The dbCAN2 metaserver (https://bcb.unl.edu/dbCAN2/) was used for the annotation of Carbohidrate Active Enzyme (CAZymes) in all the strains (H. Zhang et al., 2018). Only the CAZymes annotated by two out of the three tools (HMMER, Hotpep and Diamond) used by dbCAN2 were considered. The putative glycoside hydrolases involved in the mucin degradation were chosen according to the classes reported by Tailford *et al*., 2015.

	Strain	GenBank Accession number	Isolation source
C. celatum	**WC 0700**	**GCA 021412605.1**	**Human feces**
	D43t1_170807_C1	GCA_015547475.1	Human feces
	DSM 1785	GCA_000320405.1	Human feces
	MGYG-HGUT-01423	GCA 902374805.1	Human gut
P. bifermentans	**WC 0705**	**GCA_021412665.1**	**Human feces**
	SU1074NT	GCA 002104065.1	Soil
	MHMC-14	GCA_003585605.1	Human feces
	parabai	GCA_006782905.1	Unknown
	ATCC 638	GCA 006802875.1	Unknown
	Med78_4-601-WT-2	GCA_012843415.1	Pig feces
	BSD2780061688_150302_D6	GCA_015551995.1	Human feces
	BSD2780061688_150302_X7	GCA 015668845.1	Human feces
	S01	GCA_017288015.1	Soil
	N73	GCA 018917045.1	Monkeys' feces
	DSM 14991	GCA_019916025.1	Soil
	MGYG-HGUT-00024	GCA_902362415.1	Human gut
	ATCC 19299	GCA 000452225.2	Unknown
	PAGU 1678	GCA_019650315.1	Biobreeding rat feces
	WYM	GCA_000498455.1	Landfill leachate sludge
C. tertium	**WC0709**	**GCA_018967925.1**	**Human feces**
	Gcol.A2	GCA_003284625.1	Human feces
	Gcol.A43	GCA_003284645.1	Human feces
	BSD2780120874_150323_E10	GCA_015560625.1	Human feces
	BSD2780120875b_170604_A12	GCA_015667715.1	Human feces
	1001275B_160808_H2	GCA_015668435.1	Human feces
	DSM 2485	GCA_017873245.1	Unknown
	Clostridium tertium	GCA_900217175.1	Preterm infant faeces
	MGYG-HGUT-01328	GCA_902373955.1	Human gut

Table 1 List of the strains utilized in this work. In bold the strains isolated during this project.

3.2.2 Physiological characterization

3.2.2.1 Growth at different pH

Cultures of *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709 were incubated in anaerobiosis at 37° C in MM medium. The initial pH was adjusted to 5.5, 6.0, 6.5, 7.0, 7.5, and 8.0 with 1 M KOH and 1 M HCl solutions. The cultures were seeded (10% v/v) with 48h MM cultures. After 24, 48, and 72 hours of fermentation, supernatants were recovered after centrifugation (13,000 xg, 5 min, 4° C) and stored at -20° C. Products such as lactic, succinic, propionic, acetic, butyric, valeric, isovaleric, formic acid, 2,3-butanediol and ethanol were detected using a HPLC device (Agilent technologies, Waldbronn, Germany) provided with refractive index detector (RID) and Aminex HPX-87 H ion exclusion colon. Isocratic elution was performed with H2SO4 0.005 M at 0.6 mL/min according to Amaretti et al. (2013).

3.2.2.2 API 50 CHL

The strains *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709 were tested for the fermentation of 49 carbohydrates using API 50 CH test strips (bioMerieux, Marcy-l'Etoile, France). Bacterial biomass from the surface of M17-agar plates (BD Difco, Sparks, USA) was harvest with a swab and used to produce a suspension of 2 McFarland in 10 ml of API 50 CHL Medium. The suspension was used to inoculate each tube of the strip in the anaerobic chamber. The results were read after 48 h of incubation at 37°C.

3.2.2.3 Biofilm formation

Biofilm-forming ability and mucin adherence of the strains were tested for the three isolates *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709. Purified mucin type III was dissolved in phosphate-buffered saline (PBS) pH 7.4 to a final concentration of 1 mg/mL, 200 µl of mucin suspension were loaded in each well of 96-well microtiter plates, that were maintained overnight at 4° C. Excess of mucin was removed and wells were washed with PBS (Sadiq et al., 2021). Cultures grown anaerobically in M17 (BD Difco, Sparks, USA) medium for 48 h were diluted

(10% v/v) in fresh M17 medium and seeded in mucin-coated and uncoated wells. Plates were incubated under anaerobic condition for 48 h and biofilm formation was assayed by crystal violet (CV) staining according to Sicard et al., (2018). Briefly, unattached cells were discarded and each well washed three times with PBS. Biofilm was stained with CV solution at 0.1% (wt/vol), excess of staining was removed, and washed three times with PBS. De-staining solution (80% v/v ethanol, and 20% v/v acetone) was added to release the stain, and biofilm was quantified by measuring OD_{570}. Specific biofilm formation (SBF) index was calculated as the ratio between CV absorbance at 570 nm and culture's turbidity at 620 nm, setting a threshold of 1. *E. coli* 03.73 and *K. pneumoniae* 11.71 were used as positive and negative controls respectively (Amaretti et al., 2020; Raimondi, et al., 2019b).

3.2.2.4 Tolerance to oxygen and temperature

Tolerance to oxygen and to high temperature were tested with cultures grown in M17 broth. Pure cultures were inoculated (10% v/v) and incubated 48 h at 37° C in anaerobic condition. Portion of the same cultures were:

- (I) heated at 80°C for 30 minutes.
- (II) transferred in a 10x volume baffled flask and exposed to air for an hour during incubation in an orbital shaker at 180 rpm at 37° C;
- (III) exposure to air (II) than heated at 80°C for 30 minutes.

After each treatment, serial dilutions were made, all incubated in anaerobic condition. Most Probable Number (MPN) technique was used to estimate microbial population size. Untreated anaerobic cultures were used as controls.

3.2.2.5 Antimicrobial susceptibility test

The susceptibility to antibiotic of *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709 was assayed with broth microdilution method according to the International Standard Organisation (ISO 20776-1:2019). Cultures grown in M17 medium for 48 h at 37° C, were diluted in fresh M17 broth to obtain a final concentration of 5 x 10^5 CFU/mL. The antibiotics were tested at doubling dilutions, from 0.0625 to 64 mg/L for ampicillin, gentamicin, and chloramphenicol, and from 16 to 0.015 mg/L for penicillin G and tetracycline.

The antibiotic susceptibility was determined after 48 h of incubation at 37°C in anaerobiosis. MICs were defined according to the European Committee on Antimicrobial Susceptibility Testing breakpoints (EUCAST- https://www.eucast.org/clinical breakpoints/ version 11.0) for Gram-positive anaerobes, except for tetracycline and penicillin G, for which breakpoints were defined according to Clinical and Laboratory Standards Institute (CLSI- https://clsi.org/meetings/microbiology/ M100-Ed31:2021). No defined breakpoints were available for gentamicin in neither of the two databases. Indicatively, a gentamicin breakpoint of 16 was indicated in the work of Wei et al., (2020), referring to the CLSI-M100-S23:2019 database.

3.2.2.6 Growth on mucins' monosaccharides

The capability of *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 to growth fermenting the single monomeric units composing the oligosaccharides of mucin was assessed on basal M17 medium, formulated without any carbohydrate, and supplemented of D+Glucose (Sigma), L-Fucose (Carbosynth), *N*-Acetylneuraminic acid (Carbosynth), D+Galactose (Sigma), *N*-acetyl-D-glucosamine (Carbosynth), and *N*-acetyl-D-galactosamine (Carbosynth), all at the final concentration of 20 mM. M17 medium without lactose was used as negative control, while supplementation with glucose was used as positive control. All the media were adjusted with HCl 1 M to pH of 6.9-7.2.

C. celatum WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 were grown for two consecutive steps on M17 medium with lactose (1 g/L) at 37 °C for 48 h, in order to obtain inocula where the carbon source was depleted. These cultures were used to seed (10% v/v) tubes containing M17 supplemented with the diverse carbohydrates, and the medium without supplement as control. For each medium, three passages were carried out, each one in triplicate. Cultures were incubated at 37 °C for 48 h. At the end of each step, OD_{600} was determined.

3.2.3 *In vivo* study

The *in vivo* study has been performed in UNIPD, under the supervision of Prof. Ignazio Castagliuolo. The protocol was approved by the Animal Care and Use Ethics Committee of the University of Padova under license from the Italian Ministry of Health, in compliance with the national and European guidelines for handling and use of experimental animals.

3.2.3.1 Preparation of cryoprotected cultures

P. bifermentans WC 0705 and *C. tertium* WC 0709 were grown anaerobically for 48 h in M17 broth. The biomass of each strain was collected after centrifugation and suspended in M17 broth containing 10% v/v glycerol to obtain single-dose aliquots. The cryoprotected cultures were stored at – 80 °C.

Because of the volumes, strict anaerobiosis was not possible, and oxygen impacted at different extent on the viability of the cryoprotected cultures. According to the initial charge evaluated by MPN, the theoretical charge of the frozen sample (0.1 ml) was expected to be 10^{10}. However, when the glycerinates were thawed and enumerated, the live bacteria were 10^8 in both *C. tertium* WC 0709 and *P. bifermentans* WC 0705. Aliquots of each were thawed one by one and daily used.

3.2.3.2 Mice and treatment

Male BALB/c 8 weeks old were purchased from Envigo (San Pietro al Natisone, Udine, Italy). Mice were housed in groups of 4 per cage and maintained in a temperature-controlled environment (22 ± 2°C) under a 12-hour light/dark cycle with regular chow food and tap water provided *ad libitum*. Following a week of acclimation, mice were randomly assigned to receive vehicle alone (control group), 10^8 CFU of *C. tertium* WC 0709 or *P. bifermentans* WC 0705. The treatments were administered once a day for 10 days by intragastric gavage (24 gauge, 9-cm catheter) in a total volume of 100 μl.

3.2.3.3 Isolation of mononuclear cells

At the end of the treatment, mice were sacrificed. The colon was removed, longitudinally opened, washed to eliminate the luminal content, and immediately placed in ice-cold RPMI medium. Using glass coverslips, the mucosa was surgically removed and digested for 30 minutes in 0.15% DTT/HBSS at room temperature under gentle shaking (70 rpm) to remove residual mucus. The mucosal strips were rinsed in HBSS and incubated at room temperature in 1 mM EDTA/HBSS for 30 minutes to eliminate epithelial cells. The samples were then rinsed in HBSS, cut into 5-mm-square pieces, and placed in 5 mL enzymatic digestive solution containing 10 mg/mL collagenase type II from *Clostridium histolyticum* (Merck) and 10 μg/mL type I deoxyribonuclease (Merck). Digestion was carried on at 37°C for 30 min under continuous shaking. Cells were harvested by centrifugation (6 minutes, 1600 rpm), resuspended in PBS, and filtered through a 100 μm cell strainer to remove tissue debris. Cell suspensions were counted and used for flow cytometry analysis.

3.2.3.4 Flow cytometry

Freshly prepared cells ($1x10^6$/mL) were incubated for 30 min in ice-cold PBS containing 2% BSA (w/vol) to block unspecific bindings. Cells were washed by centrifugation in ice-cold PBS (1600 rpm, 6 min, twice) and incubated at 4°C for 30 minutes with antibodies (Table 2) diluted in blocking buffer. When required, cells were washed and subsequently incubated with secondary fluorescent labeled antibodies for 30 minutes. The fluorescent signals were recorded using BD-FACSCalibur (Becton Dickinson) and CellQuest software (Becton Dickinson). The results were analyzed using the Flowing Software 2.

Antibody	Clone	Source
CD3 PE conjugate	17A2	eBioscience
CD4 Cy7 conjugate	50134-R001	Sino Biological Inc
CD8 Cy7 conjugate	53-6.7	Invitrogen
IL4 FITC conjugate	BVD4-1D11	Invitrogen
IFNγ FITC conjugate	XMG1.2	eBioscience
CD25 APC conjugate	3G10	Invitrogen
Foxp3 FITC conjugate	polyclonal	Invitrogen
IL17 FITC conjugate	TC11-8H4	Molecular probes
CD11c Cy7 conjugate	N418	eBioscience
CD80 FITC conjugate	B7-1	Invitrogen
TLR2	polyclonal	Abcam
TLR4	polyclonal	SantaCruz
F4/80 FITC conjugate	CI:A3-1	eBioscience
anti-rabbit FITC		Invitrogen

Table 2 Antibodies used in flow cytometry analysis

3.2.4 Statistical analysis

For counts of enriched samples in MM and MacConkey medium, one-way ANOVA followed by Tukey's post-hoc analysis was utilized. Paired samples t-test was utilized to compare MM and MacConkey counts. In all the experiments, the data reported are the means of three independent experiments, each carried out in triplicate. The statistical significance was analysed with t-test ($P<0.05$). Data from flow cytometry analysis are reported as the mean ± standard error. Statistical analysis was performed using GraphPad Prism 7 software (GraphPad, San Diego, CA, United States). Statistical differences were assessed by one-way ANOVA and Bonferroni multicomparison post hoc tests. Differences were considered significant at $P < 0.05$.

3.3 Phylogenomic approaches

3.3.1 Metagenomic samples

Two datasets have been used to investigate gut species belonging to the order Clostridiales: the Human Microbiome Project (HMP- Turnbaugh et al., 2007) and 60 publicly available metagenomes of intestinal microbiota from healthy subjects. Samples of the 60 metagenomes collected from NCBI Sequence Read Archive (SRA) were obtained from healthy subjects from five different countries: China (CHN), Ethiopia (ETH), Spain (ESP), United States of America (USA), and Sweden (SWE), and had the accession numbers: PRJNA557323, PRJNA504891, PRJEB33098, PRJEB27308, PRJEB7369, respectively. The selected metagenomes were sequenced through whole-genome shotgun sequencing on Illumina paired-end platforms and produced reads ranging between 100 and 150 bp in length.

3.3.2 Dataset definition and analysis

Public available composition profiles of 554 HMP stool samples analyzed with MetaPhLan2 (Truong et al., 2015) were used. The 60 metagenomes retrieved from NCBI SRA were profiled with the same tool, using default parameter, and ignoring viruses and archaea profiling.

590 reference genomes of species belonging to *Clostridiales* were downloaded from NCBI Datasets, as of October 2021, and used to filter species found in gut microbiota. Type strain genomes of *Eubacterium rectale*, *Eubacterium siraeum* and *Clostridium leptum* were added to the defined subset in order to include other species that are typically related to the human gut microbiota (Kabeerdoss et al., 2013; Mukherjee, 2020; Newman et al., 2021).

The Average Nucleotide Identity (ANI) and Average Amino acid Identity (AAI) between all the genomes were calculated with the scripts ANI Calculator and AAI calculator of the enveomics collection (http://enve-omics.ce.gatech.edu/enveomics/) by all-against-all approach (L. Rodriguez-R & Konstantinidis, 2016). Whole-genome based Average Nucleotide Identity (gANI) and Alignment Fraction (AF) were calculated using the MiSI (Microbial Species Identifier; https://ani.jgi.doe.gov/html/home.php?) as proposed by Varghese et al. (2015). For genera demarcation, the thresholds proposed by Barco et al. (2020) were utilized (ANI > 73.98; AF> 0.331).

The genomes in the subset were annotated with Prokka (Seemann, 2014), the 16S rRNA sequences were extracted from each genome and aligned with Clustal Omega (Sievers et al., 2011). 16S rRNA gene alignment was utilized to generate a phylogenetic tree constructed with the maximum likelihood method using RAxML tool (Stamatakis, 2014). ANI and AAI distance matrices were used to compute Unweighted Pair Group Method with Arithmetic mean (UPGMA) unrooted phylogenetic trees with DendroUPGMA tool (http://genomes.urv.cat/UPGMA/, Garcia-Vallvé et al., 1999). The genomes in the phylogenetic trees were visualized with iTOL (Letunic & Bork, 2019).

The tree of AAI was compared with those constructed from the alignment of 16S rRNA gene, ANI and gANI distance matrices utilizing the following 'generalized' Robinson-Foulds metrics: the Jaccard–Robinson–Foulds (JRF), computed with k = 1 (Nye et al., 2006), and the information-based measures of Mutual Clustering Information (MCI) and Shared Phylogenetic Information (SPI) (M. Smith, 2020). JRF, MCI, and SPI were computed with the R package 'TreeDist', archived at https://dx.doi.org/10.5281/zenodo.3528123. Comparison relied on matching each informative split within a tree (i.e. an internal branch of the tree, with at least 2 leaves at each extremity) with an informative split within the other tree. For each metrics, the comparison yielded a normalized similarity score in the range 0–1 between two trees and a score for each match of paired tree splits.

4 Results

4.1 Mucin enrichment

4.1.1 Isolation of mucin degraders

The faeces of five healthy adults were subjected to three steps of strictly anaerobic enrichment in a medium with only mucins as carbon and nitrogen sources Enumeration and isolation of mucin degrading bacteria were carried out from enriched samples after the three enrichment steps in MM broth. To discriminate anaerobic bacteria from mucinolytic Enterobacteriaceae, for the enumeration step were used MM plates in anaerobiosis and MacConkey plates in aerobiosis (Tab 3). The obtained isolated from MM plates were confirmed to be strictly anaerobes. Anaerobic mucin degraders were found in the magnitude of 10^8 cfu/mL (mean ± SD = 8.3 ± 0.3 Log_{10} cfu/mL), significantly more abundant ($P < 0.05$) than mucinolytic Enterobacteriaceae, that ranged between 10^7 and 10^8 cfu/mL (mean ± SD = 7.6 ± 0.4 Log_{10} cfu/mL). The viable counts of both anaerobic mucin degraders and mucinolytic Enterobacteriaceae did not differ among the enriched samples (ANOVA, $P > 0.05$).

Sample	Anaerobic mucinolytic bacteria Log (cfu/mL)	Enterobacteriaceae Log (cfu/mL)	Isolates
EC 1	7.8 ± 0.9	7.2 ± 0.5	WC0700
			WC0701
EC 2	8.4 ± 0.7	7.3 ± 0.4	WC0702
			WC0703
EC 3	8.5 ± 0.5	7.9 ± 0.5	WC0704
EC 4	8.3 ± 0.6	7.6 ± 0.6	WC0705
			WC0706
EC 5	8.4 ± 0.5	8.0 ± 0.3	WC0707
			WC0708
			WC0709

Table 3 Cultivable mucin degraders in EC samples. The viable counts of anaerobic mucinolytic bacteria and Enterobacteriaceae respectively counted onto plates of MM and MacConkey medium is reported. Counts are reported as mean Log (cfu/mL) ± standard deviation, n = 3. For both the media, no significant difference was observed among samples (ANOVA, $P > 0.05$). Cultivable anaerobic mucinolytic bacteria were significantly more numerous than Enterobacteriaceae (paired samples t-test, $P < 0.05$).

Based on RAPD-PCR fingerprinting, the anaerobic isolates clustered in 1–3 distinct biotypes *per* sample, for a total of 10 strains (Tab. 4). The partial sequences of 16S rRNA genes showed identity > 98.6% to the following species: *Clostridium disporicum* (6 biotypes), *Clostridium tertium* (1), and *Paraclostridium benzoelyticum* (3).

Sample	Isolate	ref_seq ID	BLAST result	Identity
EC 1	WC0700	NR_026491.1	*Clostridium disporicum*	98.70%
		NR_144696.1	*Clostridium saudiense*	97.98%
	WC0701	NR_026491.1	*Clostridium disporicum*	99.14%
		NR_144696.1	*Clostridium saudiense*	98.14%
EC 2	WC0702	NR_026491.1	*Clostridium disporicum*	99.90 %
		NR_144696.1	*Clostridium saudiense*	98.28%
	WC0703	NR_026491.1	*Clostridium disporicum*	99.44%
		NR_144696.1	*Clostridium saudiense*	97.85%
EC 3	WC0704	NR_026491.1	*Clostridium disporicum*	99.08%
		NR_144696.1	*Clostridium saudiense*	98.01%
EC 4	WC0705	NR_148815.1	*Paraclostridium benzoelyticum*	99.43%
		NR_113323.1	*Paraclostridium bifermentans*	99.32%
	WC0706	NR_026491.1	*Clostridium disporicum*	99.50%
		NR_144696.1	*Clostridium saudiense*	98.15%
EC 5	WC0707	NR_148815.1	*Paraclostridium benzoelyticum*	99.74%
		NR_113323.1	*Paraclostridium bifermentans*	99.65%
	WC0708	NR_148815.1	*Paraclostridium benzoelyticum*	99.77%
		NR_113323.1	*Paraclostridium bifermentans*	99.65%
	WC0709	NR_113325.1	*Clostridium tertium*	99.27%

Table 4 Taxonomic attribution of the isolates. The taxonomy attribution was obtained by alignment of the V1-V4 regions of 16S rRNA gene with the NCBI refseq_rna database using BLAST. The accession number of best hits and the sequence identity (%) obtained with MUSCLE are reported.

4.2 Genomic and physiological characterization

4.2.1 Genome features

4.2.1.1 Revised taxonomy

Three isolates, WC 0700, WC 0705, and WC 0709 were selected to whole genome sequencing analysis. The taxonomic attribution made using SpeciesFinder has defined only the isolate WC 0709 to the *C. tertium* species but failed for the other two strains. A second approach of attribution, based on ANI value, was carried out using TypeMat tool of MiGA server and permitted to ascribe the WC 0700 and WC 0705 isolates to the species *C. celatum* and *P. bifermentans*, respectively. The 16S rRNA gene sequence of the isolate WC 0700 had an identity score of 97.02% with the strain *C. celatum* DSM 1785 (NR_026167.1), lower than the score obtained with *C. disporicum* DS1 (NR_026491.1), so the latter was the preliminary taxonomy attribution, that resulted wrong on the basis of genome sequencing.

To confirm the taxonomic assignment, digital DNA-DNA hybridization (dDDH) and average nucleotide identity (ANI) values were calculated comparing each isolate with the relative reference strain available (Tab. 5). All the isolates were over the thresholds for species demarcation, 70% and 95% for dDDH and ANI, respectively (Richter & Rossello, 2009).

Isolate	Reference strain	dDDH (%)	ANI (%)
WC 0700	*Clostridium celatum* DSM 1785	99.2	99.92
WC 0705	*Paraclostridium bifermentans* ATCC 638	89.6	98.94
WC 0709	*Clostridium tertium* DSM 2485	88.7	98.70

Table 5 Digital DNA-DNA hybridization (dDDH) and Average Nucleotide Identity (ANI) of the three isolates and their relative reference strain.

4.2.1.2 Genome features

Genome of *C. celatum* WC 0700, *P. bifermentans* WC 0705 and *C. tertium* WC 0709 were submitted to NCBI with the accession number: PRJNA781812, PRJNA781829, PRJNA737738.

The sequencing run of *C. tertium* WC 0709 produced 15.916.139 paired-end reads. The draft genome encompassed 42 contigs with the largest being 1.213.566 nt. The N_{50} length was 1.159.315 bp, while the L_{50} value 2. The assembled genome of *C. tertium* WC 0709 had a size of 3.89 Mbp, with a GC content of 27.72 % and 1230-fold coverage. A total of 3597 coding sequences (CDS) were annotated, comprising 6 rRNA genes (three 5S, two 16S, and one 23S) and 74 tRNAs.

For *C. celatum* WC 0700 the sequencing produced 6.577.366 paired-end reads 150-bp long. The assembled genome harboured 138 contigs. The N_{50} length was 70.096 bp, while the L_{50} count is 15. The strain had a genome of 3.63 Mbp with 27.75 % of GC content and 542-fold coverage. 3305 CDS were identified and included 3 rRNA genes (one of each subunit) and 62 tRNAs.

The genome of *P. bifermentans* WC 0705 had 75 contigs with the largest being 858.215 nt. The N50 value was 510.630 bp and L50 count was 3. The genome was 3.54 Mbp with a GC content of 28.02% and a 589-fold coverage. 3464 CDS were annotated, among which 8 rRNA (six 5S, one 16S and one 23S) and 46 tRNAs.

None of the three analysed genomes of this work possessed CRISPR-Cas system. Only in the genome of *C. celatum* WC 0700 a prophage sequence was identified. CARD revealed the presence of tetracycline resistance in *P. bifermentans* WC 0705, while KEGG detected β-lactam resistance in *C. tertium* WC 0709.

The presence of mobile elements, like genes putatively encoding insertion sequences (ISs) and transposases was investigated with ISFinder. Seven ISs were identified in *C. celatum* WC 0700, belonging to the families IS1595, IS6 and IS200/IS605, and presenting similarities with *Campylobacter jejuni* and *C. disporicum* (IS1595), *C. tetani* (IS6) and *C. perfringens* (IS200/IS605). *P. bifermentans* WC 0705 encompassed two ISs sequences of the families IS607 (*C. botulinum*) and IS1182 (*Bacillus cytotoxicus*). ISs of the families IS3 and IS200/IS605 were found in *C. tertium* WC 0709 and presented similarities with the counterparts of *C. beijerinckii* and *Staphylococcus epidermidis,* respectively.

Genes encoding the UviB bacteriocin were detected by BAGEL 4 in the genome of *P. bifermentans*. No match was found in *C. tertium* and *C. celatum* genomes, even if in both the genome was present UviB bacteriocin gene, with an identity of 41.00 % and 53.12%, respectively, with the counterpart UviB of *C. perfringens*. *C. celatum* genome harboured the gene *bcn5* encoding another bacteriocin, with an identity of 41% with the corresponding one of *C. perfringens*.

According to Prokka annotation, *C. tertium* WC 0709 has 41.03 % of ORF not annotated, *C. celatum* WC 0700 and *P. bifermentans* WC 0705 has 43.53% and 68.76 %, respectively.

4.2.1.3 Pangenome

To analyze the pangenomes of the species *C. celatum*, *P. bifermentans*, and *C. tertium*, all the other high-quality genomes belonging to these species available in GenBank were retrieved. General genomes information about the other available strains are reported in Tab. 6.

	Strain	Genome size (bp)	No. of scaffold	N50	L50	G+C (%)	No. of CDSs	No. of tRNA
C. celatum	**WC 0700**	**3.639.821**	**138**	**70.096**	**15**	**27.74**	**3305**	**62**
	D43t1_170807_C1	3.637.246	113	63.159	18	27.7	3157	75
	DSM 1785	3.553.721	122	69.892	15	27.7	3076	27
	MGYG-HGUT-01423	3.553.721	122	69.892	15	27.7	3076	27
P. bifermentans	**WC 0705**	**3.547.401**	**75**	**510.638**	**3**	**28.02**	**3464**	**46**
	SU1074NT	3.522.921	40	535.917	2	28.1	3360	43
	MHMC-14	3.552.671	63	208.933	5	28.2	3429	83
	parabai	3.916.009	29	530.207	3	28.43	3707	111
	ATCC 638	3.578.393	39	213.403	3	28.1	3461	24
	Med78_4-601-WT-2	3.508.699	31	351.239	4	28.1	3318	27
	BSD2780061688_150302_D6	3.623.244	26	732.858	2	28.1	3460	48
	BSD2780061688_150302_X7	3.629.130	60	119.504	10	28.1	3456	28
	S01	3.524.257	43	1.993.941	1	28.2	3306	98
	N73	3.615.727	41	418.592	3	28.2	3478	59
	DSM 14991	3.566.216	5			28.61	3354	102
	MGYG-HGUT-00024	3.643.390	25	2.249.377	1	28.2	3469	84
	ATCC 19299	3.535.241	58	153.781	9	28.1	3400	27
	PAGU 1678	3.471.060	126	82.995	13	28.1	3402	69
	WYM	3.475.995	180	41.803	26	28	3312	52
C. tertium	**WC0709**	**3.895.757**	**42**	**1.159.315**	**2**	**27.72**	**3525**	**74**
	Gcol.A2	3.897.924	60	281.242	6	27.8	3519	83
	Gcol.A43	3.801.844	55	178.423	6	27.8	3382	82
	BSD2780120874_150323_E10	3.796.527	88	96.757	13	27.8	3335	55
	BSD2780120875b_170604_A12	3.992.931	56	186.453	6	28	3534	66
	1001275B_160808_H2	3.851.958	74	107.425	11	27.8	3348	62
	DSM 2485	3.877.854	39	184.539	6	27.8	3549	76
	Clostridium tertium	3.970.462	49	258.765	5	27.8	3662	88
	MGYG-HGUT-01328	3.813.122	5	3.226.588	1	27.8	3428	74

Table 6 General genomes features. In bold the three strains isolated in this work.

The pangenome of *C. celatum* encompassed 3986 genes, 2883 in the core genome and 1103 in the accessory (Fig 14-panel A). Given the low number of available genomes of *C. celatum* it wasn't possible to calculate the α value to establish if it was a closed or open pangenome (Fig 14-panel B).
The pangenome of *P. bifermenta*ns encompassed 8398 orthologous genes, among which 2178 genes constituted the core genome (~26% of total genes). The accessory genome was composed by 2456 shell genes and 3764 cloud genes. In the *C. tertium* pangenome 2765 were the genes of core genome, while 1316 were the shell genes. Cloud genes, identified in a single strain (less than 15% of the strains), were 1975.
According to Heap's law, the *P. bifermentans* and *C. tertium* pangenomes were considered open, with a γ value of 0.41 and 0.57 respectively (Fig.14) (Tettelin et al., 2008).

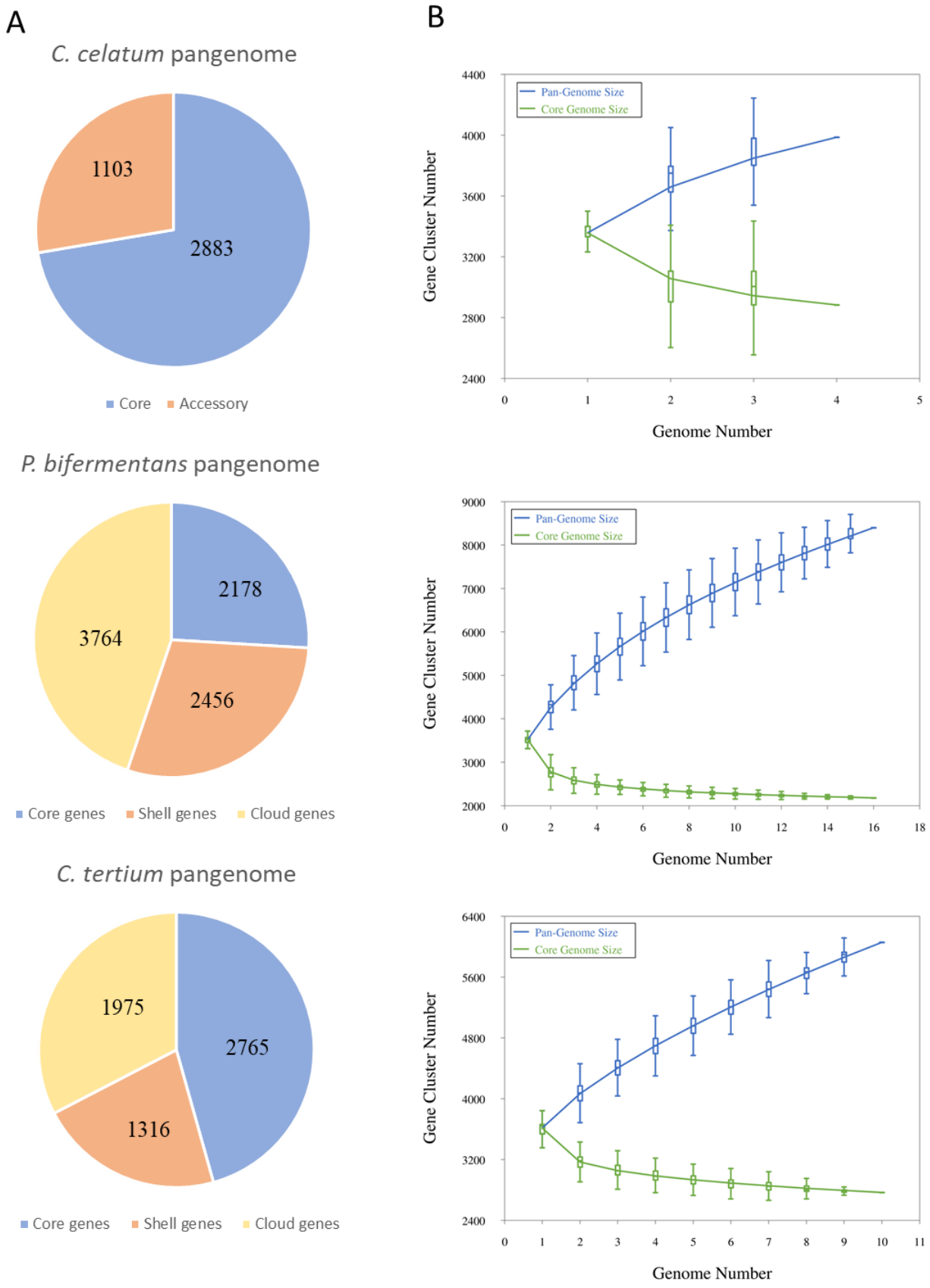


Figure 14 Pangenomes of the three species. (A) Distribution of genes found in each of the three pangenome: core (blue), shell genes (orange), cloud genes (yellow). (B) Estimation of the pangenome (blue) and the core genome (green) by including genomes one by one

4.2.1.4 Reconstructed metabolic pathway

Pathway reconstruction was carried out analyzing the genomes of *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709, utilizing KEGG gene annotation, in order to predict metabilic functions, including membrane transporters.

The three genomes shared seven complete phosphotransferase system (PTS) transporters of the families for glucose, GlcNAc, maltose, sucrose, cellobiose-diacetyl chitobiose, GalNAc, and fructose family. Other complete PTS systems were not shared by all the genomes: transporters of maltose/glucose family, α-glucoside, galactosamine, galactilol, and nitrogen families were found only in *C. tertium* WC 0709; transporters for β-glucoside and trehalose were predicted only in *C. celatum* WC 0700 and *C. tertium* WC 0709; a transporter for L-ascorbate was found in *C. tertium* WC 0709 and *P. bifermentans* WC 0705. Genes encoding PTS components of incomplete transporters were also identified, such as EIIC components with similarity to *agaC* and EIIBC component, that may partecipate in the uptake of other substrates in relation with others PTS components. Prokka gene annotation recognized genes encoding other complete PTS systems in the three genome examined, such as sorbose and mannitol families in *C. tertium* WC 0709, mannose family in *C. tertium* WC 0709 and *P. bifermentans* WC 0705, β-glucoside in *P. bifermentans* WC 0705.

	Phosphotransferase system (PTS)	*C. celatum* WC 0700	*P. bifermentans* WC 0705	*C. tertium* WC 0709
PtsI	PEP-protein phosphotransferase			
PtsH	PTS phosphocarrier protein HPr			
Glc family	glucose (*only EIICBA)		*	
Glc family	N-Acetyl-D-glucosamine			
Glc family	Maltose/ Glucose (* only EIIA component)	*	*	
Glc family	Maltose			
Glc family	Sucrose			
Glc family	β-glucoside			
Glc family	α-glucoside (* only EIIA component)	*		
Glc family	Trehalose			
Glc family	N-Acetyl-muramic acid (* only EIIA component)	*		*
Lac family	Cellobiose/ Diacetylchitobiose			
Man family	Manose (*only EIID component)	*		*
Man family	N-acetylgalactosamine			
Man family	Galactosamine (* only EIIA component)	*	*	
Man family	Sorbose			
Fru family	Fructose			
Fru family	Mannitol			
Gat family	Galactitol			
L-asc family	L-ascorbate			
	Nitrogen (* only EIIA)			*

Table 7 Predicted phosphotransferase system (PTS) of the strain *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709 according to KEGG gene annotation. In green the complete transport system, white the absent PTS, yellow and light yellow the incomplete systems lacking different component, and blue the PTS system manually checked with Prokka gene annotation.

The three genomes shared complete ABC transporters for the uptake of oligopeptides, basic amino acids, the biogenic amines spermidine and putrescine, choline (referred to as osmopretectant transporter), ribose and xylose, nucleosides, biotin, phosphate, molybdate, and sodium. Other ABC systems were predicted only in one or two genomes: complete transporters for oligosaccharides (raffinose, stachyose, and melbiose), methyl-galactoside, D-methionine, and zinc were found only in *C. tertium* WC 0709 and *C. celatum* WC 0700; tranporters for galctose oligomers and maltooligosaccharides, lysine, teichoic acid, and lantibiotics were found only in *C. tertium* WC 0709. Furthermore, all the three genomes presented several genes encoding ABC transporters with functions unknown or poorly characterized, including three ABC systems belonging to the eucaryotic-type subfamily and putatively involved in multidrug efflux system (EfrA/B, MdlA, BplA).

	C. celatum WC 0700	*P. bifermentans* WC 0705	*C. tertium* WC 0709
ABC transporters			
Mineral and organic ion transporter			
ABC transporters Prokaryotic type tungstate			
ABC transporters Prokaryotic type molybdate transport			
ABC transporters Prokaryotic type spermidine/putrescine transporter			
ABC transporters Prokaryotic type osmoprotectant transporter			
Oligosaccharide, polyol, and lipid transporter			
ABC transporters Prokaryotic type galactose oligomer/ maltooligosaccharide			
ABC transporters Prokaryotic type raffinose/stachyose/melibiose			
ABC transporters Prokaryotic type lactose/ L-arabinose			
ABC transporters Prokaryotic type Aldouronate			
ABC transporters Prokaryotic type N-Acetylglucosamine			
ABC transporters Prokaryotic type Nucleoside			
Monosaccharide transporter			
ABC transporters Prokaryotic type Ribose/ Autoinducer2/ D-xylose			
ABC transporters Prokaryotic type Methyl-galactoside			
Phospate and amico acid transporters			
ABC transporters Prokaryotic type Cystine			
ABC transporters Prokaryotic type Arginine/ Lysine/ Histidine			
ABC transporters Prokaryotic type Lysine			
ABC transporters Prokaryotic type D-methionine			
Peptide and nickel transporters			
ABC transporters Prokaryotic type Oligopeptide			
Metallic cation, iron-siderophore and vitamin B12 transporters			
ABC transporters Prokaryotic type Zinc			
ABC transporters Prokaryotic type Cobalt			
ABC transporters Prokaryotic type Nickel			
ABC transporters Prokaryotic type Biotin			
ABC-2 and other transporters			
ABC transporters Prokaryotic type Na+			
ABC transporters Prokaryotic type Teichoic acid			
ABC transporters Prokaryotic type Lantibiotics			
ABC transporters Eukaryotic type ABCB subfamily , EfrA/B (multidrug efflux)			
ABC transporters Eukaryotic type ABCB subfamily , MdlA (multidrug efflux)			
ABC transporters Eukaryotic type ABCC subfamily , BplA (bacteriocin export)			

Table 8 ABC transporters of the strain *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709 according to KEGG gene annotation. Green box indicates complete system, white box the missing system, light yellow incomplete systems.

Pathway reconstruction of central carbohydrate catabolism in *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709 yielded complete routes for Embden-Meyerhof glycolysis, the non-oxidative phase of pentose phosphate shunt (including PRPP synthesis), and pyruvate oxidation to acetyl-CoA. The alternative glycolytic route of Entner-Doudoroff and the tricarboxylic and glyoxylate cycles were extensively incomplete in all the strains. Leloir's pathway for galactose utilization was complete in all the genomes, on the other hand the pathways for the degradation of ascorbate, glucuronate, and galacturonate were incomplete or missing.

The whole route for the production of UDP-glucose, UDP-galactose and UDP-GlcNAc were recognized in all the genomes. In addiction, glycogen biosynthesis and degradation occurred in *C.*

tertium WC 0709 and *C. celatum* WC 0700, but did not in *P. bifermentans* WC 0705 that lacked 4-α-glucotransferase (malQ) or pullulanase (pulA) in the degradation pathway. The three genomes harbored the whole set or most of the genes for the biosynthesis of nucleotides and deoxy-ribonucleotides of purines and pyrimidines. Some incompletness in these routes were found only in *P. bifermentans* WC 0705, and may be attributed to deficiency in KEGG annotation of the genes.

With regards to the biosynthesis of aminoacids, pathway reconstraction with KEGG annotation indicated that the anabolic potential of the three bacteria was largely incomplete. Therefore it should be further investigated whether pathways incompletness arised from defects in gene annotation or whether it could actually lead to nutritional auxotrophies. The three genomes hardbored all the genes necessary for the biosynthesis of lysine, through the succinyl-DAP pathway (*C. celatum* and *C. tertium*) or the DAP aminotransferase pathway (*P. bifermentans*). The metabolic route branching from lysin biosynthesis and yielding homoserine and then threonine was complete only in *C. celatum* WC 0700 and *C. tertium* WC0709, while the one transforming homoserine into methinine was incomplete in all the three genomes. The pathway of *de novo* biosynthesis of serin was interrupted in all the genomes, that however were all equipped with the genes necessary to transform preformed sei into cystein. The pathways yielding valine, leucine, and isoleucine from pyruvate were complete in *C. celatum* WC 0700 and *C. tertium* WC0709, while any gene for the *de novo* biosynthesis of branched chain amino acids were not identified in *P. bifermentans* WC 0705. Likewise, the pathways for the transformation of glutamate into proline and into arginine via ornithine were complete in *C. celatum* WC 0700 and *C. tertium* WC0709, but were incomplete or missing in *P. bifermentans* WC 0705. A complete shikimate pathway was predicted in all the three genomes, however the routes of transformation of chorismate into phenylalanine and tyrosine was always incomplete and the one leading to tryptophan was always missing. A complete pathway for histidine degradation was identified only in *P. bifermentans* WC 0705. Genes involved in the interconversions between glutamate, ornithine, arginine, spermidine, and putrescine, were identified in the three genomes, however the metabolic modules channeling them to GABA, and finally toward succinyl-CoA for degradation were interrupted.

As regards the biosynthesis of vitamins and cofactors, the pathways were for the most part incomplete, with only a few exceptions. *De novo* cobalamin production is the result of three consecutive metabolic modules, the transformation of glutamate into siroheme, the transformation of siroheme into cobyrinate a,c-diamide, and the final production of cobalamin, consisting of 6, 11, and 7 enzymatic reactions. *P. bifermentans* WC 0705 harbored all the three modules, lacking only one enzyme in siroheme production module. In *C. celatum* WC 0700 and *C. tertium* WC0709, the central module was greatly incomplete, despite the first one and the last one were complete. CoA production

from panthotenate was predicted in all the genomes. *C. tertium* WC0709 harbored the pathway for pyridoxal, while *C. celatum* WC 0700 and *P. bifermentans* WC 0705 had pathway for thiamine mono phosphate salvage (TMP).

4.2.1.5 Mucin utilization

The analysis of CAZymes using the dbCAN2 metaserver revealed a total of 162 putative ORFs encoding proteins involved in metabolism, and transport of carbohydrates in the genome of *C. celatum* WC 0700, 95 ORFs in *P. bifermentans* WC 0705, and 200 ORFs in *C. tertium* WC 0709. Only the CAZymes annotated by two out of the three tools (HMMER, Hotpep and Diamond) used by dbCAN2 were considered for the comparative analysis (Tab. 9).

	C. celatum **WC 0700**		***P. bifermentans*** **WC 0705**		***C. tertium*** **WC 0709**	
Cazy class	# of identified genes	# of genes with signal peptide	# of identified genes	# of genes with signal peptide	# of identified genes	# of genes with signal peptide
Glycoside Hydrolase (GH)	64	28	24	1	78	15
Carbohydrate esterase (CE)	9	2	9	4	10	4
Glycosyl transferase (GT)	22	0	16	0	30	0
Carbohydrate-binding module (CBM)	27	19	3	1	16	9
Auxiliary activity (AA)	0	0	0	0	0	0
Polysaccharide lyase (PL)	0	0	0	0	4	0

Table 9 CAZymes identified in the three genomes with at least two out the three dbCAN tools.

The comparative genomic analysis (Fig. 15) identified in the *C. celatum* strains (Tab. 1) at least 25 GHs involved in mucin degradation, mostly containing a signal peptide. In particular, there were 2 sialidases (GH33), 3 fucosidases (GH95), 5 galactosidases (GH2) and 10 hexosaminidases (GH20/84/89). Among this last group, 5 to 7 GHs were endo-α-*N*-acetylgalactosaminidases (GH101) that catalyze the hydrolysis of core 1 Galβ1-3GlcNAc from mucin protein backbone (Josenhans et al., 2020).

In *C. tertium* there were at least nine GHs that target mucins *O*-glycans: a sialidase (GH33), 3 or 4 different galactosidases of the family GH2, and in some cases of the families GH35 and GH42.

The *C. tertium* strains presented a variable number of hexosaminidases including one endo-α-*N*-acetylgalactosaminidase (GH101).

P. bifermentans had the smallest set of GHs potentially involved in mucin degradation, with only one galactosidase (GH2), one hexosaminidase GH20 or in some cases one GH85.

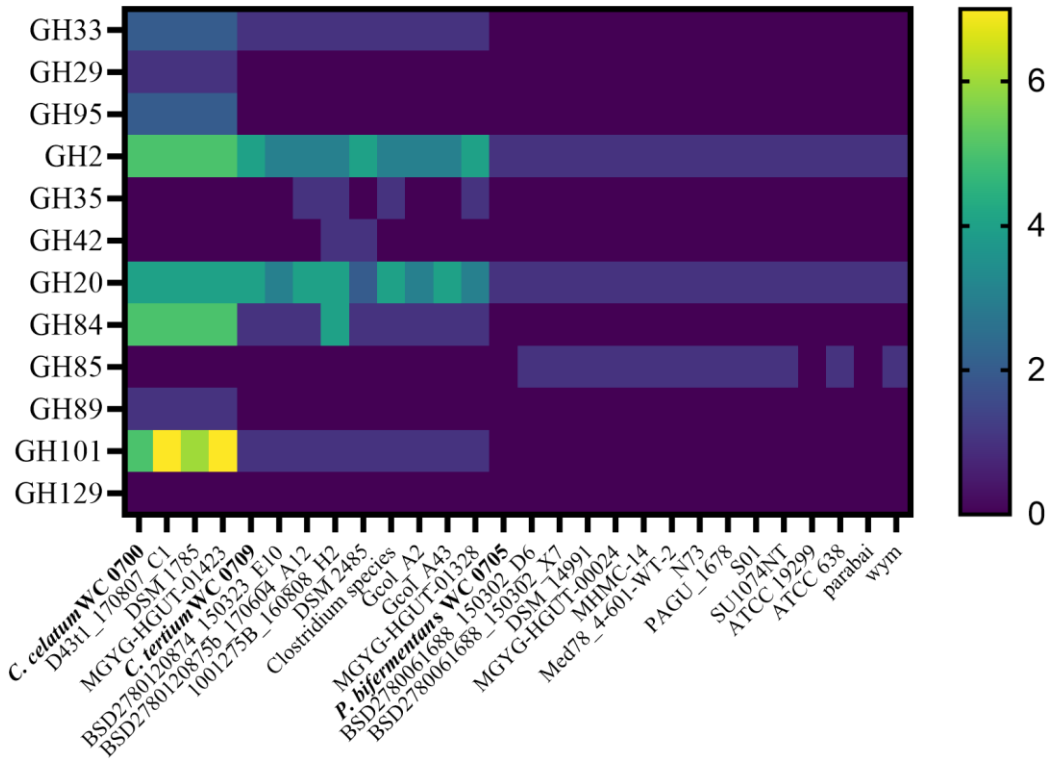


Figure 15 Heat map showing the gene copy number of each GH in all the examined genomes. Sialidase (GH33), fucosidases (GH29, GH95), galactosidases (GH2, GH35, GH42), hexosaminidases (GH20, GH84, GH85, GH89) and α-*N*-acetylgalactosaminidases (GH101, GH129).

To better understand if *C. celatum* WC 0700, *P. bifermentans* WC 0705 and *C. tertium* WC 0709 were able not only to cleave mucin glycans but also to catabolize the corresponding monosaccharides, a comparative approach of metabolic pathway was carried out based on what was predicted by Ravcheev & Thiele (2017). (Fig 16)

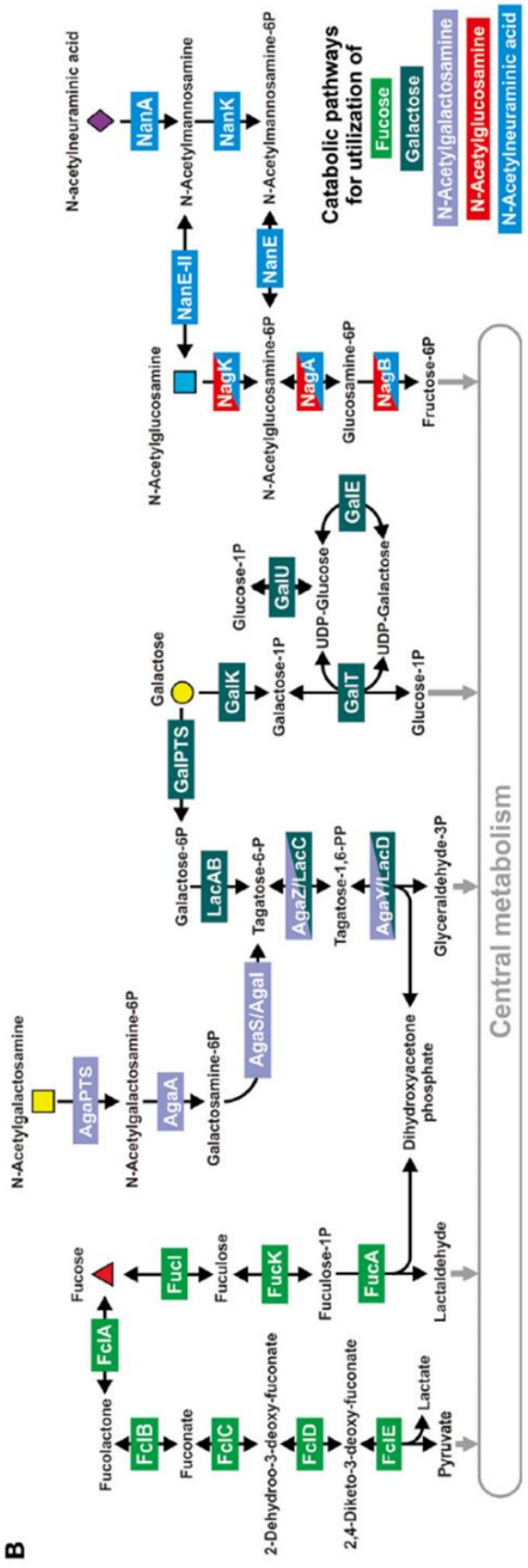


Figure 16 Predicted metabolic pathways for the degradation and utilization of the single monosaccharides from mucin's structure (Ravcheev & Thiele, 2017)**.**

According to genome analysis, the distribution of genes required for utilization of galactose, *N*-acetylgalactosamine and ***N***-acetylglucosamine were widespread among the three strains *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705. In particular, it was predicted that galactose was utilized only by the Leloir pathway, which included the expression of *galK, galU, galT* and *galE* genes and use Gal through UDP galactose (Afzal et al., 2015). None of the strains examined were able to use Gal through the tagatose-6-phosphate pathway, lacking the galactose-6P isomerase (LacAB) that is the first enzyme of the metabolic chain.

Among the three putative different transport systems for *N*-acetylgalactosamine (Ravcheev & Thiele, 2017), only the AgaPTS system was found in the three genomes, that imports GalNAc-6-P into the cells. The metabolic pathway involves *N*-acetylgalactosamine-6-phosphate deacetylase encoded by *agaA*, galactosamine-6-phosphate isomerase/deaminase (*agaS*), tagatose-6P-kinase (*agaZ*), and tagatose-1,6-biphosphate aldolase (*agaY*), and results in the production of glyceraldehyde-3P. Prokka annotated genomes lacked of *agaA* gene but all the three strains they had one or two *nagA* (*N*-acetylglucosamine-6-phosphate deacetylase) genes. In the genome of *C. celatum* DSM 1785 both the genes *agaA* and *nagA* were previously identified (Ravcheev & Thiele, 2017). The annotated *agaA* sequence was used to make a blast search in our genomes. *C. celatum* WC 0700 harboured a gene, annotated by Prokka as nagA_2, presenting an identity of 99%, while *C. tertium* WC 0709 had an identity of 83,48% with the Prokka annotated nagA_1 (accession number: MBU6134103.1). No match was found in *P. bifermentans* WC 0705 that had only one *nagA* gene. Taking into account the predicted availability of all the enzyme of the metabolic pathway and the high number of hypotetical proteins, it is possible that, like in *E.coli*, the enzyme encoded by *nagA* funfill the function of *agaA* (Hu et al., 2013).

Different GlcNAc-specific transport systems were identified: PTS NagE system was only found in *C. tertium* WC 0709 that possess, togheter with *C. celatum* WC 0700, also the ABC transport system NgcEFG. NagP permease was identified in *C. celatum* WC 0700 and *P. bifermentans* WC 0705. The other enzymes of the *N*-acetylglucosamine utilization pathway including *nagK*, *nagA* and *nagB* were detected in all the strain examined.

L-Fucose utilization pathway was present only in *C.celatum* WC 0700, that harboured the predicted ABC-type fucose transport system. This monosaccharide is the substrate of fucose isomerase (*fucI*), fuculokinase (*fucK*) and fuculose phosphate aldolase (*fucA*) genes that produce lactaldehyde at the end of the metabolic chain.

Even if sialidase GH33 was observed in *C.celatum* and *C. tertium* strains, only the latter expressed all the enzymes involved in *N*-acetylneuraminic acid utilization. Two different pathway discernable by the sugar epimerase have been described previously and both involve the *nanT*

permease and *nanA* aldolase (Brigham et al., 2009). In *C. tertium* WC 0709 were identified the *nanK* kinase (annotated by Prokka as bglK in *C. tertium* WC 0709) and the *nanE* epimerase genes that convert *N*-acetylmannosamine-6P to *N*-acetylglucosamine-6P.

To verify the predicted ability of *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 to ferment mucin *O*-glycan monomers, the strains were cultured in a medium supplemented with the single monosaccharide composing mucin: L-fucose, *N*-acetylneuraminic acid, galactose, *N*-acetyl-D-galactosamine, and *N*-acetyl-D-glucosamine. Medium supplemented of glucose was used as positive control.

C. celatum WC 0700 confirmed the predicted metabolic capabilities, growing in M17 supplemented with L-fucose, galactose, and GlcNAc. As expected, C. *tertium* WC 0709 grew on *N*-Acetylneuraminic acid, galactose, and GlcNAc. Interestingly, the strain was also able to use fucose even if the corresponding pathway was not predicted from genome annotation and pathway reconstruction. In *C. celatum* WC 0700 and *C. tertium* WC 0709, growth in presence of GalNAc did not occur, suggesting that the nagA proteins identified in their genome do not have the *N*-acetylgalactosamine-6-phosphate deacetylase function, or are not expressed/functional.

On the other side, *P. bifermentans* WC 0705 grew only in the medium supplemented with GalNAc.

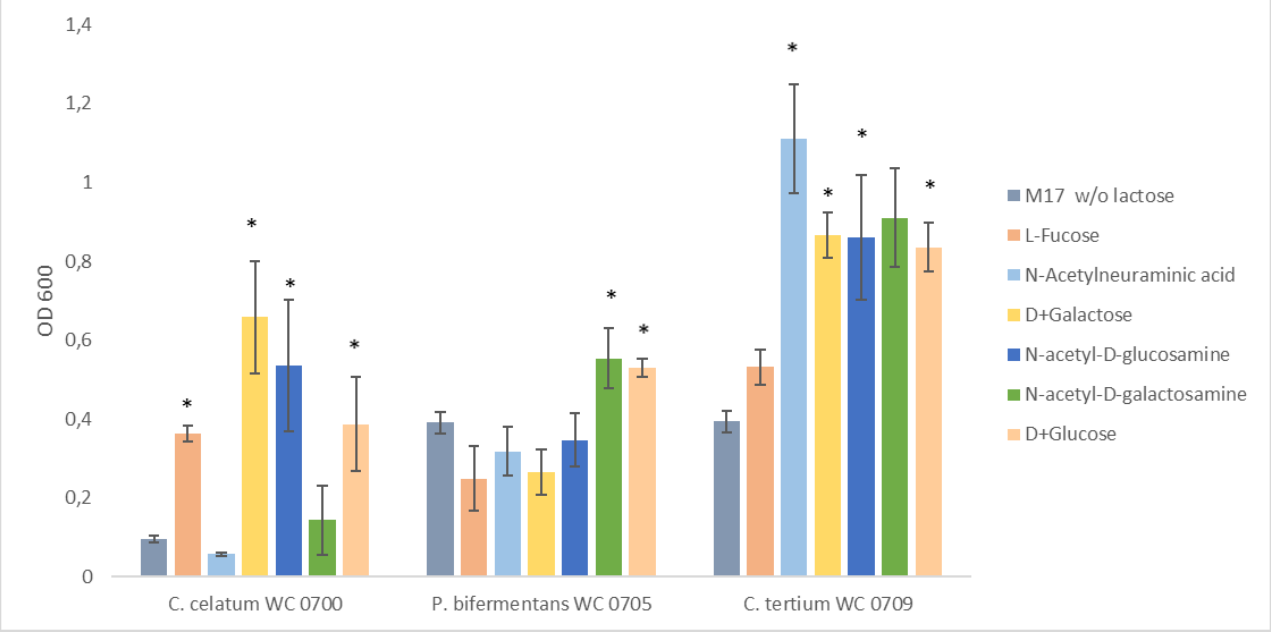


Figure 17 Growth of the three strains on mucin *O*-glycan monosaccharides. Bacteria were grown in medium supplemented with individually monosaccharides (20 mM). Graphs show the mean value ± SD acquired from three independent experiments. (t-test, $P < 0.05$)

4.2.2 Physiological features

4.2.2.1 Fermentation at various pH condition

Fermentation products of *C. celatum* WC 0700, *P. bifermentans* WC 0705 and *C. tertium* WC 0709 were analysed after 24, 48, and 72 h of incubation in MM set at different initial pH. For all the tested strains, under all the pH conditions, the maximum biomass yield was generally reached after 24 h (Fig. 18). On the other side, the highest biomass production was obtained at pH values close to neutrality, the optimal range being 6.5–7.5 for *C. celatum* WC 0700, and narrower for *C. tertium* WC 0709 (6.5–7.0) and *P. bifermentans* WC 0705 (7.0). Sugar fermentation generated organic acids (formic, acetic, propionic, butyric, valeric, isovaleric, and succinic acids), 2,3 butandiol, and ethanol. The abundance of each fermentation product depended on the strain and the pH, with acetic acid being generally the main metabolite generated by the strains in most of the conditions.

At the optimal pH (7.0), *C. celatum* WC 0700 produced acetic, formic, and propionic acids during growth, that reached 0.5, 0.2, and 0.1 g/L after 48 h, respectively. Growth was accompanied also by the generation of isovaleric and lactic acids, both peaking at 24 h (0.2 and 0.1 g/L respectively), and slightly decreasing towards the end of the fermentation, and by minor amounts of the other metabolites (always < 0.1 g/L).

At pH 7.0, the main products of *P. bifermentans* WC 0705 were, in decreasing order, acetic, propionic, and succinic acids, that reached 0.25, 0.15, and 0.08 g/L, respectively at 48 h. Isovaleric acid reached 0.15 g/L after 24 h, then slightly decreased, while all the other metabolites always remained below 0.05 g/L.

In the cultures of *C. tertium* WC 0709, acetic acid was the most abundant fermentation product, with 0.3 g/L after 48 h. Propionic and isovaleric acids were also abundantly produced in the first 24 h (0.2 and 0.1 g/L, respectively) but decreased toward the end of the fermentation. Formic acid reached 0.05 g/L at 24 h, while all the other metabolites always remained at lower concentration.

Analysis of the headspace by a Gas-Analyzer detected in all the processes CO_2 and H_2 as volatile fermentation products.

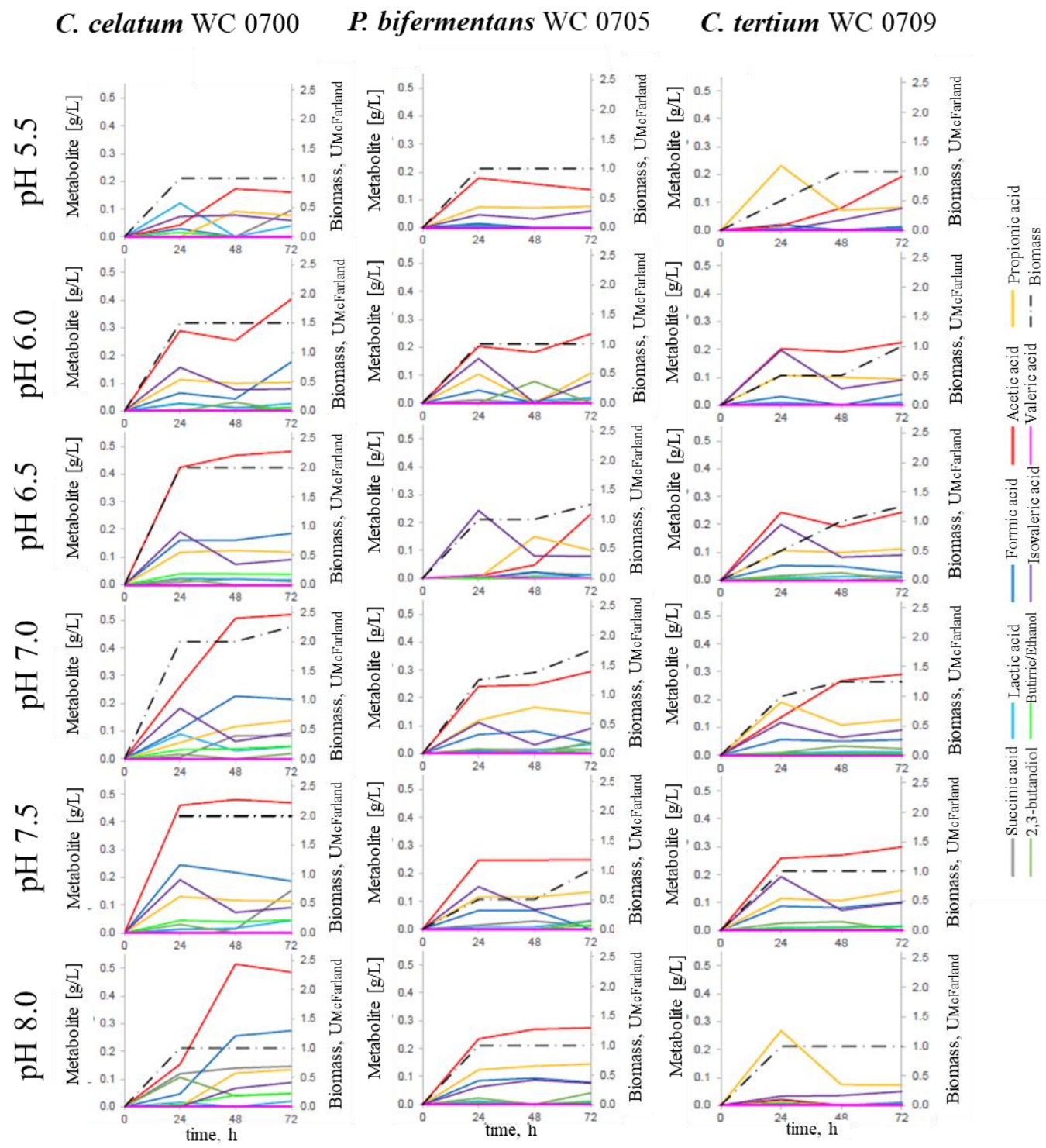


Figure 18 Time-course of fermentation products. Production of metabolite after 24, 48 and 72 h of incubation in MM broth.

4.2.2.2 Biochemical features

Metabolic capabilities of *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 were assessed using API 50 CH. The three strains were able to ferment D-ribose, D-galactose, D-glucose, D-fructose, D-mannose, D-cellobiose, D-maltose, D-lactose, and esculin. Only *C. tertium* WC 0709 and *P. bifermentans* WC 0705 were able to utilize glycerol, D-xylose, D-mannitol, and D-sorbitol. *C. tertium* WC 0709 and *C. celatum* WC 0700 also fermented amygdalin, D-saccharose, D-melibiose, gentiobiose. *C. tertium* WC 0709 exhibited the widest range of substrates, fermenting also D-xylose, L-sorbose, N-acetylglucosamine, arbutin, salicin, D-melezitose, and D-tagatose. *P. bifermentans* WC 0705 was the sole strain growing on starch, xylitol, gluconate, and 2-ketogluconate, while only *C. celatum* WC 0700 fermented D-trehalose.

On the other hand, the three strain were not able to utilize 21 substrates, among which erythritol, D- and L-arabinose, L-xylose, methyl-β-D-xylopyranoside, dulcitol, inositol, methyl-α-D-mannopyranoside, methyl-α-D-glucopyranoside, inulin, D-raffinose, glycogen, D-turanose, D-lyxose, D- and L-fucose, D- ad L-arabitol, and 5-ketogluconate.

	Control	Glycerol	Erythritol	D-arabinose.	L-arabinose	D-ribose	D-xylose	L-xylose	D-adonitol	Methyl-β-D-x ylopyranoside	D-galactose	D-glucose	D-fructose	D-mannose	L-sorbose	L-rhamnose	Dulcitol	Inositol	D-mannitol	D-sorbitol	Methyl-α-D-mannopyranoside	Methyl-α-D-glucopyranoside	N-acetylglucosamine	Amygdalin	Arbutin
C. celatum WC 0700	-	-	-	-	-	+	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	+	-
C. tertium WC 0709	-	+	-	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	+	+	-	-	+	+	+
P. bifermentans WC 0705	-	+	-	-	-	+	-	-	+	-	+	+	+	+	-	-	-	-	+	+	-	-	-	-	-

	Esculin ferric citrate	Salicin	D-cellobiose	D-maltose	D-lactose	D-melibiose	D-saccharose	D-trehalose	Inulin	D-melezitose	D-raffinose	starch	Glycogen	Xylitol	Gentiobiose	D-turanose	D-lyxose	D-tagatose	D-fucose	L-fucose	D-arabitol	L-arabitol	Potassium gluconate	Potassium 2-ketogluconate	Potassium 5-ketogluconate
C. celatum WC 0700	+	-	+	+	+	+	+	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
C. tertium WC 0709	+	+	+	+	+	+	+	-	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-
P. bifermentans WC 0705	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-

Table 10 API 50 CH tests performed on *C. celatum* WC 0700, *C. tertium* WC 0709, *and P. bifermentans* WC 0705.

4.2.2.3 Biofilm formation

The three strains were tested for biofilm formation in presence and absence of mucin (Fig. 19). The SBF of *C. tertium* WC 0709 was under the threshold of 1, suggesting the inability to form biofilm in both conditions. Compared to the positive control *E. coli* 03.73, slight biofilm occurred in *C. celatum* WC 0700 and *P. bifermentans* WC 0705, with scores of 2.0 and 2.8, respectively. No statistically difference (paired samples t-test, $P<0.05$) was observed in presence of mucin coating.

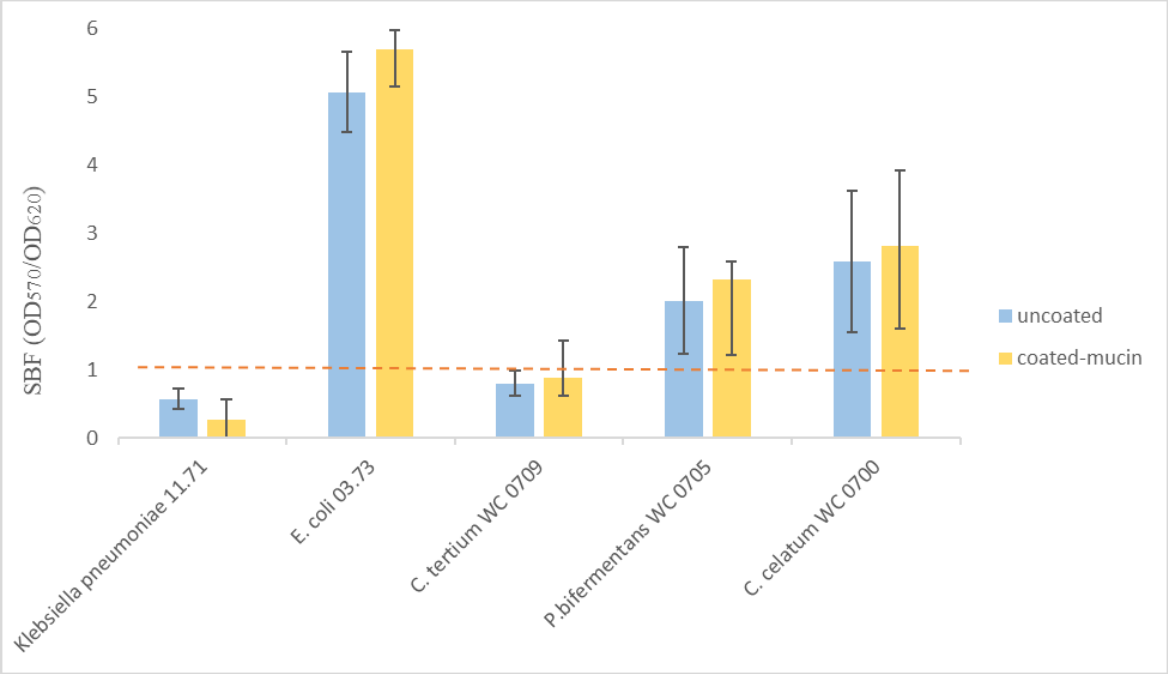


Figure 19 Biofilm formation in coated and un-coated mucin wells. The SBF is calculated as the ratio between the crystal violet absorbance at 570 nm and the culture turbidity at 620 nm, setting a threshold of 1 (red dashed line). The reported data are means ± standard deviations of at least three independent experiments, each carried out in triplicate.

4.2.2.4 Tolerance to Oxygen and Temperature

Like other members of Clostridia class, the three isolates of this study *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709 produce spore that allow them to resist to stress conditions. Survival under oxygen and high temperature exposure was assessed. Estimation of microbial population size was made with MPN method after the following treatments: (I) exposure to 80°C for 30 minutes, (II) exposure to oxygen for an hour, and (III) exposure to oxygen for an hour

followed by heat treatment at 80°C for 30 minutes. Anaerobic cultures were used as controls (untreated).

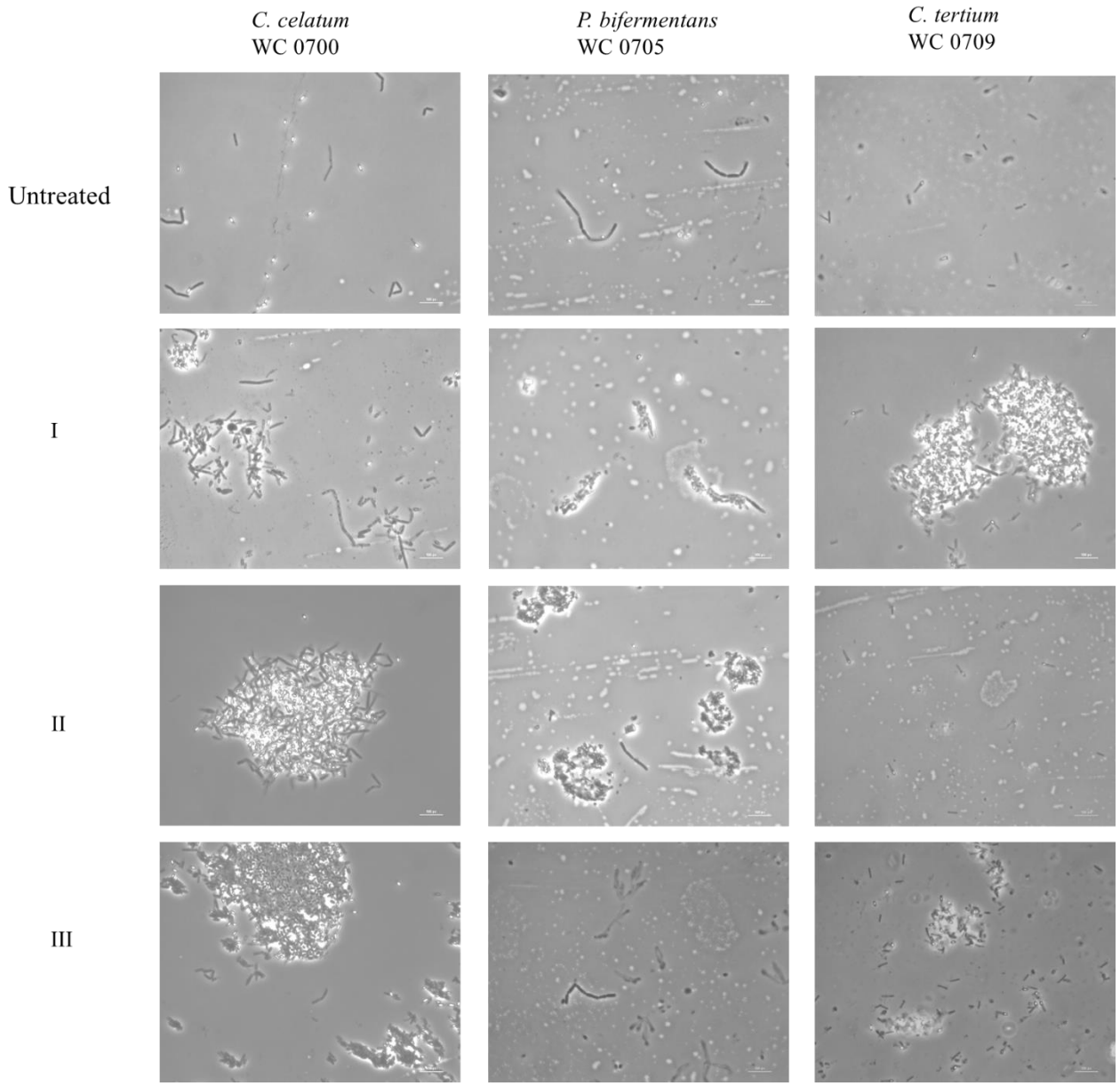


Figure 20 Survival under oxygen and high temperature exposure. Estimation of microbial population after the following treatments: (I) exposure to 80°C for 30 minutes, (II) exposure to oxygen for an hour and (III) exposure to oxygen for an hour followed by heat treatment at 80°C for 30 minutes. Untreated: control cultures grown anaerobically at 37°C.

C. celatum WC 0700 was more sensitive to both oxygen and temperature stress, with a decrease of one Log_{10} (untreated). Oxygen exposure (II) did not affect microbial size of *C. tertium* WC 0709 and *P. bifermentans* WC 0705, according to MPN enumeration. Heat treatment at 80 °C (I and III) caused a decrease of one Log_{10} in *C. tertium* WC 0709.

4.2.2.5 Antimicrobial susceptibility test

Genetic determinants for antibiotic resistances were searched in the genome of *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709. The CARD analysis revealed the presence of tetracycline resistance in *P. bifermentans* WC 0705, while KEGG detected β-lactam resistance in *C. tertium* WC 0709. To validate genomic data, broth microdilution method for MIC determination were performed.

C. tertium WC 0709 resulted resistant to ampicillin (Tab. 11), with a MIC breakpoint of 8, and only intermediate to penicillin according to CLSI breakpoints (M100-ED31:2021). All the strains were susceptible to the chloramphenicol, with breakpoints of 2 in both *C. celatum* WC 0700 and *C. tertium* WC 0709, or 4 mg/L in the case of *P. bifermentans* WC 0705.

Both, *C. tertium* WC 0709 and *P. bifermentans* WC 0705 were resistant to tetracycline with breakpoints of 16. *C. celatum* WC 0700 was susceptible to all the antimicrobials tested. Interestingly, all the strains exhibited resistance to high concentration of gentamicin referring to a breakpoint of 16 indicated in the work of Wei et al. (2020), according to the CLSI-M100-S23:2019 database.

Antimicrobial	**MIC range (mg/L)**	**Reference breakpoints**			***C. celatum* breakpoints**	***P. bifermentans* breakpoints**	***C. tertium* breakpoints**
		S	I	R			
Ampicillin	64-0.06	≤4	-	>8	1	0.125	8
Gentamicin	64-0.06	-	-	-	32	32	>64
Chloramphenicol	64-0.06	≤8	-	>8	2	4	2
Penicillin G	16-0.01	≤0.5	1	≥2	0,25	0.03	1
Tetracycline	16-0.01	≤4	8	≥16	0,06	16	16

Table 11 Antimicrobial susceptibility of the three isolates. Ampicillin and chloramphenicol breakpoints were defined according to EUCAST (https://www.eucast.org/clinical breakpoints/ version 11.0) for Gram-positive anaerobes, while penicillin G and tetracycline breakpoint were defined according to Clinical and Laboratory Standards Institute (CLSI-https://clsi.org/meetings/microbiology/ M100-Ed31:2021). S: susceptible; I: intermediate; R: resistant

4.2.3 *In vivo* effect on the immune system

To evaluate the impact of *C. tertium* WC 0709 and *P. bifermentans* WC 0705 on the immune system, the phenotype of dendritic cells (DC) and lymphocytes in Peyer's patches was analyzed. 10 days of oral supplementation with either *C. tertium* WC 0709 or *P. bifermentans* WC 0705 produced quite different effects on the mucosal immune system.

Albeit the significant decrease of the percentage of $CD3^+CD8^+$ lymphocytes observed after the treatment with *C. tertium* WC0709, the ratio of cells producing IFNγ and IL-17 increased (Fig. 21 panel B), suggesting a higher activation of these cells. No changes on $CD3^+CD8^+$ lymphocytes population were recorded after the supplementation with *P. bifermentans* WC 0705, no either a diverse level of cytokines expressed.

After the treatment with *C. tertium* WC 0709 or *P. bifermentans* WC 0705, the level of $CD3^+CD4^+$ lymphocytes did not change (Fig. 21 panel A). Since IFN-γ/IL-4 ratio was higher in mice after the *C. tertium* WC 0709 intervention compared to controls, it seemed that *C. tertium* WC 0709 mediated the cell-mediated Th1 immune response.

Only the supplementation with *P. bifermentans* WC 0705 exerted a significant effect on Treg cells, significantly enhancing the percentage in the gut lamina propria (Fig. 22 panel B), and the expression of $FoxP3^+$.

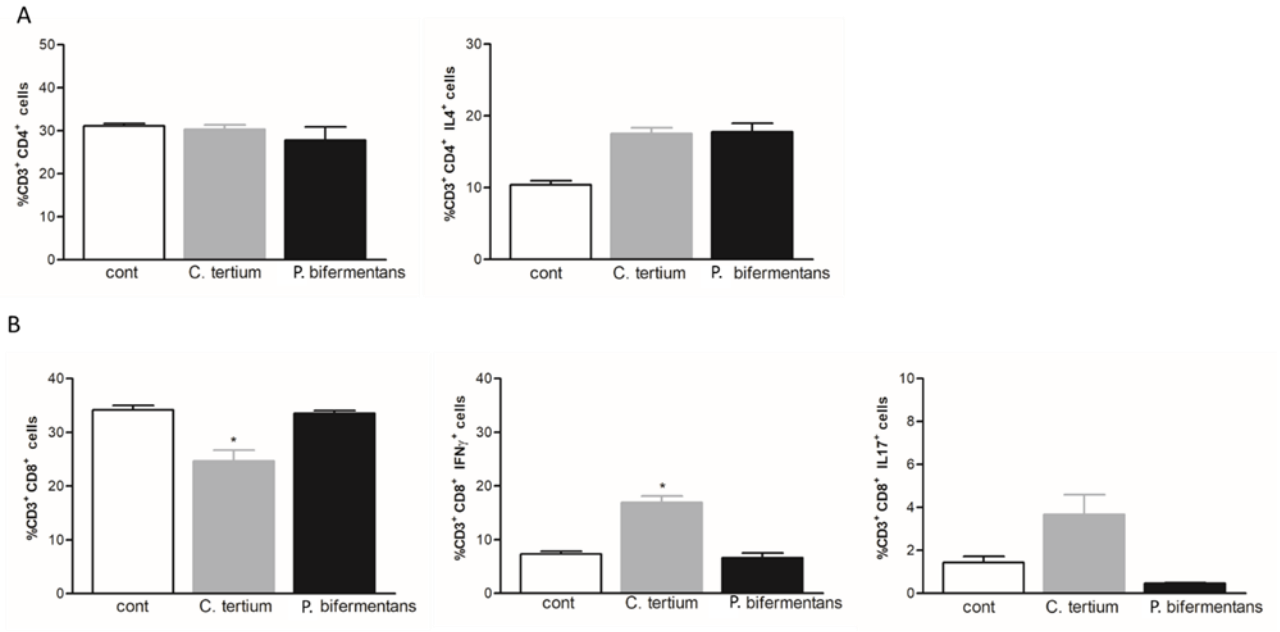


Figure 21 Effect of daily oral supplementation with 10^8 CFU of *C. tertium* WC0709, and *P. bifermentans* WC0705 on intestinal mucosa immune cells, determined by FACS. CD3 $^+$ lymphocytes. * Indicates significant difference versus the control (t-test, P< 0.05).

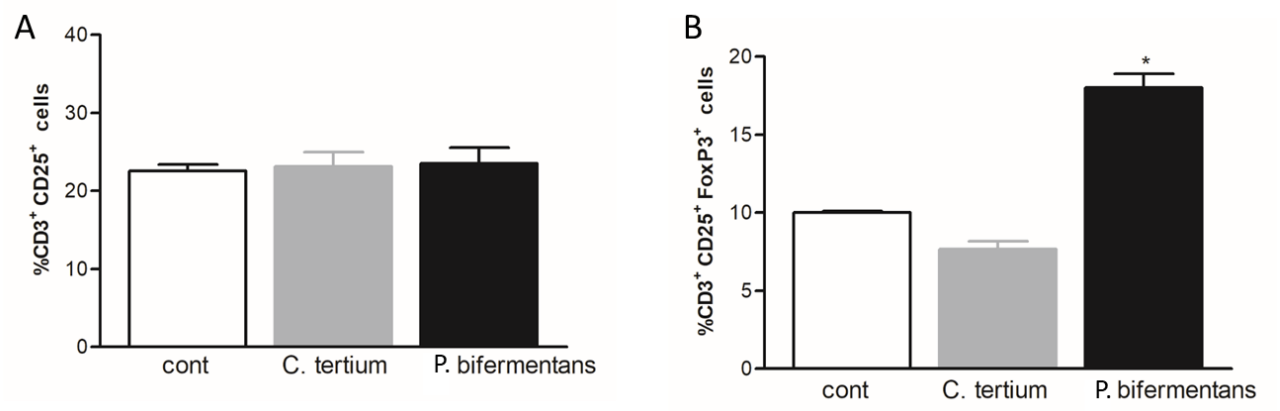


Figure 22 Effect of daily oral supplementation with 10^8 CFU of *C. tertium* WC0709, and *P. bifermentans* WC0705 on intestinal mucosa immune cells, determined by FACS. CD3 $^+$CD25$^+$ cells * Indicates significant difference versus the control (t-test, P< 0.05).

Both *C. tertium* WC 0709 and *P. bifermentans* WC 0705 determined a significant decrease of the markers of DC maturation CD80 and TLR2, with major effect exerted by *P. bifermentans* WC 0705. On the other side, only *P. bifermentans* WC 0705 drastically reduced the expression level of TLR4 (Fig. 23).

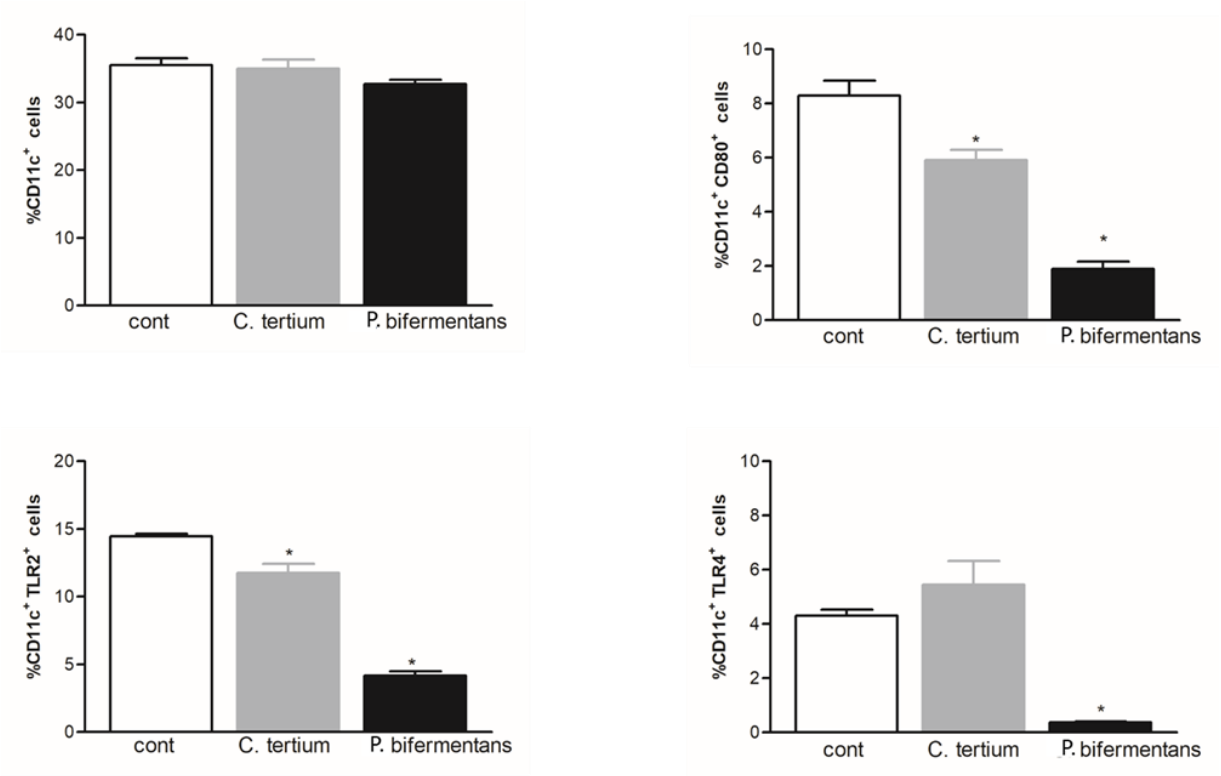


Figure 23 Effect of daily oral supplementation with 10^8 CFU of *C. tertium* WC0709, and *P. bifermentans* WC0705 on intestinal mucosa immune cells, determined by FACS. dendritic cells $CD11^+$. * Indicates significant difference versus the control (t-test, $P < 0.05$).

4.3 Human gut Clostridiales

The aim of this second part of this project was to investigate the population of Clostridiales detected in the gut microbiota, using 60 metagenomes of intestinal microbiota from healthy subjects of five different countries, to determine the overwhelming and the most recurrent species. Once defined, reference genomes of the Clostridiales species pinpointed in the gut metagenomes have been compared, to assess the phylogenetic relationships and to verify the role of main lineages within Clostridiales in the gut microbiota function and composition.

4.3.1 Who and how many

The order of Clostridiales encompasses 26 families, and 281 genera, according to NCBI database (http://www.bacterio.net). The species of Clostridiales present in the gut microbiota were searched in two different datasets: 553 samples of the HMP (Human Microbiome Project- Turnbaugh et al., 2007), and the set of 60 intestinal metagenome of healthy subjects from five geographically different cohorts assembled by Candeliere et al. (2022). As a whole, 159 species of Clostridiales were pinpointed in the gut microbiota. Albeit the higher number of metagenomes of the former dataset, all the species herein identified (74) were present also in some metagenomes of the latter group.

In terms of relative abundance, in the 60 metagenomes Firmicutes accounted for 46.55% of total bacteria, ranging from 10.42 to 85.42 % (Fig. 24). Most of them were Clostridiales (42.09%), and in sole two samples Clostridiales were outnumbered by other Firmicutes (Fig. 24).

Among the 159 species detected, 126 had reference genomes in the NCBI database, whereas the genomes of the type strains *Eubacterium rectale*, *Eubacterium siraeum*, and *Clostridium leptum* were added to the defined subset to include these relevant species. Furthermore, also the reference genome of *C. tertium* was added to the subset.

33.45% of the abundance was referred to the 130 species detected at least once in one of the two datasets (Fig. 25). Within the 8.64% of Clostridiales that did not reach the species level, relevant was the weight of *Sudoligranulum* unclassified (5.05%). The remainder 3.59% was ascribed to 99 other OTUs

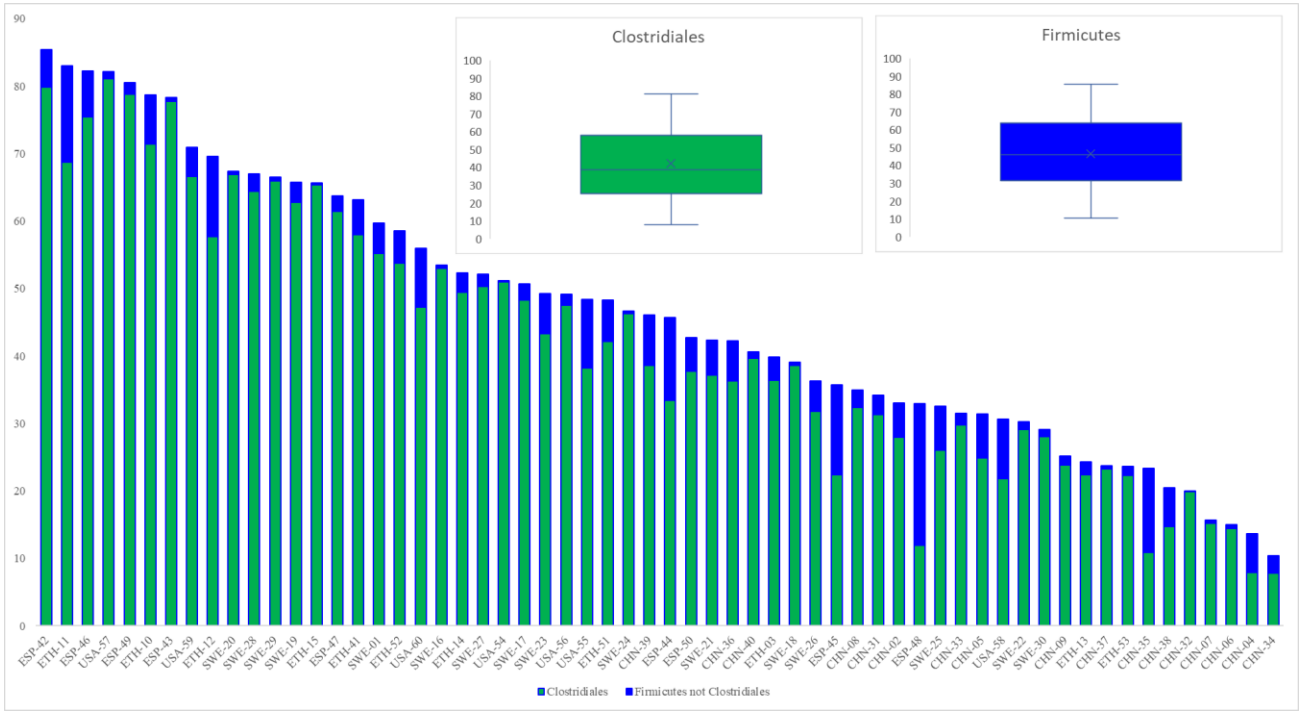


Figure 24 Distribution of Firmicutes and Clostridiales among the 60 analysed metagenomes. Abundance of each is displayed in the box plot.

The contribution of single species to the relative abundance revealed that *Faecalibacterium prausnitzii* was the most abundant one (10.99% of total bacteria), followed *by Eubacterium rectale* (7.18%), *Ruminococcus bromii* (2.35%), *Roseburia intestinalis*, *E. siraeum*, *R. torques*, *R. inulinivorans*, *Butyrivibrio crossotus*, *E. eligens*, *Dorea longicatena*, *Anearobutyricum halii*, *Blautia obeum*, *Coprococcus eutactus*, *C. comes* (1.18-0.51%) (Fig. 25). Only *F. prausnitzii*, *R. torques*, and *R. obeum* were present in the whole set of 60 samples, and in 537, 534, and 510 out of the 553 samples of the HMP, respectively. Other recurrent species not claimed as the most abundant were *D. formicigenerans* (57/60; 0.41%), *R. hominis* (55/60; 0.50%), *C. catus* (55/60; 0.29%), *E. ramulus* (52/60; 014%), *Anaerostipes hadrus* (52/60; 0.07%), and *E. ventriosum* (51/60; 0.52%). On the other side, the species *B. crossotus* and *C. eutactus*, that have been reported among the most abundant on the basis of the whole relative abundance, catched the attention for the extremely rare presence (20/60 and 36/553 *B. crossotus*; 20/60 and 33/553 *C. eutactus*). Interestingly, they reached very high relative abundance in single samples: *B. crossotus* 34.93% in SWE-01 and *C. eutactus* 16.15% in SWE-24.

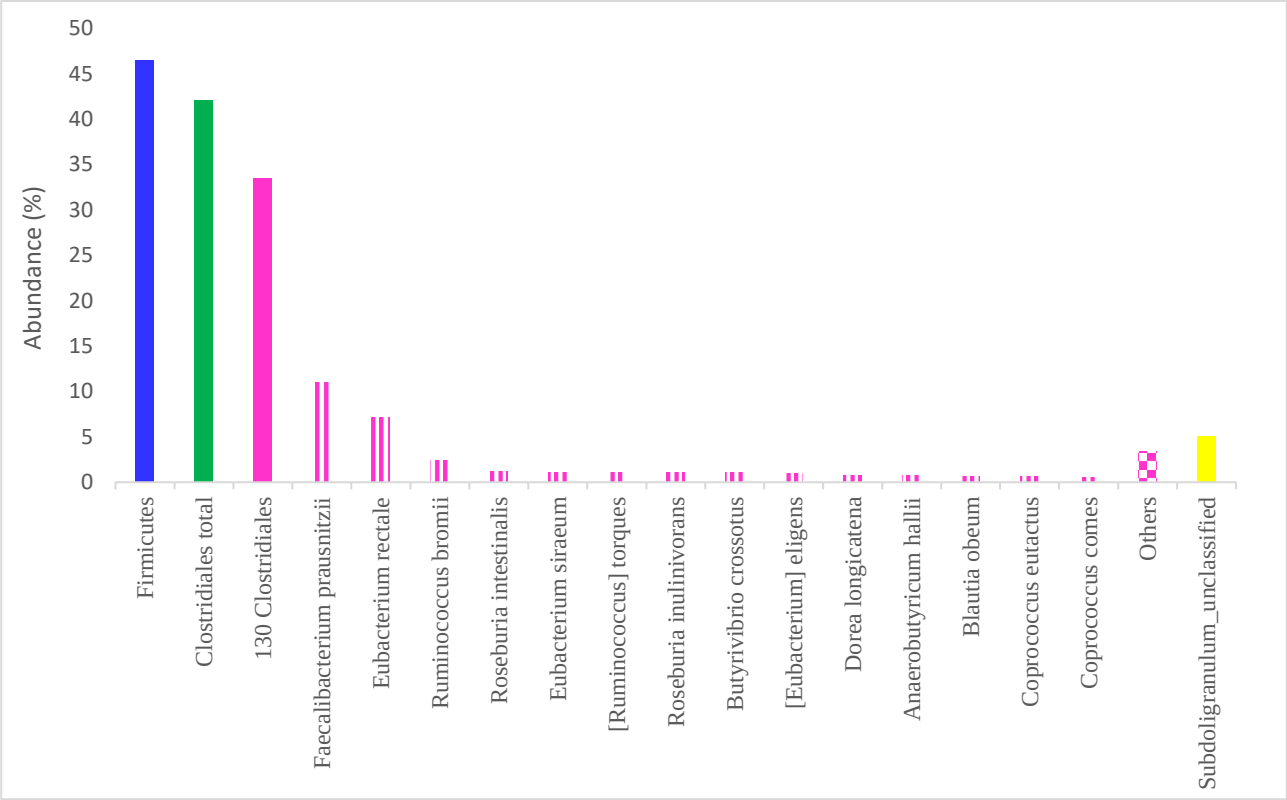


Figure 25 Mean abundance of Firmicutes, Clostridiales, the sum of the 130 species investigated, and the 14 more abundant species.

4.3.2 Phylogenomic analysis of gut Clostridiales

We wanted to infer phylogenesis of the intestinal Clostridiales species from metagenome data, trying to organize the boundaries on the basis of genome similarity. Phylogenetic relationships were searched within a set of 133 genomes, that included also the 3 genomes of *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 sequenced in this PhD Project.

Pairwise AAI values were calculated for the data set. AAI represents pairwise comparison between all protein-coding sequences between a pair of genomes and offers robust resolution between strains of the same or closely related species. Taking into account to the genus boundaries set at 55-60% AAI (Rodriguez-R & Konstantinidis, 2014), 35 species did not cluster with others and seem to belong each to a diverse genus (Fig. 26).

Only the genera *Acetivibrio* (3 species), *Butyrivibrio* (2 species), *Desulfotomaculum* (4 species), *Peptostreptococcus* (2 species), *Ruminoclostridium* (2 species), *Thermaerobacter* (2 species), and *Anaerostipes* (2 species) are consistent with the current taxonomy, forming clusters that encompassed all the species (at least among those pinpointed in gut microbiota) ascribed to the genus.

15 species of Clostridium resulted not associated with any other Clostridiales at genus level. On the other side, three putative genera were identified, encompassing only *Clostridium* spp. (1: *C. butyricum, C. saccharobutylicum, C. beijerinckii, C. saccharoperbutylacetonicum;* 2*: C. celatum, C. sartagoforme, C. tertium*; 3: *C. novyi, C. colicanis;* 4*: C. pasteurianum, C. arbusti;* 5*: C. carboxidivorans, C. tyrobutyricum, C. kluyveri, C. autoethanogenum, C. ljungdahlii;).*
Other *Clostridium* spp. were included on polyphyletic clusters encompassing species currently ascribed to diverse genera.

Ruminococcus gnavus, R. torques, R. lactaris, Tyzzarela nexilis, C. hylemonae, C. scindens, Coprococcus comes, Dorea longicatena, and *D. formicigenerans* represented the putative largest genus identified. Other expected genera encompass:

- *Eubacterium rectale, Roseburia inulinivorans, R. hominis,* and *R. intestinalis;*
- *Enterocloster asparagiformis, E. citroniae, E. boltae*, and *C. clostridiforme*;
- *Flavonifractor plautii* and *Pseudflavonifractor capillosus*;
- *Terrisporobacteri glycolicus, Clostridioides difficile, Paeniclostridium sordellii*, and *Paraclostridium bifermentans*;
- the species ascribed to the genera *Desulfitobacterium* (4) and *Desulfosporosinus* (4);
- *Blautia producta* and *B. hansenii*, the latter also presenting a strong relationship with *R. torques*, whereas *B. obeum* and *B. hydrogenotrophica*, according to the AAI scores, belong to two other diverse genera.

F. prausnitzii did not cluster with other species, and its closest relative was *Subdoligranulum variabile* (AAI = 59.06%).

13 other phylogenetic relationships at higher taxonomic level than genus have been identified (green boxes in Fig. 26). *Desulfotomaculum* clusters with *Desulfoscipio, Desulfofarcimen,* and *Desulfurispora. Eubacterium limosum* and *Acetobacterium woodii* are related.

The putative genus encompassing *Terrisporobacter glycolicus, Clostridioides difficile, Paeniclostridium sordellii*, and *Paraclostridium bifermentans* is related to *Peptacetobacter hiranonis* and *Intestinibacter bartlettii*.

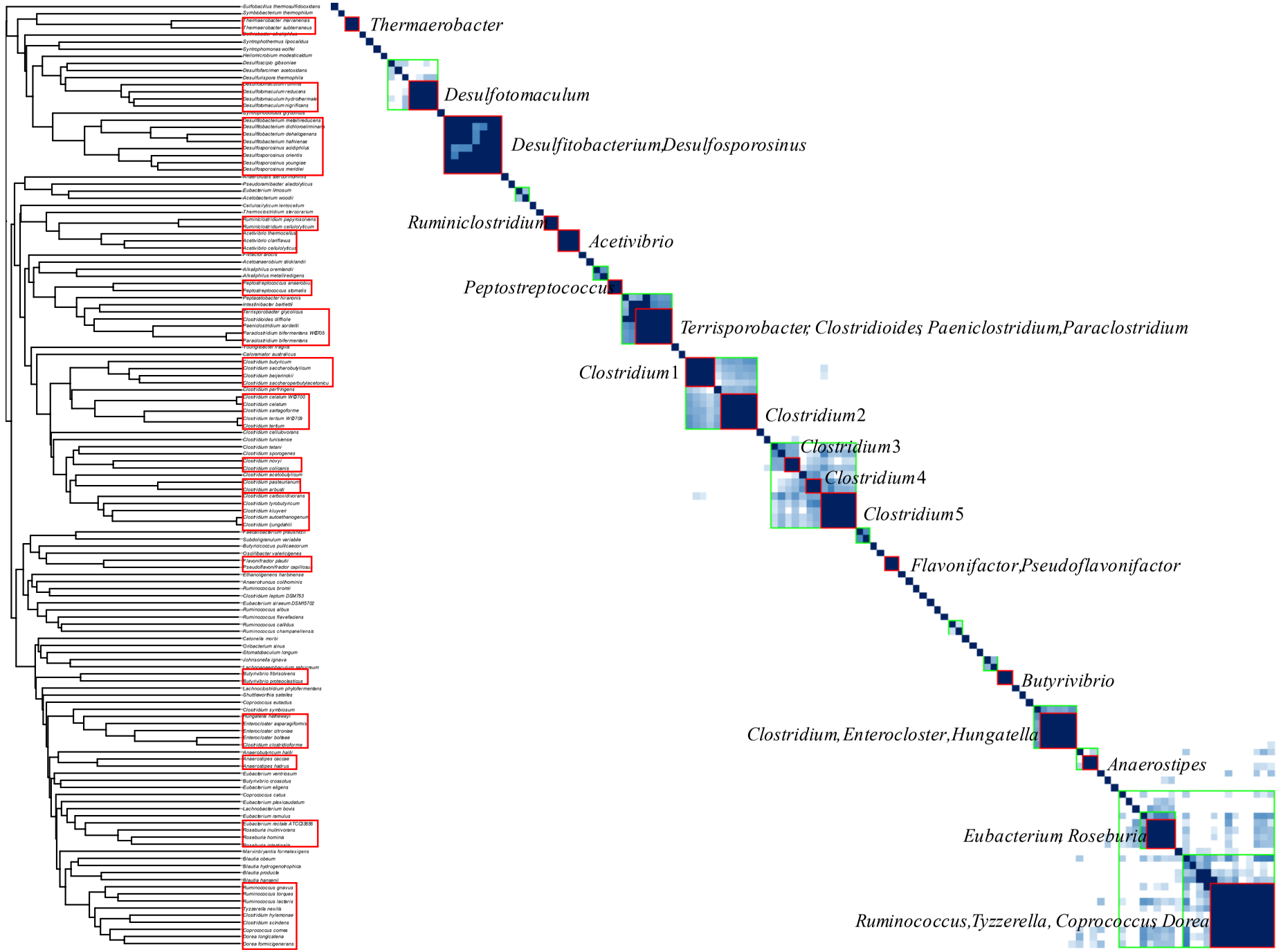
C. perfringens is somehow related to the two Clostridium genera 1. and 2., while a larger cluster composed by Clostridium genera 3., 4., and 5., also encompasses *C. tetani, C. sporogenes*, and *C. acetobutylicum.*

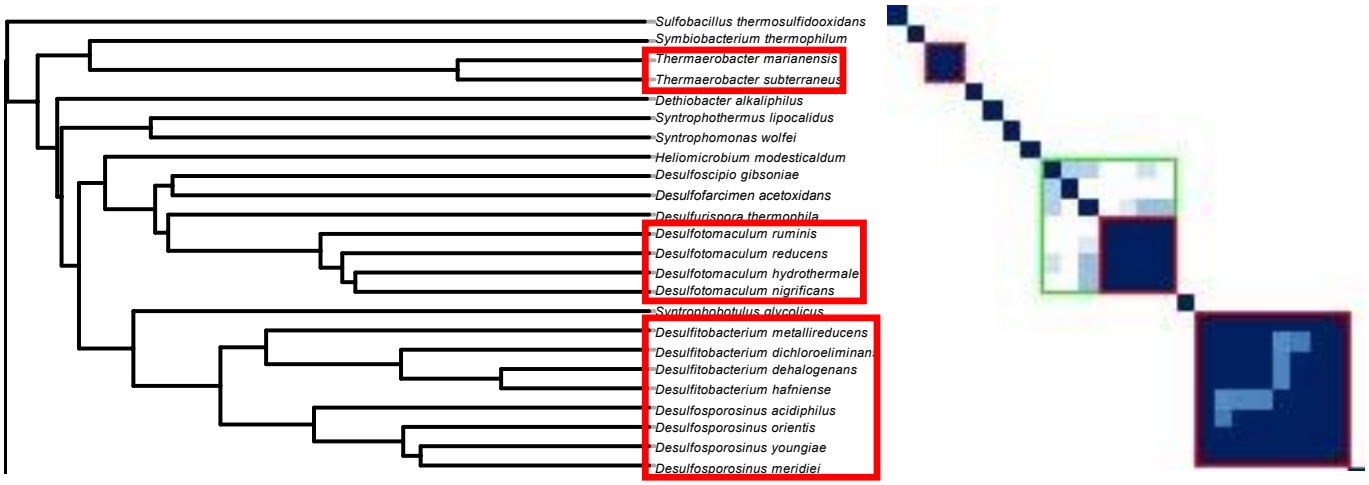
The species *Coprococcus catus, Eubacterium plexicaudatum, Lachnobacterium bovis, E. ramulus, E. rectale, Roseburia inulinivorans, R. hominis, R. intestinalis, Marvinbryantia formatexigens, B. obeum, B. hydrogenotrophica, B. producta, B. hansenii, Ruminococcus gnavus, R.*

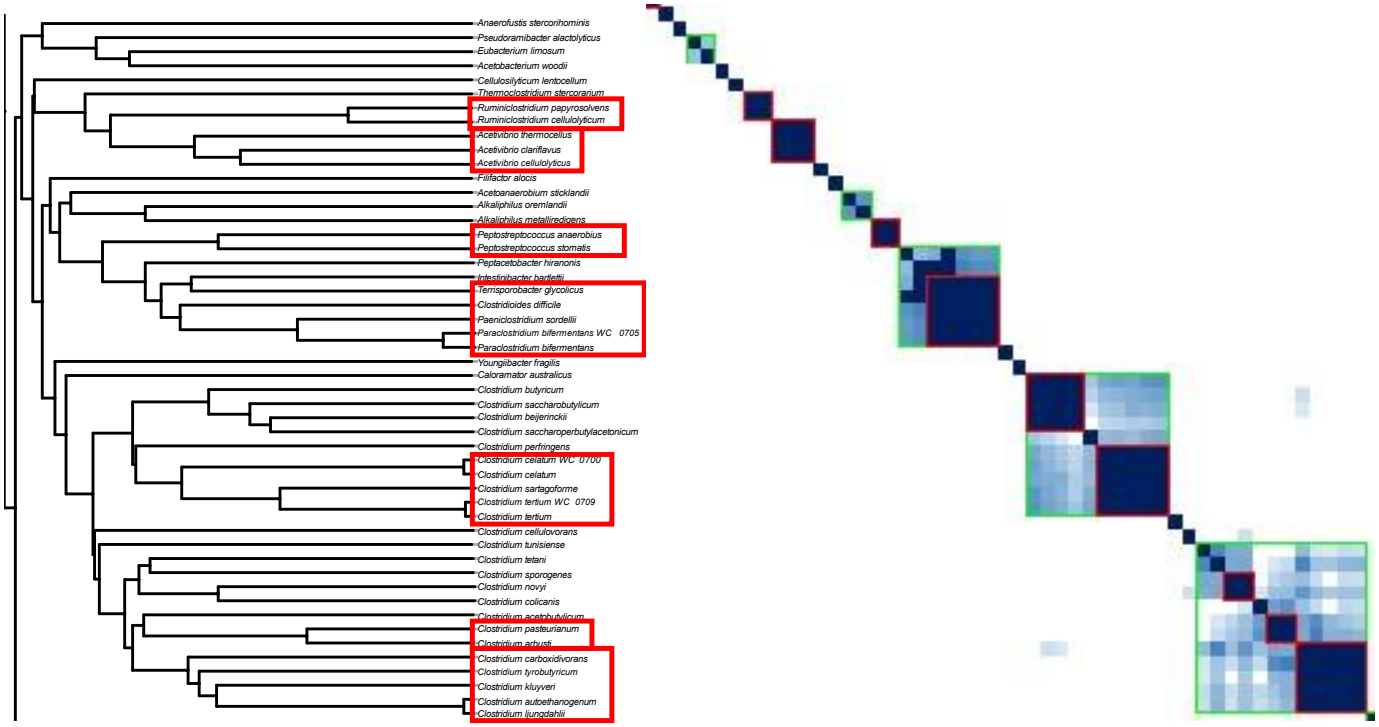
torques, R. lactaris, Tyzzerella nexilis, C. hylemonae, C. scindens, Coprococcus comes, Dorea longicatena, and *D. formicigenerans* are in some way related among each other, and albeit some species clearly belong to independent genera, can be tracked back to a common evolutive origin.

The genus *Anaerostipes*, including the intestinal species *A. caccae* and *A. hadrus*, is related to *Anaerobutyricum halii*.

Some evolutionary relationship is present within the couples of species *Johnsonella ignava* and *Lachnoanaerobaculum saburreum*, *Ruminococcus callidus* and *R. champanellensis*, *Alkaliphilus oremlandii* and *A. metalliredigens*, *Eubacterium limosum* and *Acetobacterium woodii*.


Thermaerobacter
Desulfotomaculum
Desulfitobacterium,Desulfosporosinus
Ruminiclostridium
Acetivibrio
Peptostreptococcus
Terrisporobacter, Clostridioides, Paeniclostridium,Paraclostridium
Clostridium1
Clostridium2
Clostridium3
Clostridium4
Clostridium5
Flavonifactor,Pseudoflavonifactor
Butyrivibrio
Clostridium, Enterocloster, Hungatella
Anaerostipes
Eubacterium, Roseburia
Ruminococcus, Tyzzerella, Coprococcus, Dorea


Sulfobacillus thermosulfidooxidans
Symbiobacterium thermophilum
Thermaerobacter marianensis
Thermaerobacter subterraneus
Dethiobacter alkaliphilus
Syntrophothermus lipocalidus
Syntrophomonas wolfei
Heliomicrobium modesticaldum
Desulfoscipio gibsoniae
Desulfofarcimen acetoxidans
Desulfurispora thermophila
Desulfotomaculum ruminis
Desulfotomaculum reducens
Desulfotomaculum hydrothermale
Desulfotomaculum nigrificans
Syntrophobotulus glycolicus
Desulfitobacterium metallireducens
Desulfitobacterium dichloroeliminans
Desulfitobacterium dehalogenans
Desulfitobacterium hafniense
Desulfosporosinus acidiphilus
Desulfosporosinus orientis
Desulfosporosinus youngiae
Desulfosporosinus meridiei


Anaerofustis stercorihominis
Pseudoramibacter alactolyticus
Eubacterium limosum
Acetobacterium woodii
Cellulosilyticum lentocellum
Thermoclostridium stercorarium
Ruminiclostridium papyrosolvens
Ruminiclostridium cellulolyticum
Acetivibrio thermocellus
Acetivibrio clariflavus
Acetivibrio cellulolyticus
Filifactor alocis
Acetoanaerobium sticklandii
Alkaliphilus oremlandii
Alkaliphilus metalliredigens
Peptostreptococcus anaerobius
Peptostreptococcus stomatis
Peptacetobacter hiranonis
Intestinibacter bartlettii
Terrisporobacter glycolicus
Clostridioides difficile
Paeniclostridium sordellii
Paraclostridium bifermentans WC 0705
Paraclostridium bifermentans
Youngiibacter fragilis
Caloramator australicus
Clostridium butyricum
Clostridium saccharobutylicum
Clostridium beijerinckii
Clostridium saccharoperbutylacetonicum
Clostridium perfringens
Clostridium celatum WC 0700
Clostridium celatum
Clostridium sartagoforme
Clostridium tertium WC 0709
Clostridium tertium
Clostridium cellulovorans
Clostridium tunisiense
Clostridium tetani
Clostridium sporogenes
Clostridium novyi
Clostridium colicanis
Clostridium acetobutylicum
Clostridium pasteurianum
Clostridium arbusti
Clostridium carboxidivorans
Clostridium tyrobutyricum
Clostridium kluyveri
Clostridium autoethanogenum
Clostridium ljungdahlii

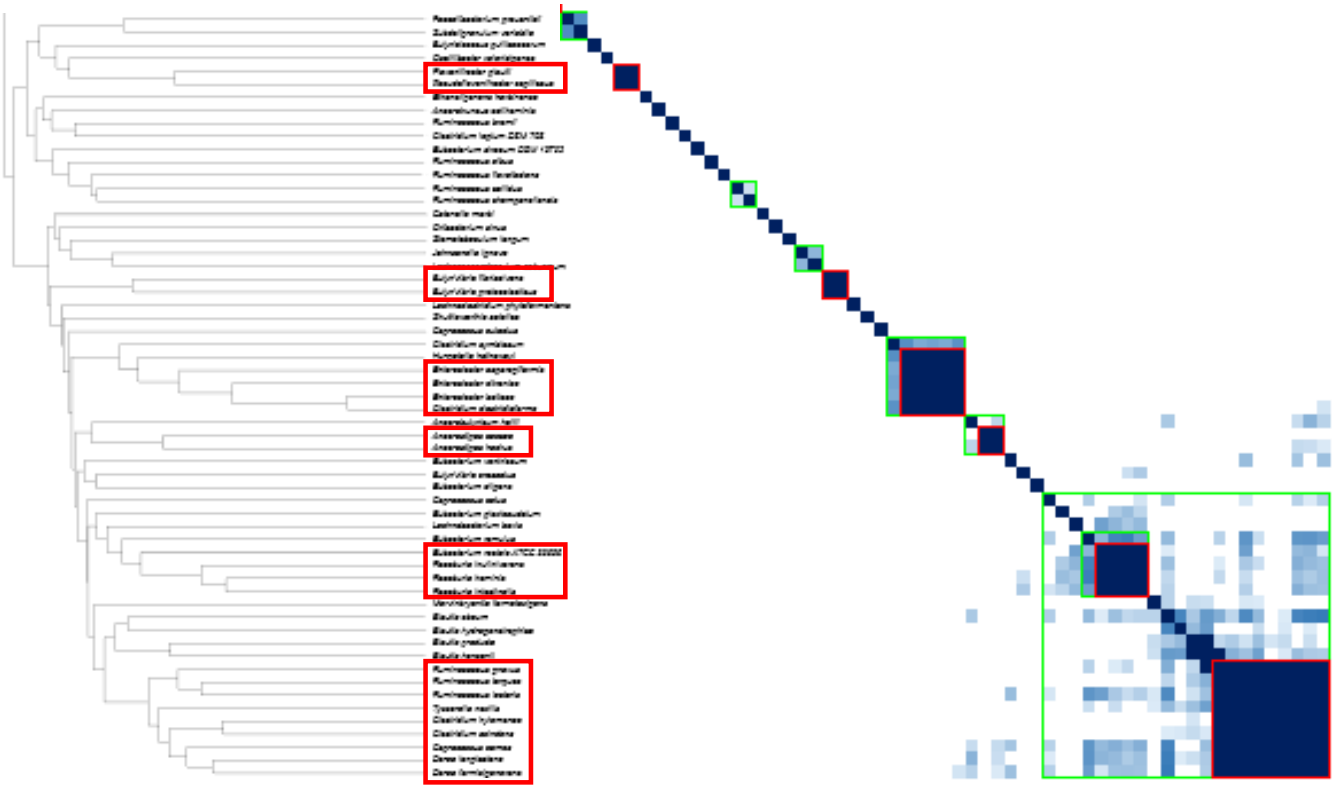


Figure 26 AAI heatmap with correlated tree. Close-ups on three different portions of the tree are reported**.**

The current taxonomy of the 130 species herein considered was analysed in terms of families, utilizing the output on the metagenome analysis (Fig. 27). Most of the species (41) were ascribed to the family of Clostridiaceae, with species forming a consistent cluster that corresponds to Cluster I of the old grouping (Collins et al., 1994).

Two clades of Clostridiaceae were also present within the previously recognized Cluster XIVa, that encompassed all the Lachnospiraceae (with a sole exception), Eubacteriaceae, and a sole species of Ruminococcaceae. Major members of the old Cluster IV were Ruminococcaceae and Oscillospiraceae.

In Cluster XI were included Peptostreptococcaceae and two Clostridiaceae. Peptococcaceae formed a consistent clade, that included the sole species of Heliobacteriaceae. Cluster VIII encompassed the three Syntrophomonadaceae, whereas the three species previously ascribed to Clostridiales family XVII *incertae sedis* and the sole species of Clostridiales family XVIII *incertae sedis* clustered together.

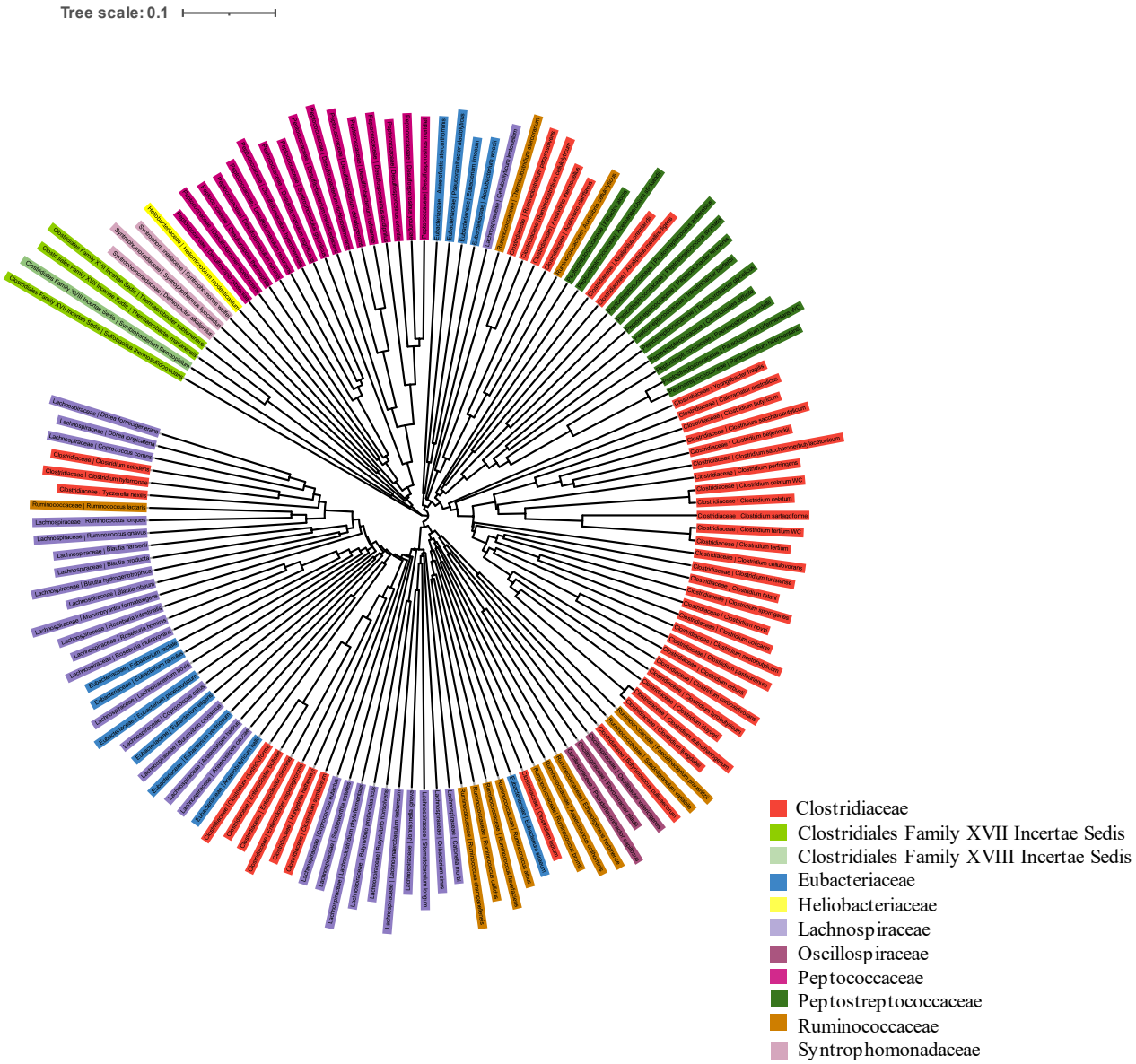


Figure 27 AAI phylogenetic tree with highlighted families

To further define the genera relations, the MiSI method (Barco et al. 2020) was applied to the 130 unique species to calculate gANI and AF in all-against-all approach. The tool yielded 16.770 relations. Only 130 showed both values of gANI > 73.98 and AF > 0.331, while the remaining 16.640 presented values lower than the thresholds (Fig. 28). Among the 130 analyzed species, 71 showed no relation at genus level with any other. The results of MiSI and AAI approaches were consistent, but MiSI resulted more stringent than AAI.

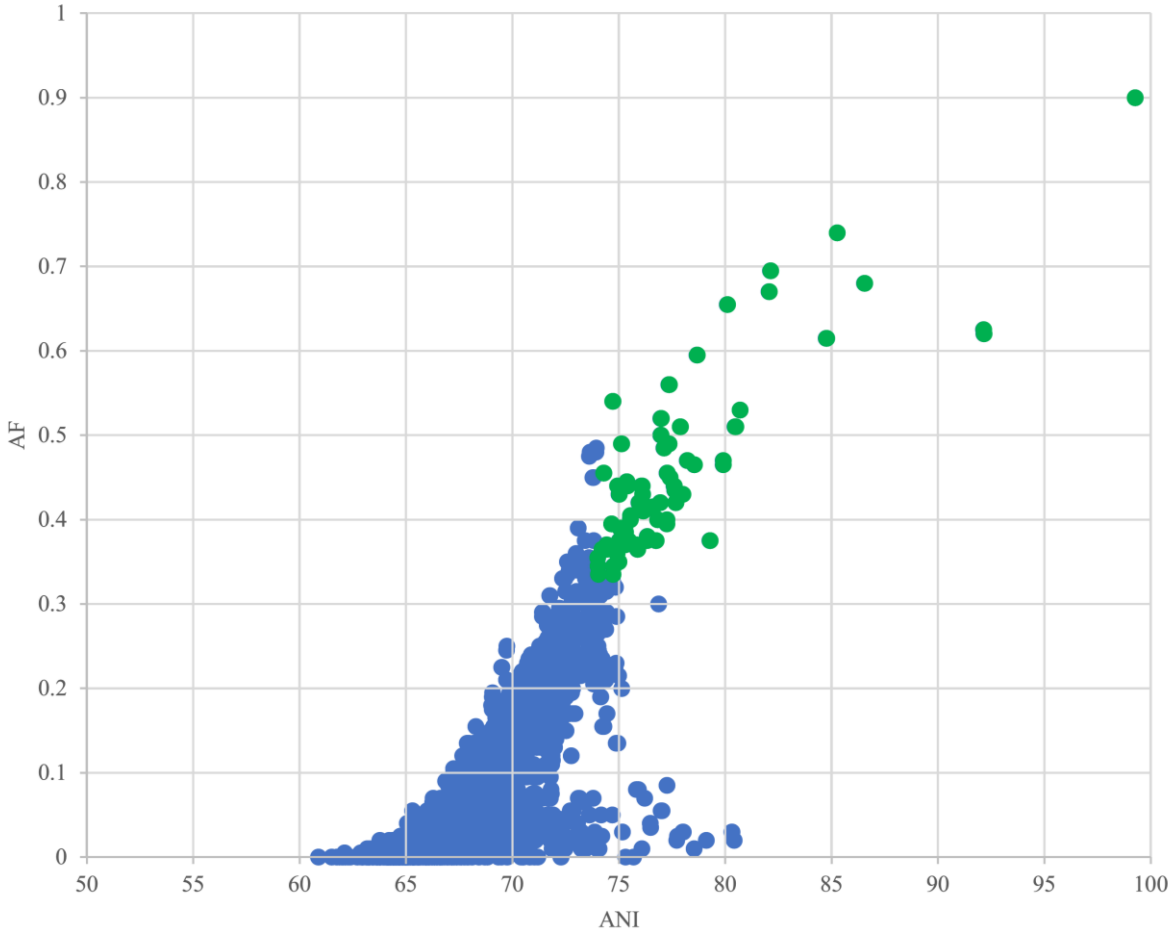


Figure 28 AF and gANI pairwise genome comparison. Green dots indicate relation with AF and gANI values higher than genus thresholds.

4.3.3 Comparison of phylogenetic trees

The tree of AAI (Fig. 26) was compared with those constructed with ANI (Fig. 29) and gANI (Fig. 30) identity scores and with the alignment of 16S rRNA gene (Fig. 31). Comparison was carried out with 'generalized' Robinson-Foulds metrics: the Jaccard–Robinson–Foulds (JRF) and the information-based measures of Mutual Clustering Information (MCI) and Shared Phylogenetic Information (SPI). For each metrics, the comparison yielded a normalized similarity score in the range 0–1 between two trees, reported in Tab. 12. With all the metrics, the tree of 16S rRNA gene presented the greatest similarity with the AAI one, indicating that it has the highest similarity in the topology of nodes, branches, and leaves. Comparison with gANI tree yielded scores slightly lower than 16S rRNA gene, whereas the ANI tree presented the lowest degree of similarity.

	JRF	**MCI**	**SPI**
AAI × ANI	0.484	0.441	0.361
AAI × gANI	0.701	0.637	0.567
AAI × 16S	0.746	0.687	0.652

Table 12 Normalized similarity scores obtained from the comparison of ANI, gANI, and 16S rRNA gene trees with AAI tree, utilizing the generalized' Robinson-Foulds metrics Jaccard-Robinson-Foulds (JRF), Mutual Clustering Information (MCI), and Shared Phylogenetic Information (SPI).

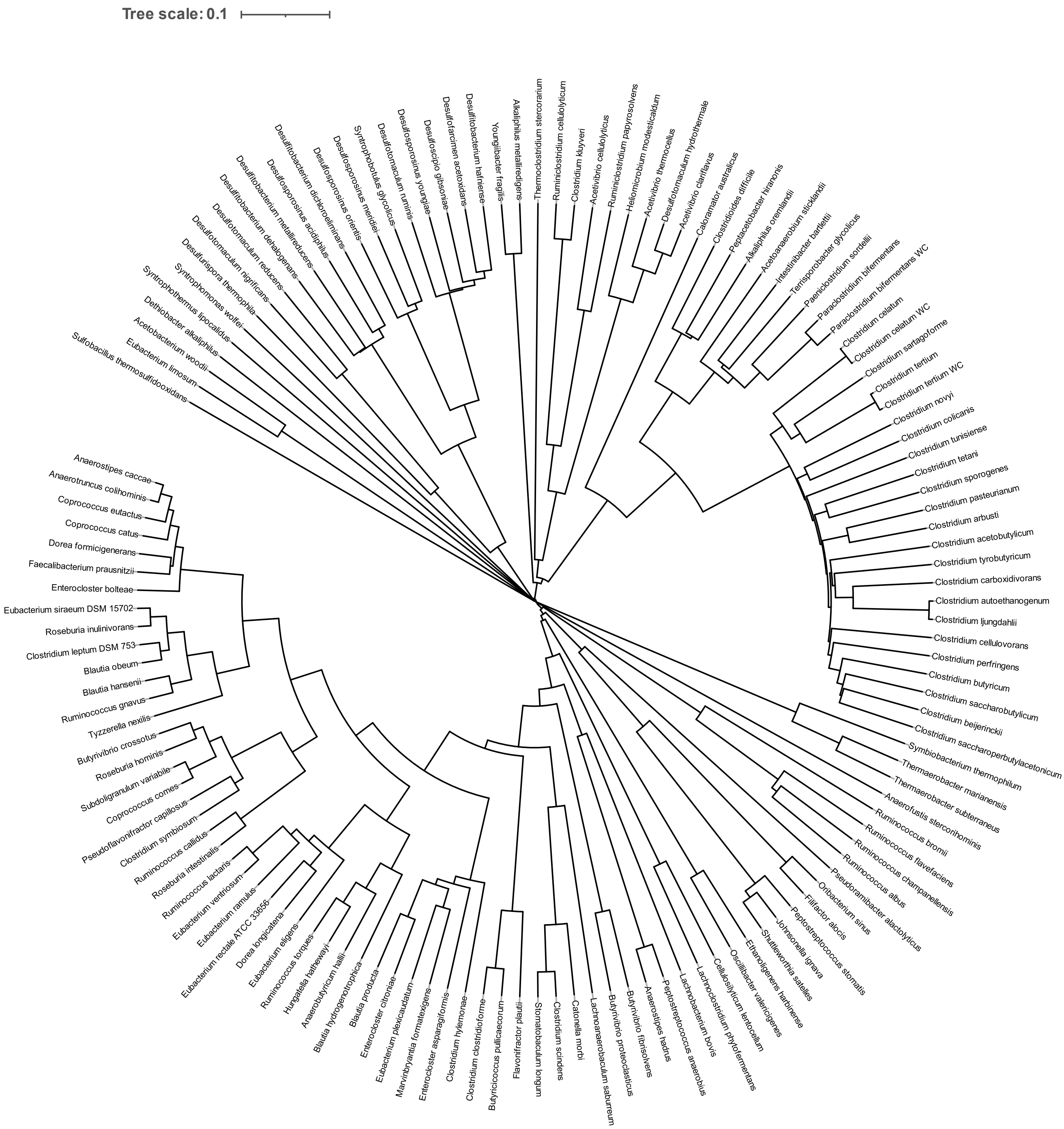


Figure 29 Phylogenetic tree based on ANI

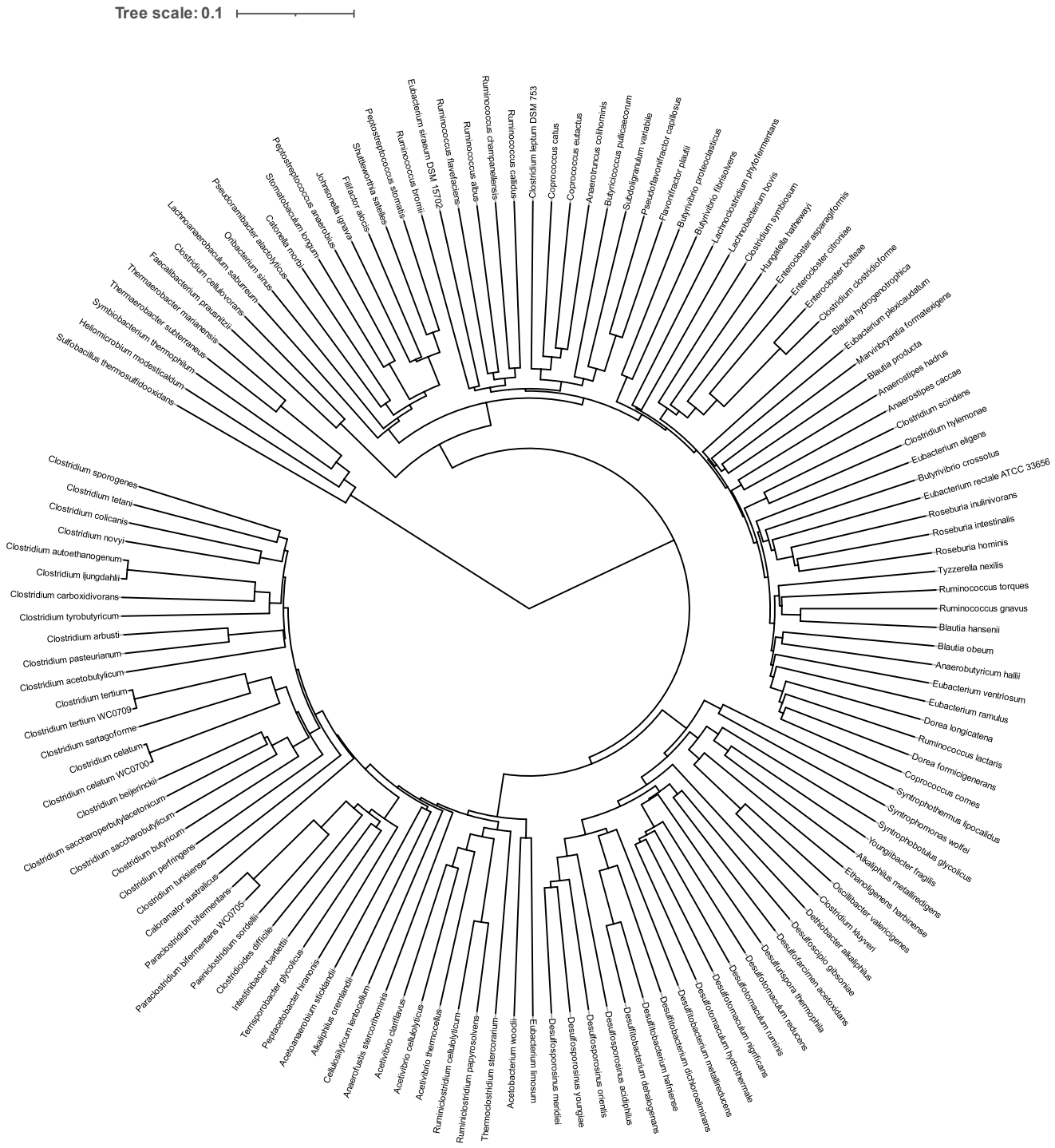


Figure 30 Phylogenetic tree based on gANI

Figure 31 Phylogenetic tree based on 16s rRNA

5. Discussion

The gastrointestinal tract is the largest surface of the human body, and is the site where the highest amount of commensal bacteria are located, particularly in the colon (Pelaseyed et al., 2014). The epithelial surface of GI tract, formed by a single layer of active cells, separates immune cells from a wide range of external stimuli (Melhem et al., 2021). This barrier is represented by the mucosa surface that covers the entire gastrointestinal tract. The mucus layer provides a physical barrier against harmful pathogens and a beneficial environment in terms of nutritional resources and colonization sites for commensal bacteria (Josenhans et al., 2020). As a whole, most of the bacteria-host interactions take place on the mucosal surface (Flaherty et al., 2010).

Considering the central role exerted by the intestinal mucus, it is crucial to deeply investigate mucin degraders of the gut to unveil how mucins influence selection of commensal bacteria and how specific bacteria shape the mucus layer. A healthy microbiota residing in the mucus permits a balance between mucus degradation and biosynthesis of its components, with optimal defence against pathogens and negative stimuli. Then, the description of mucus associated bacteria is mandatory albeit, to date, only few gut commensal bacteria have been indicated as mucin degraders, such as *Akkermansia muciniphila*, *Allobaculum mucolyticum*, *Bacteroides* spp., *Ruminococcus* spp., and *Bifidobacterium* spp. (Berkhout et al., 2021; Derrien et al., 2004; van Muijlwijk et al., 2021).

In order to fill this gap, this project aimed to unveil gut bacterial population capable to degrade mucins through a culture dependent and independent approach. Intestinal mucin degraders of healthy subject were enriched in a medium containing only mucin as the sole carbon and nitrogen source.

The 16S rRNA gene profiling (BioProject ID: PRJNA649741) revealed that the main taxa that took advantages on mucin substrate were ascribed to the phylum Proteobacteria, and were related to Enterobacteriaceae, according to their role as protein degraders in the gut microbiota. Increased abundance of several taxa of Firmicutes *(e.g Enterococcus*, Oscillospiraceae Peptococcaceae, *Lachnoclostridium*, and *Anaerotruncus*) was observed in the enriched samples (Raimondi et al., 2021). Even though metagenomic analysis disclosed a complex population of bacteria able to use mucins, only isolates belonging to three anaerobic species of mucin degrading bacteria (*C. celatum*, *C. tertium,* and *P. bifermentans*) were isolated from the enriched samples of 5 different microbiota. *C. celatum* and *C. tertium*, according to the taxonomical classification belong to the Clostridia cluster I (*Clostridium sensu stricto*), whereas *P. bifermentans* to the family Peptospreptococcaceae. The OTU of *Clostridium sensu stricto* I increased in all the samples where it was present, although it did not reach a significant differential abundance (Raimondi et al., 2021).

In the gut, a complex community of saccharolytic and proteolytic bacteria is expected to participate in mucin breakdown, with linkage-specific glycosidases and proteases of primary degraders providing other bacteria with simpler oligosaccharides, peptides, and monomeric moieties. It is plausible that several taxa that get enriched in mucin medium, detected by the 16S rRNA gene analysis, were able to grow on mucins only when co-cultured within a bacterial community where each species contributed with different hydrolytic activities, while they could not grow in pure cultures where they did not take advantages from crossfeeding relationships (Raimondi et al., 2021).

The three mucinolytic species isolated in this project are poorly characterized, and only few informations were available in scientific literatures (Hauschild & Holdeman, 1974; Jyothsna et al., 2016a; D. L. Miller et al., 2001).

To better understand their functional role in gut, three strains, one for each species identified, were subjected to biochemical, physiological, and genomic characterization. The genome analysis of *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 revealed that the three strains produce a wide set of mucin *O*-glycan targeting enzymes that allow them to degrade mucins. In particular, *C. tertium* WC 0709 had the highest number (200) of putative ORFs encoding proteins involved in metabolism and transport of mucin carbohydrates, followed by *C. celatum* WC 0700 (162), and by *P. bifermentans* WC 0705 (95). *C. tertium* WC 0709 and *C. celatum* WC 0700 encoded diverse glycosyl hydrolases (GH), many of which targeting similar glycoside linkage. For example, four β-galactosidases (GH2) and four β-hexosaminidases (GH20) were detected in the *C. tertium* WC 0709 repertoire of CAZymes, while *C. celatum* WC 0700 presented five β-galactosidases (GH2), four and five β-hexosaminidases of GH20 and GH84 families, respectively. *C. celatum* WC 0700 also encoded five endo-α-*N*-acetylgalactosaminidases (GH101) that catalyze the hydrolysis of core 1 Galβ1-3GlcNAc from mucin protein backbone (Josenhans et al., 2020). This redundancy of CAZymes with similar function may be related to the high variability of mucin structures in terms of glycan composition and linkages along the gastrointestinal tract (Robbe et al., 2004; Tailford et al., 2015). *P. bifermentans* WC 0705 had the smallest set of GHs, with only one galactosidase (GH2) and one hexosaminidase (GH20). Furthermore, most of the identified GHs in all the three strains had a signal peptide, indicating that these enzymes are mainly associated to the bacterial surface or secreted into the environment, acting in the extracellular compartment because of the large size and degree of polymerisation of the mucin structure (Berkhout et al., 2021).

Pathways reconstruction was performed, focusing the interest on mucin *O*-glycans utilization, according to what was predicted by Ravcheev & Thiele (2017). In this approach, it is noteworthy to take into account that in the three genomes a large ratio of genes was not annotated, keeping hidden

a wide part of metabolic and functional potential. Genes required for utilization of galactose, *N*-acetylgalactosamine, and *N*-acetylglucosamine were widespread among the three strains, while only *C. celatum* WC 0700 encoded all the putative enzyme for fucose utilization, and *C. tertium* WC 0709 for *N*-acetylneuraminic acid degradation. Even if two sialidases GH33 were identified in the genome of *C. celatum* WC 0700, the corresponding proteins for sialic acid catabolism were not detected. It is possible that the expression of GH33 from *C. celatum* WC 0700 is linked to the removal of *N*-acetylneuraminic acid from the terminal position of mucin chain, allowing the degradation of the underlying glycans. The results obtained for *C. celatum* WC 0700 were in accordance with the comparative analysis made by Ravcheev & Thiele (2017). The authors indicated *C. celatum* DSM 1785 as a generalist mucin degrader, having a glycan-cleavage pattern that includes 50 glycans, lacking only of α-galactosidase activity. Furthermore, the strain *C. celatum* DSM 1785 can form beneficial pairs with *Bifidobacterium* spp., *Lactobacillus* spp., and with Enterobacteriaceae.

The predicted ability of *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 to ferment mucin *O*-glycan monomers, was verified culturing the strains in a medium supplemented with the single monosaccharides that compose mucin's structure. Interestingly, growth of *C. tertium* WC 0709 and *P. bifermentans* WC 0705 was determined also without the presence of a carbohydrate, indicating the capability of these species to ferment other carbon sources, such as amino acids. Generally, biomass yield of *C. tertium* WC 0709 was more abundant than that obtained for *C. celatum* WC 0700 and *P. bifermentans* WC0705. For the latter, the presence of fermentable carbohydrates such as glucose or *N*-acetyl-galactosamine determined a limited, but statistically significant, increase of biomass yield. On the other side, albeit also *C. tertium* WC 0709 grew without supplement of carbohydrate, for this strain the availability of fermentable monosaccharides determined the doubling of the OD_{600}, suggesting a better affinity for carbohydrates. However, on cultures grow in mucin, the highest biomass yield was obtained by *C. celatum* WC 0700, in agreement with a major metabolic capability to ferment mucin components.

Cultures of *C. celatum* WC 0700 on single monosaccharides confirmed the predicted metabolic capabilities, with successful growth in M17 supplemented with L-fucose, galactose, or *N*-acetylglucosamine. *C. tertium* WC 0709 grew on *N*-acetylneuraminic acid, galactose, *N*-acetylglucosamine, and fucose. On the other side, *P. bifermentans* WC 0705 grew only in the medium supplemented with *N*-acetylgalactosamine. Albeit the potential to use fucose was highlighted by both genome analysis and cultures on this carbohydrate, growth on fucose failed for the three strains in the API test. The use of this carbon source maybe requires some adaptation that is not possible within the strict time limits of API test. Conversely, this bias can be overcome by the three subsequent steps of cultivation carried on in cultures with the single monosaccharides.

The pathway reconstruction was also carried out to predict metabolic functions, including membrane transporters of *C. celatum* WC 0700, *P. bifermentans* WC 0705, and *C. tertium* WC 0709. All the strains shared seven complete phosphotransferase system (PTS) transporters for glucose, GlcNAc, maltose, sucrose, cellobiose-diacetyl chitobiose, GalNAc, and fructose family. In addition, in *C. tertium* WC 0709 PTS systems for α-glucoside, galactosamine, galactitol, sorbose, mannitol, and mannose were predicted. Both *C. tertium* WC 0709 and *C. celatum* WC 0700 encoded putative PTS system involved in β-glucoside, trehalose, and ABC transporters for raffinose, stachyose, melbiose, methyl-galactoside, D-methionine, zinc, galactose oligomers, lysine, teichoic acid, and lantibiotics. PTS systems involved in L-ascorbate, mannose, and β-glucoside transport were predicted in the genome of *P. bifermentans* WC 0705. The metabolism of several of these substrates were checked with API 50 CH tests. All the three strains were able to ferment D-ribose, D-galactose, D-glucose, D-fructose, D-mannose, D-cellobiose, D-maltose, D-lactose, and esculin. *C. tertium* WC 0709 resulted positive for the metabolism of D-xylose, L-sorbose, N-acetylglucosamine, arbutin, salicin, D-melezitose, and D-tagatose. *P. bifermentans* WC 0705 was the sole strain growing on starch, xylitol, gluconate, and 2-ketogluconate, while only *C. celatum* WC 0700 fermented D-trehalose.

According to reconstructed central carbohydrate catabolism, some pathway seemed incomplete, for example the degradation of ascorbate, glucuronate, and galacturonate, biosynthesis of aminoacids, and transformation of chorismate into phenylalanine and tyrosine. Particularly, pathway incompleteness was mostly observed in *P. bifermentans* WC 0705. This observation may be related to nutritional auxotrophies or to the deficiency in gene annotation. Indeed, according to Prokka annotation, *P. bifermentans* WC 0705 had the highest number (68.76%) of ORFs not annotated, followed by *C. celatum* WC 0700 and *C. tertium* WC 0709 having 43.53% and 41.03% of hypothetical proteins, respectively. The proportion of protein-coding genes for which a molecular function cannot be predicted, is higher when these are encoded by species not well described and characterized (Heintz-Buschart & Wilmes, 2018).

Phenotypical description of the isolates *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 revealed that all the strains reached the highest biomass production after 24 h of growth at 37° C in anaerobiosis, with optimal pH values close to neutrality (6.5-7). The three strains fermented mucin producing mainly acetic, formic, propionic acids, and hydrogen, while the production of isovaleric acids was higher in the first 24 h and then slightly decreased in all the strains.

Several studies indicated that gut bacteria inhabit mucus as polymicrobial biofilm, providing benefits to the host thanks to the exchange of nutrients on the epithelial surface, and protecting from pathogens invasion (S. Macfarlane & Dillon, 2007; J. Yang et al., 2020). The strains *C. celatum* WC

0700 and *P. bifermentans* WC 0705 produced a weak biofilm but no differences were observed in mucin coated surface, indicating the absence of capability to adhere to the terminal residues of the mucin structure. The biofilm forming ability of *P. bifermentans* WC 0705 was in line to the definition of this species as weak biofilm former given by Sadiq et al., 2021.

Some resistance to antibiotics was highlighted. *P. bifermentans* WC 0705 and *C. tertium* WC 0709 were resistant to tetracycline and ampicillin, respectively, as predicted by genome analysis. Many clinical cases of bacteraemia due to *C. tertium* were described in literature, and highlights its resistance to aminoglycosides and third generation cephalosporins (D. L. Miller et al., 2001; Vanderhofstadt et al., 2010). Interestingly, gentamicin resistance was a common trait among all the three strains although it was not predicted by none of the antibiotic resistance identifier tools.

Endospore formation is a common feature among clostridia, that allow the survive in harsh conditions (Bhattacharjee & Sorg, 2018). The endospore formation of *C. celatum* WC 0700, *C. tertium* WC 0709, and *P. bifermentans* WC 0705 was tested under temperature and oxygen stress condition. After the exposure to oxygen for one hour or heat treatment at 80°C for 30 minutes, the strains resumed vegetative growth when anaerobic environment was restored, with only a decrease of one Log_{10} (MPN/ml) in *C. celatum* after both the treatments, and in *C. tertium* after the heat treatment. The ability to resist to oxygen exposure could represents a positive trait for the mucus layer colonization. Indeed, it was demonstrated that although the O_2 concentration decrease along the gastrointestinal tract, an opposite trend was observed from the colon lumen to the intestinal epithelium where the O_2 tension increase from 0.4% to 5% (Kint et al., 2020).

Oral supplementation with either *C. tertium* WC 0709 or *P. bifermentans* WC 0705 produced quite different effects on the mucosal immune system. *C. tertium* reduced the number of CD3+CD8+ lymphocytes, but their expression of cytokines IFNγ and IL-17, involved in the immune response toward infective agents, increased. Since IFN-γ/IL-4 ratio was greater in mice with *C. tertium* WC 0709 intervention compared to controls, the data suggest that *C. tertium* tends to mediate the cell-mediated Th1 immune response. This polarization of mucosal immune system has been reported with some probiotic strains such as *Lactobacillus rhamnosus* GG, and is claimed as a positive effect (C. Shi et al., 2020). Moreover, an increase in IL-17 promotes the ability of the host to resist infection by bacteria, mycobacteria, and fungi (Curtis & Way, 2009). Indeed, luminal bacteria such as *L. rhamnosus* GG can promote the differentiation of Th0 cells into Th17 cells, promoting the expression of IL-17, regulate immunity and enhance the defences of the host (R. Chen et al., 2016). However, while *L. rhamnosus* GG improve the differentiation and promote the balance of Th-17 and Treg cells to alleviate inflammatory mediated tissue injury (R. Chen et al., 2016; C. Shi et al., 2020), *C. tertium* WC 0709 favours a Th1 response.

Supplementation with *P. bifermentans* WC 0705 exerted a significant effect on Treg cells, enhancing the percentage in the gut lamina propria. The increase of Treg cells in intestinal mucosa is generally viewed as an extremely favourable event since these cells exert a potent immune-modulatory effect reducing the severity of inflammatory responses (Jacobse et al., 2021). Regulation of development and/or effector functions of different immune cell populations (such as Treg cells vs Th17 cells) depends on the activity of different members of the commensal community. Thus, it is well accepted that the relative abundance of immunomodulatory members of gut microbiota can directly influence the host mucosal and systemic immune system (Ivanov & Honda, 2012). Indeed, previous studies have shown that segmented filamentous bacteria can induce interleukin-17-producing CD4+ T cells whereas combination of indigenous clostridia to conventionally reared mice or derived from human microbiota can induce Treg cells in the mouse colonic lamina propria and thereby contribute to protecting mice against colitis and allergic responses (Atarashi et al., 2011; Ivanov et al., 2009; Narushima et al., 2014).

Interestingly, the effect of indigenous clostridia to regulate Treg cells seem dependent on productions of short-chain fatty acids (SCFAs) such as that butyrate, that inhibits histone deacetylases increasing transcription at the Foxp3 promoter, and propionate, that signals through GPR43 to induce FOXP3+ Treg cells (van der Hee & Wells, 2021). Indeed, *C. tertium* effectively produces SCFA from indigested fibers and from mucin, representing an attractive source of butyrate and propionate that exert favourable effects on gut mucosal immune system.

Since three species of Clostridiales (*C. celatum*, *C. tertium*, and *P. bifermentans*) became the central topic of this project, we wanted to determine their closest evolutionary relationships and perform a phylogenomic study of Clostridiales colonizing the intestine of healthy subjects. Two datasets were utilized: a larger one encompassing 553 metagenomes from the HPM project (Turnbaugh et al., 2007), and another built in this laboratory to make an inventory of β-glucuronidases (Candeliere et al., 2022). Interestingly, the second dataset, albeit with a lower numerosity, extracted a higher number of species than the HMP. Then, the further description of the gut Clostridiales was made using this database.

This analysis identified 133 species, the importance of which was investigated according to the whole abundance, to the highest relative amount reached in a sample, and to the frequency of identification in the metagenomes. *Faecalibacterium prausnitzii*, an efficient butyrate producer (Louis & Flint, 2009), was the most abundant Clostridiales among all the samples, followed by *Eubacterium rectale* and *Ruminococcus bromii*. Interestingly, a *Subdoligranulum* not classified at species level was one of the dominant bacteria within the gut microbiota. *F. prausnitzii*, *R. torques*, and *R. obeum* were detected in all the 60 samples. Other recurrent species not recognized among the

most abundant were *D. formicigenerans, R. hominis, C. catus*, *E. ramulus*, *Anaerostipes hadrus*, and *E. ventriosum*.

Diverse phylogenomic approaches were applied to organize the Clostridiales order and validate current genera and families. AAI was chosen as reference method since ANI is less performing for distant relationships. Our results had not the ambition to re-organize the order but wanted to provide some consistent information concerning phylogenesis of intestinal Clostridiales. Most of the Clusters recognized by Collins (1994) were consistent, and also some families formed separate clades. However, most of the genera included species that did not cluster together.

As a whole, this preliminary study shed light on the species of Clostridiales detected in human gut microbiota, with information about the frequency and the abundance. Working on a reduced number of species and not on the whole set of Clostridiales, allowed to provide a preview of the reorganization of the order, necessary to obtain a more consistent and robust proposal of genera and to identify the actual limits of family demarcation.

6. References

Aagaard, K., Ma, J., Antony, K., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Sci Transl Med*, *6*(237). https://doi.org/10.1126/scitranslmed.3008599

Abbeele, P. Van Den, Belzer, C., Goossens, M., Kleerebezem, M., Vos, W. M. De, Thas, O., Weirdt, R. De, Kerckhof, F., & Wiele, T. Van De. (2013). Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. *The ISME Journal*, *7*, 949–961. https://doi.org/10.1038/ismej.2012.158

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Afzal, M., Manzoor, I., & Kuipers, P. (2015). GalR Acts as a Transcriptional Activator of galKT in the Presence of Galactose in Streptococcus pneumoniae. *J Mol Microbiol Biotechnol*, *25*(6), 363–371. https://doi.org/10.1159/000439429

Agans, R., Rigsbee, L., Kenche, H., Michail, S., Khamis, H., & Paliy, O. (2011). Distal gut microbiota of adolescent children is different from that of adults. *FEMS Microbiol Ecol*, *77*(2), 404–412. https://doi.org/10.1111/j.1574-6941.2011.01120.x

Agergaard, C. N., Hoegh, S. V., Holt, H. M., & Justesen, U. S. (2016). Two serious cases of infection with clostridium celatum after 40 years in hiding? *Journal of Clinical Microbiology*, *54*(1), 236–238. https://doi.org/10.1128/JCM.01938-15

Agus, A., Clément, K., & Sokol, H. (2021). Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut*, *70*(6), 1174–1182. https://doi.org/10.1136/gutjnl-2020-323071

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., … McArthur, A. G. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *48*(D1), D517–D525. https://doi.org/10.1093/nar/gkz935

Alemao, C. A., Budden, K. F., Gomez, H. M., Rehman, S. F., Marshall, J. E., Shukla, S. D., Donovan, C., Forster, S. C., Yang, I. A., Keely, S., Mann, E. R., El Omar, E. M., Belz, G. T., & Hansbro, P. M. (2020). Impact of diet and the bacterial microbiome on the mucous barrier and immune disorders. *Allergy: European Journal of Allergy and Clinical Immunology*, *July*, 1–21.

https://doi.org/10.1111/all.14548

Allen, A., & Flemström, G. (2005). Gastroduodenal mucus bicarbonate barrier: Protection against acid and pepsin. *American Journal of Physiology - Cell Physiology*, *288*(1 57-1). https://doi.org/10.1152/ajpcell.00102.2004

Almagro-moreno, S., & Boyd, E. F. (2009). Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol Biol*, *16*, 1–16. https://doi.org/10.1186/1471-2148-9-118

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., Pollard, K. S., Sakharova, E., Parks, D. H., Hugenholtz, P., Segata, N., Kyrpides, N. C., & Finn, R. D. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology*, *39*(1), 105–114. https://doi.org/10.1038/s41587-020-0603-3

Amaretti, A., Bernardi, T., Leonardi, A., Raimondi, S., Zanoni, S., & Rossi, M. (2013). Fermentation of xylo-oligosaccharides by Bifidobacterium adolescentis DSMZ 18350 : kinetics , metabolism , and β -xylosidase activities. *Appl Microbiol Biotechnol*, *97*(7), 3109–3117. https://doi.org/10.1007/s00253-012-4509-y

Amaretti, A., Gozzoli, C., Simone, M., Raimondi, S., Righini, L., Pérez-brocal, V., García-lópez, R., & Moya, A. (2019). Profiling of Protein Degraders in Cultures of Human Gut Microbiota. *Front Microbiol*, *10*, 1–13. https://doi.org/10.3389/fmicb.2019.02614

Amaretti, A., Righini, L., Candeliere, F., Musmeci, E., Bonvicini, F., Gentilomi, G. A., Rossi, M., & Raimondi, S. (2020). Antibiotic Resistance , Virulence Factors , Phenotyping , and Genotyping of Non- Escherichia coli Enterobacterales from the Gut Microbiota of Healthy Subjects. *International Journal of Molecular Sciences*, *21*(5), 1847. https://doi.org/10.3390/ijms21051847

Anand, S., Kaur, H., & Mande, S. S. (2016). Comparative In silico Analysis of Butyrate Production Pathways in Gut Commensals and Pathogens. *Front Microbiol*, *7*, 1945. https://doi.org/10.3389/fmicb.2016.01945

Anantharaman, K., Brown, C., Hug, L., Sharon, I., Castelle, C., Probst, A., Thomas, B., Singh, A., Wilkins, M., Karaoz, U., Brodie, E., Williams, K., Hubbard, S., & Banfield, J. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun*, *7*(13219). https://doi.org/10.1038/ncomms13219

Andrews, S. (2010). *FastQC: a quality control tool fro high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Antonini, M., Conte, M. Lo, Sorini, C., & Falcone, M. (2019). How the interplay between the commensal microbiota, gut barrier integrity, and mucosal immunity regulates brain autoimmunity. *Frontiers in Immunology*, *10*(AUG). https://doi.org/10.3389/fimmu.2019.01937

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Ashida H, Maki R, Ozawa H, Tani Y, Kiyohara M, Fujita M, Imamura A, Ishida H, Kiso M, Y. K., & Yamamoto, K. (2008). Characterization of two different endo-alpha-N-acetylgalactosaminidases from probiotic and pathogenic enterobacteria, Bifidobacterium longum and Clostridium perfringens. *Glycobiology*, *18*(9), 727–734. https://doi.org/10.1093/glycob/cwn053

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., Taniguchi, T., Takeda, K., Hori, S., Ivanov, I., Umesaki, Y., Itoh, K., & Honda, K. (2011). Induction of colonic regulatory T cells by indigenous Clostridium species. *Science*, *331*(6015), 337–341. https://doi.org/10.1126/science.1198469

Bach, E., Sant'Anna, F., Magrich Dos Passos, JF Balsanelli, E., de Baura, V., Pedrosa, F., de Souza, E., & Passaglia, L. (2017). Detection of misidentifications of species from the Burkholderia cepacia complex and description of a new member, the soil bacterium Burkholderia catarinensis sp. nov. *Pathog Dis*, *75*, 6. https://doi.org/10.1093/femspd/ftx076

Baker, M. (2012). De novo genome assembly: what every biologist should know. *Nature Methods*, *9*, 333–337. https://doi.org/10.1038/nmeth.1935

Baldelli, V., Scaldaferri, F., Putignani, L., & Chierico, F. Del. (2021). The Role of Enterobacteriaceae in Gut Microbiota Dysbiosis in Inflammatory Bowel Diseases. *Microorganisms*, *9*(4), 1–15.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bankole, E., Read, E., Curtis, M. A., Neves, J. F., & Garnett, J. A. (2021). The Relationship between Mucins and Ulcerative Colitis : A Systematic Review. *J Clin Med*, *10*(9), 1935.

Bansil, R., & Turner, B. S. (2006). Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid and Interface Science*, *11*(2–3), 164–170. https://doi.org/10.1016/j.cocis.2005.11.001

Bansil, R., & Turner, B. S. (2018). The biology of mucus: Composition, synthesis and organization. *Advanced Drug Delivery Reviews*, *124*, 3–15. https://doi.org/10.1016/j.addr.2017.09.023

Barakat, M., Hernandez, S., Benoit, J., Pourshahid, S., Mamoon, Y., & Martin, G. T. (2018). Clostridium tertium: An Unusual Cause of Pyogenic Liver Abscess. *ACG Case Reports Journal*,

5(1), e30. https://doi.org/10.14309/crj.2018.30

Barco, R. A., Garrity, G. M., Scott, J. J., Amend, J. P., Nealson, K. H., & Emerson, D. (2020). A genus definition for bacteria and archaea based on a standard genome relatedness index. *MBio*, *11*(1). https://doi.org/10.1128/MBIO.02475-19

Barrett, L. F., Saragadam, S. D., Dimaria, C. N., & Delgado-daza, A. (2020). Infection of a prosthetic knee joint with Clostridium bifermentans. *Oxf Med Case Reports*, *8*, 256–258. https://doi.org/10.1093/omcr/omaa057

Beaumont, M., Portune, K. J., Steuer, N., Lan, A., Cerrudo, V., Audebert, M., Dumont, F., Benamouzig, R., Mancano, G., Khodorova, N., Andriamihaja, M., Airinei, G., Tom, D., Davila, A., Claus, S. P., & Sanz, Y. (2017). *Quantity and source of dietary protein influence metabolite production by gut microbiota and rectal mucosa gene expression : a randomized , parallel , double-blind trial in overweight humans*. 1005–1019. https://doi.org/10.3945/ajcn.117.158816

Béchon, N., & Ghigo, J. (2021). Gut biofilms : Bacteroides as model symbionts to study biofilm formation by intestinal anaerobes. *FEMS Microbiology Reviews*, 1–17. https://doi.org/10.1093/femsre/fuab054

Bell, A., Juge, N., Bell, A., & Juge, N. (2020). *Mucosal glycan degradation of the host by the gut microbiota*. *00*(00), 1–6. https://doi.org/10.1093/glycob/cwaa097

Belzer, C., & De Vos, W. M. (2012). Microbes insidefrom diversity to function: The case of Akkermansia. *ISME Journal*, *6*(8), 1449–1458. https://doi.org/10.1038/ismej.2012.6

Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., & Tabak, L. A. (2012). Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, *22*(6), 736–756. https://doi.org/10.1093/glycob/cwr182

Bergey, D. H., Harrison, F. C., Breed, R. S., Hammer, B. W., & Huntoon, F. M. (1923). *Bergey ' s Manual of Determinative Bacteriology* (The Williams and Wilkins Co. (ed.); 1st ed.). Baltimore.

Bergstrom, K. S. B., & Xia, L. (2013). Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology*, *23*(9), 1026–1037. https://doi.org/10.1093/glycob/cwt045

Berkhout, M. D., Plugge, C. M., & Belzer, C. (2021). How microbial glycosyl hydrolase activity in the gut mucosa initiates microbial cross-feeding. *Glycobiology*, *cwab105*. https://doi.org/10.1093/glycob/cwab105

Beukema, M., Faas, M. M., & de Vos, P. (2020). The effects of different dietary fiber pectin structures on the gastrointestinal immune barrier: impact via gut microbiota and direct effects on immune cells. *Experimental and Molecular Medicine*, *52*(9), 1364–1376. https://doi.org/10.1038/s12276-020-0449-2

Bhattacharjee, D., & Sorg, J. A. (2018). Conservation of the " Outside-in " Germination Pathway in

Paraclostridium bifermentans. *Front Microbiol*, *17*(9), 2487. https://doi.org/10.3389/fmicb.2018.02487

Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, Nikkïla J, Monti D, Satokari R, Franceschi C, Brigidi P, D. V. W. (2010). Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One*, *17*(5). https://doi.org/10.1371/annotation/df45912f-d15c-44ab-8312-e7ec0607604d

Bilal, T., Malik, B., & Rehman, K. (2018). Metagenomic analysis of uncultured microorganisms and their enzymatic attributes. *Journal of Microbiological Methods*, *155*(October), 65–69. https://doi.org/10.1016/j.mimet.2018.11.014

Birchenough, G. M. H., Johansson, M. E. V., Gustafsson, J. K., Bergström, J. H., & Hansson, G. C. (2015). New developments in goblet cell mucus secretion and function. *Mucosal Immunology*, *8*(4), 712–719. https://doi.org/10.1038/mi.2015.32

Björkstén, B., Sepp, E., Julge, K., Voor, T., & Mikelsaar, M. (2001). Allergy development and the intestinal microflora during the first year of life. *J Allergy Clin Immunol*, *108*(4), 516–520. https://doi.org/10.1067/mai.2001.118130

Bolte, L. A., Vila, A. V., Imhann, F., Collij, V., Gacesa, R., Peters, V., Wijmenga, C., Kurilshikov, A., Fu, J., Dijkstra, G., Zhernakova, A., & Weersma, R. K. (2021). Long-term dietary patterns are associated with pro- inflammatory and anti- -inflammatory features of the gut microbiome. *Gut*, *70*, 1287–1298. https://doi.org/10.1136/gutjnl-2020-322670

Boltin, D., Perets, T., Vilkin, A., & Niv, Y. (2013). Mucin function in inflammatory bowel disease: an update. *Journal of Clinical Gastroenterology Clin Gastroenterol*, *47*(2), 106–111. https://doi.org/10.1097/MCG.0b013e3182688e73

Bovenkamp, J. H. B. Van De, Mahdavi, J., Male, A. M. K., Büller, H. A., Einerhand, A. W. C., Borén, T., & Dekker, J. (2003). The MUC5AC Glycoprotein is the Primary Receptor for Helicobacter pylori in the Human Stomach. *Helicobacter*, *8*(5), 521–532. https://doi.org/10.1046/j.1523-5378.2003.00173.x

Brenner, D. J., Fanning, G. R., Rake, A. V., & Johnson, K. (1969). Batch procedure for thermal elution of DNA from hydroxyapatite. *Anal Biochem*, *28*, 477–459.

Brigham, C., Caughlan, R., Gallegos, R., Dallas, M. B., Godoy, V. G., Malamy, M. H., Al, B. E. T., & Acteriol, J. B. (2009). Sialic Acid (N -Acetyl Neuraminic Acid) Utilization by Bacteroides fragilis Requires a Novel N -Acetyl Mannosamine Epimerase □ †. *J Bacteriol*, *191*(11), 3629–3638. https://doi.org/10.1128/JB.00811-08

Bunesova V, Lacroix C, S. C. (2018). Mucin cross-feeding of infant Bifidobacteria and Eubacterium Hallii. *Microb Ecol*, *75*(1), 228–238. https://doi.org/10.1007/s00248-017-1037-4

Candeliere, F., Raimondi, S., Ranieri, R., Musmeci, E., Zambon, A., Amaretti, A., & Rossi, M. (2022). β-glucuronidase pattern predicted from gut metagenomes indicates potentially diversified pharmacomicrobiomics. *Frontiers in Microbiology*, *ahead of p*.

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

Chen, L., Wang, W., Zhou, R., Ng, S. C., Li, J., Huang, M., Zhou, F., Wang, X., Shen, B., Kamm, M. A., Wu, K., & Xia, B. (2014). Characteristics of fecal and mucosa-associated microbiota in chinese patients with inflammatory bowel disease. *Medicine (United States)*, *93*(8), 1–9. https://doi.org/10.1097/MD.0000000000000051

Chen, R., Xu, L., Du, S., Huang, S., Wu, H., Dong, J., Huang, J., Wang, X., Feng, W., & Chen, Y. (2016). Lactobacillus rhamnosus GG supernatant promotes intestinal barrier function, balances Treg and TH17 cells and ameliorates hepatic injury in a mouse model of chronic-binge alcohol feeding. *Toxicol Lett.*, *241*, 103–110. https://doi.org/10.1016/j.toxlet.2015.11.019

Chen, X.-J., Wang, Z.-Q., Zhou, Z.-Y., Zeng, N.-Y., Huang, Q.-F., Wang, Z.-W., & Tang, W.-L. (2020). Characterization of Peptacetobacter hominis gen . nov ., sp . nov ., isolated from human faeces , and proposal for the reclassification of Clostridium hiranonis within the genus Peptacetobacter. *Int J Syst Evol Microbiol*, *70*(5), 2988–2997. https://doi.org/10.1099/ijsem.0.003925

Chun, J., & Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *International Journal of Systematic and Evolutionary Microbiology*, *64*(PART 2), 316–324. https://doi.org/10.1099/ijs.0.054171-0

Chung, H., Pamp, S. J., Hill, J. A., Surana, N. K., Sanna, M., Troy, E. B., Reading, N. C., Villablanca, E. J., Wang, S., Mora, J. R., Umesaki, Y., Mathis, D., Benoist, C., David, A., & Kasper, D. L. (2012). Gut immune maturation depends on colonization with a host-specific microbiota. *Cell*, *149*(7), 1578–1593. https://doi.org/10.1016/j.cell.2012.04.037

Coker, J. K., Moyne, O., Rodionov, D. A., & Zengler, K. (2021). Carbohydrates great and small, from dietary fiber to sialic acids: How glycans influence the gut microbiome and affect human health. *Gut Microbes*, *13*(1), 1–18. https://doi.org/10.1080/19490976.2020.1869502

Collins, M., Lawso, P., Willems, A., Cordoba, J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., & Farrow, J. (1994). The phylogeny of the genus Clostridium : proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol*, *44*(4), 812–826. https://doi.org/10.1099/00207713-44-4-812

Cone, R. A. (2009). Barrier properties of mucus. *Advanced Drug Delivery Reviews*, *61*(2), 75–85.

https://doi.org/10.1016/j.addr.2008.09.008

Coombes, J. L., Siddiqui, K. R. R., Arancibia-, C. V, Hall, J., Sun, C., Belkaid, Y., & Powrie, F. (2007). A functionally specialized population of mucosal CD103 ⁄ DCs induces Foxp3 ⁄ regulatory T cells via a TGF- □ – and retinoic acid – dependent mechanism. *J Exp Med*, *204*(8), 1757–1764. https://doi.org/10.1084/jem.20070590

Corfield, A. P., Wagner, S. A., Clamp, J. R., Kriaris, M. S., & Hoskins, L. C. (1992). Mucin degradation in the human colon: Production of sialidase, sialate O- acetylesterase, N-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infection and Immunity*, *60*(10), 3971–3978. https://doi.org/10.1128/iai.60.10.3971-3978.1992

Corfield, Anthony P. (2015). Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochimica et Biophysica Acta - General Subjects*, *1850*(1), 236–252. https://doi.org/10.1016/j.bbagen.2014.05.003

Corfield, Anthony P. (2018). The Interaction of the Gut Microbiota with the Mucus Barrier in Health and Disease in Human. *Microorganisms*, *6*(3), 78. https://doi.org/10.3390/microorganisms6030078

Cornick, S., Tawiah, A., & Chadee, K. (2015). Roles and regulation of the mucus barrier in the gut. *Tissue Barriers*, *3*(1), 1–2. https://doi.org/10.4161/21688370.2014.982426

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, *46*(W1), W246–W251. https://doi.org/10.1093/nar/gky425

Cresci, G. A., Clinic, C., Bawden, E., Intern, D., & Clinic, C. (2016). *HHS Public Access*. *30*(6), 734–746. https://doi.org/10.1177/0884533615609899.The

Crosa, J. H., Brenner, D. J. & Falkow, S. (1973). Use of a singlestrand specific nuclease for analysis of bacterial and plasmid deoxyribonucleic acid homo- and heteroduplexes. *J Bacteriol*, *115*, 904–911.

Cruz-Morales, P., Orellana, C. A., Moutafis, G., Moonen, G., Rincon, G., Nielsen, L. K., & Marcellin, E. (2019). Revisiting the Evolution and Taxonomy of Clostridia, a Phylogenomic Update. *Genome Biol Evol*, *11*(7), 2035–2044. https://doi.org/10.1093/gbe/evz096

Cummings, J., & Macfarlane, G. (1991). The colonic flora, fermentation and large bowel digestive function. The Large Intestine: Physiology, Pathophysiology and Disease. *Raven Press*, 51–92.

Cunningham, A. L., Stephens, J. W., & Harris, D. A. (2021). Intestinal microbiota and their metabolic contribution to type 2 diabetes and obesity. *Journal of Diabetes & Metabolic Disorders*, *20*(2), 1855–1870. https://doi.org/10.1007/s40200-021-00858-4

Curtis, M., & Way, S. (2009). Interleukin-17 in host defence against bacterial, mycobacterial and fungal pathogens. *Immunology*, *126*(2), 177–185. https://doi.org/10.1111/j.1365-2567.2008.03017.x.

Daniel, E. J. ., Las Rivas, M., Lira-Navarrete, E., García-García, A., Hurtado-Guerrero, R., Clausen, H., & Gerken, T. A. (2020). Ser and Thr acceptor preferences of the GalNAc-Ts vary among isoenzymes to modulate mucin-type O-glycosylation. *Glycobiology*, *30*(11), 910–922. https://doi.org/10.1093/glycob/cwaa036

De Ley, J. (1970). Re-examination of the association between melting point, buoyant density, and chemical base composition of deoxyribonucleic acid. *J Bacteriol*, *101*, 738–754.

de Vos, W. M., & De Vos, E. A. J. (2012). Role of the intestinal microbiome in health and disease: From correlation to causation. *Nutrition Reviews*, *70*(SUPPL. 1), 45–56. https://doi.org/10.1111/j.1753-4887.2012.00505.x

Derrien, M., Vaughan, E. E., Plugge, C. M., & Vos, W. M. De. (2004). *Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium*. 1469–1476. https://doi.org/10.1099/ijs.0.02873-0

Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Stappenbeck, T. S., & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, *167*(5), 1339–1353. https://doi.org/10.1016/j.cell.2016.10.043

Di Tommaso, N., Gasbarrini, A., & Ponziani, F. R. (2021). Intestinal Barrier in Human Health and Disease. *Int J Environ Res Public Health*, *18*(23), 12836. https://doi.org/10.3390/ijerph182312836

Donaldson, G. P., Lee, S. M., & Mazmanian, S. K. (2015). Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology*, *14*(1), 20–32. https://doi.org/10.1038/nrmicro3552

Duncan, K., Carey-ewend, K., & Vaishnava, S. (2021). Spatial analysis of gut microbiome reveals a distinct ecological niche associated with the mucus layer. *Gut Microbes*, *00*(00), 1–21. https://doi.org/10.1080/19490976.2021.1874815

Egan, M., Dempsey, E., Ryan, C. A., Ross, R. P., & Stanton, C. (2021). The Sporobiota of the Human Gut. *Gut Microbes*, *13*(1), 1–17. https://doi.org/10.1080/19490976.2020.1863134

Egert, M., Graaf, A. A. De, Smidt, H., Vos, W. M. De, & Venema, K. (2006). Beyond diversity : functional micro- biomics of the human colon. *Trends Microbiol*, *14*(2), 86–91. https://doi.org/10.1016/j.tim.2005.12.007

Engevik, M. A., Engevik, A. C., Engevik, K. A., Auchtung, J. M., Chang-graham, A. L., Ruan, W., Luna, R. A., Hyser, J. M., Spinler, J. K., & Versalovic, J. (2020). Mucin-Degrading Microbes Release Monosaccharides That Chemoattract Clostridioides difficile and Facilitate Colonization

of the Human Intestinal Mucus Layer. *ACS Infect Dis*, *7*(5), 1126–1142. https://doi.org/10.1021/acsinfecdis.0c00634

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J., Druart, C., Bindels, L., Guiot, Y., Derrien, M., Muccioli, G., Delzenne, N., de Vos, W., & Cani, P. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A*, *110*(22), 9066–9071. https://doi.org/10.1073/pnas.1219451110

Fan, Y., & Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology*, *19*(1), 55–71. https://doi.org/10.1038/s41579-020-0433-9

Favier, C. F., Vaughan, E. E., De Vos, W. M., & Akkermans, A. D. L. (2002). Molecular monitoring of succession of bacterial communities in human neonates. *Applied and Environmental Microbiology*, *68*(1), 219–226. https://doi.org/10.1128/AEM.68.1.219-226.2002

Fernández-Tomé, S., Ortega Moreno, L., Chaparro, M., & Gisbert, J. (2021). Gut Microbiota and Dietary Factors as Modulators of the Mucus Layer in Inflammatory Bowel Disease. *Int J Mol Sci*, *22*(19), 10224. https://doi.org/10.3390/ijms221910224

Flaherty, S. O., Saulnier, D., Pot, B., & Versalovic, J. (2010). How can probiotics and prebiotics impact mucosal immunity ? *Gut Microbes*, *1*(5), 293–300. https://doi.org/10.4161/gmic.1.5.12924

Fodor, A. A., DeSantis, T. Z., Wylie, K. M., Badger, J. H., Ye, Y., Hepburn, T., Hu, P., Sodergren, E., Liolios, K., Huot-Creasy, H., Birren, B. W., & Earl, A. M. (2012). The "most wanted" taxa from the human microbiome for whole genome sequencing. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0041294

Fujita, M., Tsuchida, A., Hirata, A., Kobayashi, N., Goto, K., Osumi, K., Hirose, Y., Nakayama, J., Yamanoi, T., Ashida, H., & Mizuno, M. (2011). Glycoside Hydrolase Family 89 α-N-acetylglucosaminidase from Clostridium perfringens Specifically Acts on GlcNAcα1,4Galβ1R at the Non-reducing Terminus of O -Glycans in Gastric Mucin. *J Biol Chem*, *286*(8), 6479–6489. https://doi.org/10.1074/jbc.M110.206722

Gao, B., Chi, L., Zhu, Y., Shi, X., Tu, P., Li, B., Yin, J., & Gao, N. (2021). An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies. *Biomolecules*, *11*(4), 530. https://doi.org/10.3390/biom11040530

Garcia-Vallvé, S., Palau, J., & Romeu, A. (1999). Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in Escherichia coli and Bacillus subtilis. *Mol Biol Evol*, *16*(9), 1125–1134. https://doi.org/10.1093/oxfordjournals.molbev.a026203

Gareau, M., Sherman, P., & Walker, W. (2010). Probiotics and the gut microbiota in intestinal health and disease. *Nat Rev Gastroenterol Hepatol*, *7*(9), 503–514.

https://doi.org/10.1038/nrgastro.2010.117.

Godl, K., Johansson, M. E. V., Lidell, M. E., Mörgelin, M., Karlsson, H., Olson, F. J., Gum, J. R., Kim, Y. S., & Hansson, G. C. (2002). The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *Journal of Biological Chemistry*, *277*(49), 47248–47256. https://doi.org/10.1074/jbc.M208483200

Gomila, M., Peña, A., Mulet, M., Lalucat, J., & García-Valdés, E. (2015). Phylogenomics and systematics in Pseudomonas. *Front Microbiol*, *6*, 214. https://doi.org/10.3389/fmicb.2015.00214

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*(1), 81–91. https://doi.org/10.1099/ijs.0.64483-0

Gorkiewicz, G. (2018). Gut microbiome : a new player in gastrointestinal disease. *Virchows Arch*, *472*(1), 159–172. https://doi.org/10.1007/s00428-017-2277-x

Grech, A., Collins, C. E., Holmes, A., Lal, R., Duncanson, K., Taylor, R., & Gordon, A. (2021). Maternal exposures and the infant gut microbiome: a systematic review with meta-analysis. *Gut Microbes*, *13*(1), 1–30. https://doi.org/10.1080/19490976.2021.1897210

Grimont, P., Popoff, M., Grimont, F., Coynault, C., & M, L. (1980). Reproducibility and correlation study of three deoxyribonucleic acid hybridization procedures. *Curr Microbiol*, *4*, 325–330.

Grootaert, H., van Landuyt, L., Hulpiau, P., & Callewaert, N. (2020). Functional exploration of the GH29 fucosidase family. *Glycobiology*, *30*(9), 735–745. https://doi.org/10.1093/glycob/cwaa023

Guan, Q. (2019). A Comprehensive Review and Update on the Pathogenesis of Inflammatory Bowel Disease. *Journal of Immunology Research*, *2019*. https://doi.org/10.1155/2019/7247238

Guan, Z., & Goldfine, H. (2021). Lipid diversity in Clostridia. *Biochimica et Biophysica Acta Mol Cell Biol Lipids*. https://doi.org/10.1016/j.bbalip.2021.158966

Guo, P., Zhang, K., Ma, X., & He, P. (2020). Clostridium species as probiotics : potentials and challenges. *J Anim Sci Biotechnol*, *1*, 1–10. https://doi.org/J Anim Sci Biotechnol

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hale, A., Kirby, J. E., & Albrecht, M. (2016). Fatal spontaneous Clostridium bifermentans necrotizing endometritis: A case report and literature review of the pathogen. *Open Forum Infectious Diseases*, *3*(2), 1–3. https://doi.org/10.1093/ofid/ofw095

Hall, A. B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G. K., Vatanen, T., Fornelos, N., Wilson, R., Bertha, M., Cohen, M., Garber, J., Khalili, H., Gevers, D., Ananthakrishnan, A. N., Kugathasan, S., Lander, E. S., Blainey, P., … Huttenhower, C. (2017). A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med.*, *9*(103). https://doi.org/10.1186/s13073-017-0490-5

Hamer, H., De Preter, V., Windey, K., & Verbeke, K. (2012). Functional analysis of colonic bacterial metabolism: relevant to health? *Am J Physiol Gastrointest Liver Physiol*, *302*(1), G1-9. https://doi.org/10.1152/ajpgi.00048.2011

Hansson, G. C. (2020). Mucins and the Microbiome. *Annu Rev Biochem*, *89*, 769–793. https://doi.org/10.1146/annurev-biochem-011520-105053

Hansson G.C. (2012). Role of mucus layers in gut infection and inflammation. *Curr Opin Microbiol*, *15*(1), 57–62. https://doi.org/10.1016/j.mib.2011.11.002

Haran, J., & McCormick, B. . (2020). Aging, Frailty, and the Microbiome—How Dysbiosis Influences Human Aging and Disease. *Gastroenterology*, *160*(2), 507–523. https://doi.org/https://doi.org/10.1053/j.gastro.2020.09.060

Hasnain, S., Tauro, S., Das, I., Tong, H., Chen, A., Jeffery, P., McDonald, V., Florin, T., & McGuckin, M. (2013). IL-10 Promotes Production of Intestinal Mucus by Suppressing Protein Misfolding and Endoplasmic Reticulum Stress in Goblet Cells. *Gastroenterology*, *144*(2), 357–368. https://doi.org/10.1053/j.gastro.2012.10.043

Hauschild, A. H. W., & Holdeman, L. V. (1974). Clostridium celatum sp.nov., isolated from normal human feces. *International Journal of Systematic Bacteriology*, *24*(4), 478–481. https://doi.org/10.1099/00207713-24-4-478

Hayashi Sant'Anna, F., Bach, E., Porto, R. Z., Guella, F., Hayashi Sant'Anna, E., & Passaglia, L. M. P. (2019). Genomic metrics made easy: what to do and where to go in the new era of bacterial taxonomy. *Critical Reviews in Microbiology*, *45*(2), 182–200. https://doi.org/10.1080/1040841X.2019.1569587

Heintz-Buschart, A., & Wilmes, P. (2018). Human Gut Microbiome: Function Matters. *Trends in Microbiology*, *26*(7), 563–574. https://doi.org/10.1016/j.tim.2017.11.002

Henry, C. H. (1918). An investigation of the cultural reactions of certain anaerobes found in wounds. *The Journal of Pathology and Bacteriology*, *21*(3), 344–385. https://doi.org/10.1002/path.1700210304

Hollister, E. B., Riehle, K., Luna, R. A., Weidler, E. M., Rubio-Gonzales, M., Mistretta, T. A., Raza, S., Doddapaneni, H. V., Metcalf, G. A., Muzny, D. M., Gibbs, R. A., Petrosino, J. F., Shulman, R. J., & Versalovic, J. (2015). Structure and function of the healthy pre-adolescent pediatric gut

microbiome. *Microbiome*, *3*, 36. https://doi.org/10.1186/s40168-015-0101-x

Holmén-Larsson, J. M., Karlsson, H., Sj, H., & Hansson, G. C. (2009). A complex , but uniform O - glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MS. *Glycobiology*, *19*(7), 756–766. https://doi.org/10.1093/glycob/cwp048

Holmén Larsson, J.M., Thomsson, K.A., Rodríguez-Piñeiro, A.M., Karlsson, H. & Hansson, G. C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *Am. J. Physiol. Gastrointest. Liver Physiol*, *305*, 357–363. https://doi.org/10.1152/ajpgi.00048.2013

Howe, C., MacLennan, J. D., Mandl, I., & Kabat, E. A. (1957). Enzymes of Clostridium tertium: effects on blood group and virus receptor substances. *J Bacteriol*, *74*(3), 365–376. https://doi.org/10.1128/jb.74.3.365-376.1957

Hu, Z., Patel, I. R., & Mukherjee, A. (2013). Genetic analysis of the roles of agaA , agaI , and agaS genes in the N-acetyl-D-galactosamine and D-galactosamine catabolic pathways in Escherichia coli strains O157 : H7 and C. *BMC Microbiol*, *13*, 94. https://doi.org/10.1186/1471-2180-13-94

Iseki, S., & Okada, S. (1951). *On a Specific Enzym Which Decomposes Group A Substances*. *8*, 6–9.

Ivanov, I., Atarashi, K., Manel, N., Brodie, E., Shima, T., Karaoz, U., Wei, D., Goldfarb, K., Santee, C., Lynch, S., Tanoue, T., Imaoka, A., Itoh, K., Takeda, K., Umesaki, Y., Honda, K., & Littman, D. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*, *139*(3), 485–498. https://doi.org/10.1016/j.cell.2009.09.033

Ivanov, I., & Honda, K. (2012). Intestinal commensal microbes as immune modulators. *Cell Host Microbe*, *12*(4), 496–508. https://doi.org/10.1016/j.chom.2012.09.009

Jacobse, J., Li, J., Rings, E., Samsom, J., & Goettel, J. (2021). Intestinal Regulatory T Cells as Specialized Tissue-Restricted Immune Cells in Intestinal Immune Homeostasis and Disease PMC8371910. *Front Immunol*, *12*, 716499. https://doi.org/10.3389/fimmu.2021.716499

Jensen, P. H., Kolarich, D., & Packer, N. H. (2010). Mucin-type O-glycosylation - Putting the pieces together. *FEBS Journal*, *277*(1), 81–94. https://doi.org/10.1111/j.1742-4658.2009.07429.x

Johansson, M. E. V., Holmén Larsson, J. M., & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4659–4665. https://doi.org/10.1073/pnas.1006451107

Josenhans, C., Müthing, J., Elling, L., Bartfeld, S., & Schmidt, H. (2020). How bacterial pathogens of the gastrointestinal tract use the mucosal glyco- code to harness mucus and microbiota : New ways to study an ancient bag of tricks. *International Journal of Medical Microbiology*, *310*(December 2019). https://doi.org/10.1016/j.ijmm.2020.151392

Juge, N., Tailford, L., & Owen, C. D. (2016). Sialidases from gut bacteria : a mini-review. *Biochem Soc Trans.*, *44*(1), 166–175. https://doi.org/10.1042/BST20150226

Jyothsna, T. S. S., Tushar, L., Sasikala, C., & Ramana, C. V. (2016a). *isolated from marine sediment and reclassification of Clostridium bifermentans as Paraclostridium bifermentans comb . nov . Proposal of a new genus Paeniclostridium gen . nov . to accommodate Clostridium sordellii and Clostridium ghonii*. 1268–1274. https://doi.org/10.1099/ijsem.0.000874

Jyothsna, T. S. S., Tushar, L., Sasikala, C., & Ramana, C. V. (2016b). Paraclostridium benzoelyticum gen. nov., sp. nov., isolated from marine sediment and reclassification of Clostridium bifermentans as Paraclostridium bifermentans comb. nov. Proposal of a new genus Paeniclostridium gen. nov. to accommodate Clostridium sord. *Int J Syst Evol Microbiol*, *66*(3), 1268–1274. https://doi.org/10.1099/ijsem.0.000874

Kabeerdoss, J., Sankaran, V., Pugazhendhi, S., & Ramakrishna, B. S. (2013). Clostridium leptum group bacteria abundance and diversity in the fecal microbiota of patients with inflammatory bowel disease: a case – control study in India. *BMC Gastroenterology*, *13*(1), 1. https://doi.org/10.1186/1471-230X-13-20

Kanehisa, M., & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science*, *29*(1), 28–35. https://doi.org/10.1002/pro.3711

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kaoutari, A. El, Armougom, F., Gordon, J. I., Raoult, D., & Henrissat, B. (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Publishing Group*, *11*(July), 497–504. https://doi.org/10.1038/nrmicro3050

Kelsen, J. R., Wu, G. D., Kelsen, J. R., & Wu, G. D. (2012). The gut microbiota, environment and diseases of modern society. *Gut Microbes*, *3*(4), 374–382. https://doi.org/10.4161/gmic.21333

Kidess, E., Kleerebezem, M., & Brugman, S. (2021). Colonizing Microbes , IL-10 and IL-22 : Keeping the Peace at the Mucosal Surface. *Front Microbiol*, *12*(September). https://doi.org/10.3389/fmicb.2021.729053

Kim, J., Hetzel, M., Boiangiu, C. D., & Buckel, W. (2004). Dehydration of (R)-2-hydroxyacyl-CoA to enoyl-CoA in the fermentation of alpha-amino acids by anaerobic bacteria. *FEMS Microbiol. Rev*, *28*, 455–468. https://doi.org/10.1016/j.femsre.2004.03.001

Kim, M., Oh, H. S., Park, S. C., & Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, *64*(PART 2),

346–351. https://doi.org/10.1099/ijs.0.059774-0

Kim, S., Oh, D., & Kang, H. A. (2011). Features and applications of bacterial sialidases. *Appl Microbiol Biotechnol*, *91*(1), 1–15. https://doi.org/10.1007/s00253-011-3307-2

Kim, Y. S., & Ho, S. B. (2010). Intestinal goblet cells and mucins in health and disease: Recent insights and progress. *Current Gastroenterology Reports*, *12*(5), 319–330. https://doi.org/10.1007/s11894-010-0131-2

Kint, N., Alves Feliciano, C., Martins, M., Morvan, C., Fernandes, S., Folgosa, F., Dupuy, B., Texeira, M., & Martin-Verstraete, I. (2020). How the Anaerobic Enteropathogen Clostridioides difficile Tolerates Low O2 Tensions. *MBio*, *11*(5), e01559-20. https://doi.org/10.1128/mBio.01559-20.

Kiu, R., Caim, S., Alcon-Giner, C., Belteki, G., Clarke, P., Pickard, D., Dougan, G., & Hall, L. J. (2017). Preterm infant-associated clostridium tertium, clostridium cadaveris, and clostridium paraputrificum strains: Genomic and evolutionary insights. *Genome Biology and Evolution*, *9*(10), 2707–2714. https://doi.org/10.1093/gbe/evx210

Kiyohara M, Nakatomi T, Kurihara S, Fushinobu S, Suzuki H, Tanaka T, Shoda SI, Kitaoka M, Katayama T, Yamamoto K, A. H. (2012). α-N-acetylgalactosaminidase from infant-associated bifidobacteria belonging to novel glycoside hydrolase family 129 is implicated in alternative mucin degradation pathway. *J Biol Chem*, *287*(1), 693–700. https://doi.org/10.1074/jbc.M111.277384

Kobayashi, J. (2018). Nitric Oxide Effect of diet and gut environment on the gastrointestinal formation of N -nitroso compounds : A review. *Nitric Oxide*, *73*, 66–73. https://doi.org/10.1016/j.niox.2017.06.001

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., & Knight, R. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*, *108*, 4578–4585. https://doi.org/10.1073/pnas.1000081107

Konstantinidis, K. T., Ramette, A., & Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1475), 1929–1940. https://doi.org/10.1098/rstb.2006.1920

Konstantinidis, K. T., & Tiedje, J. M. (2005). Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology*, *187*(18), 6258–6264. https://doi.org/10.1128/JB.187.18.6258-6264.2005

Konstantinidis, K., & Tiedje, J. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A*, *102*(7), 2567–2572. https://doi.org/10.1073/pnas.0409727102

Koropatkin NM, Cameron EA, M. E. (2012). How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology*, *10*(5), 323–335. https://doi.org/10.1038/nrmicro2746

Kostic, A., Xavier, R., & Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology*, *146*(6), 1489–1499. https://doi.org/10.1053/j.gastro.2014.02.009

Koutsioulis, D., Landry, D., & Guthrie, E. P. (2008). Novel endo- α - N -acetylgalactosaminidases with broader substrate specificity. *Glycobiology*, *18*(10), 799–805. https://doi.org/10.1093/glycob/cwn069

Kudelka, M. R., Stowell, S. R., Cummings, R. D., & Neish, A. S. (2020). Intestinal epithelial glycosylation in homeostasis and gut microbiota interactions in IBD. *Nature Reviews Gastroenterology and Hepatology*, *17*(10), 597–617. https://doi.org/10.1038/s41575-020-0331-7

Kutsuna, R., Tomida, J., Miyoshi-akiyama, T., Mori, K., Hayashi, M., Morita, Y., Tanaka, K., & Kawamura, Y. (2019). Description of Paraclostridium bifermentans subsp . muricolitidis subsp . nov ., emended description of Paraclostridium bifermentans (Sasi Jyothsna et al ., 2016), and creation of Paraclostridium bifermentans subsp . bifermentans subsp . nov. *Microbiol Immunol*, *63*(1), 1–10. https://doi.org/10.1111/1348-0421.12663

Larsen, M. V., Cosentino, S., Lukjancenko, O., Saputra, D., Rasmussen, S., Hasman, H., Sicheritz-Pontén, T., Aarestrup, F. M., Ussery, D. W., & Lund, O. (2014). Benchmarking of methods for genomic taxonomy. *Journal of Clinical Microbiology*, *52*(5), 1529–1539. https://doi.org/10.1128/JCM.02981-13

Larsson, J. M. H., Karlsson, H., Crespo, J. G., Johansson, M. E. V., Eklund, L., Sjövall, H., & Hansson, G. C. (2011). Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflamm Bowel Dis*, *17*(11), 2299–2307. https://doi.org/10.1002/ibd.21625

Lawson, P. A., & Rainey, F. A. (2016). Proposal to restrict the genus Clostridium Prazmowski to Clostridium butyricum and related species. *Int J Syst Evol Microbiol*, *66*(2), 1009–1016. https://doi.org/10.1099/ijsem.0.000824

Lee, J. S., Wang, R. X., Alexeev, E. E., & Colgan, S. P. (2021). Intestinal Inflammation as a Dysbiosis of Energy Procurement: New Insights into an Old Topic. *Gut Microbes*, *13*(1), 1–20. https://doi.org/10.1080/19490976.2021.1880241

Leja, K., Myszka, K., & Czaczyk, K. (2013). The ability of Clostridium bifermentans strains to lactic acid biosynthesis in various environmental conditions. *SpringerPlus*, *2*(1), 1–8.

https://doi.org/10.1186/2193-1801-2-44

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*, *47*(W1), W256-259. https://doi.org/10.1093/nar/gkz239

Lewis, A. L., & Lewis, W. G. (2012). Microreview Host sialoglycans and bacterial sialidases : a mucosal perspective. *Cell Microbiol*, *14*(8), 1174–1182. https://doi.org/10.1111/j.1462-5822.2012.01807.x

Li, K., Bihan, M., Yooseph, S., & Methe, B. A. (2012). Analyses of the Microbial Diversity across the Human Microbiome. *PLoS One*, *7*(6). https://doi.org/10.1371/Citation

Li, Y., Jin, Y., Zhang, J., Pan, H., Wu, L., Liu, D., Liu, J., Hu, J., & Shen, J. (2021). Recovery of human gut microbiota genomes with third-generation sequencing. *Cell Death and Disease*, *12*(6), 569. https://doi.org/10.1038/s41419-021-03829-y

Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V., & McGuckin, M. A. (2008). Mucins in the mucosal barrier to infection. *Mucosal Immunology*, *1*(3), 183–197. https://doi.org/10.1038/mi.2008.5

Littman, D. R., & Rudensky, A. Y. (2010). Th17 and Regulatory T Cells in Mediating and Restraining Inflammation. *Cell*, *140*(6), 845–858. https://doi.org/10.1016/j.cell.2010.02.021

Liu, C., Du, M.-X., Abuduaini, R., Yu, H.-Y., Li, D.-H., Wang, Y.-J., Zhou, N., Jiang, M.-Z., Niu, P.-X., Han, S.-S., Chen, H.-H., Shi, W.-Y., Wu, L., Xin, Y.-H., Ma, J., Zhou, Y., Jiang, C.-Y., Liu, H.-W., & Liu, S.-J. (2021). Enlightening the taxonomy darkness of human gut microbiomes with a cultured biobank. *Microbiome*, *9*(1), 119. https://doi.org/10.1186/s40168-021-01064-3

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*, *42*(November 2013), 490–495. https://doi.org/10.1093/nar/gkt1178

Lopes-Santos, L., Castro, D., Ferreira-Tonin, M., Corrêa, D., Weir, B., Park, D., Ottoboni, L., Neto, J., & Destéfano, S. (2017). Reassessment of the taxonomic position of Burkholderia andropogonis and description of Robbsia andropogonis gen. nov., comb. nov. *Antonie Van Leeuwenhoek*, *110*(6), 727–736. https://doi.org/10.1007/s10482-017-0842-6

Lopetuso, L. R., Scaldaferri, F., Petito, V., & Gasbarrini, A. (2013). Commensal Clostridia: Leading players in the maintenance of gut homeostasis. *Gut Pathogens*, *5*(1), 1. https://doi.org/10.1186/1757-4749-5-23

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

Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic

microbiota. *Environmental Microbiology*, *19*(1), 29–41. https://doi.org/10.1111/1462-2920.13589

Louis, P., Hold, G. L., & Flint, H. J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nature Reviews Microbiology*, *12*(10), 661–672. https://doi.org/10.1038/nrmicro3344

Macfarlane, G. T., & Macfarlane, S. (2011). Fermentation in the Human Large Intestine Its Physiologic Consequences and the Potential Contribution of Prebiotics. *J Clin Gastroenterol*, *45*(December), 120–127.

Macfarlane, S., & Dillon, J. (2007). Microbial biofilms in the human gastrointestinal tract. *J Appl Microbiol*, *102*(5), 1187–1196. https://doi.org/10.1111/j.1365-2672.2007.03287.x.

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., Quraishi, M. N., Kinross, J., Smidt, H., Tuohy, K. M., Thomas, L. V., Zoetendal, E. G., & Hart, A. (2016). The gut microbiota and host health: A new clinical frontier. *Gut*, *65*(2), 330–339. https://doi.org/10.1136/gutjnl-2015-309990

Marcobal, A., Southwick, A. M., Earle, K. A., & Sonnenburg, J. L. (2013). A refined palate: Bacterial consumption of host glycans in the gut. *Glycobiology*, *23*(9), 1038–1046. https://doi.org/10.1093/glycob/cwt040

Mariat, D., Firmesse, O., Levenez, F., Guimarães, V. D., Sokol, H., Doré, J., Corthier, G., & Furet, J. P. (2009). The firmicutes/bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiology*, *9*, 1–6. https://doi.org/10.1186/1471-2180-9-123

Martens, E. C., Chiang, H. C., & Gordon, J. I. (2009). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe*, *4*(5), 447–457. https://doi.org/10.1016/j.chom.2008.09.007.Mucosal

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/https://doi.org/10.14806/ej.17.1.200.

Matijaši, M., Peri, M., Cipˇ, H., Panek, M., Vraneši, D., & Keleˇ, D. L. (2016). Modulating Composition and Metabolic Activity of the Gut Microbiota in IBD Patients. *Int J Mol Sci*, *17*(4), 578. https://doi.org/10.3390/ijms17040578

McDole, J. R., Wheeler, L. W., McDonald, K. G., Wang, B., Konjufca, V., Knoop, K. A., Newberry, R. D., & Miller, M. J. (2012). Goblet cells deliver luminal anrigen to CD103+ DCs in the small intestine. *Nature*, *483*(7389), 345–349. https://doi.org/10.1038/nature10863.Goblet

Mcguckin, M. A., Lindén, S. K., Sutton, P., & Florin, T. H. (n.d.). FOCUS ON MUCOSAL MICROB Mucin dynamics and enteric pathogens. *Nature Publishing Group*. https://doi.org/10.1038/nrmicro2538

Meier-Kolthoff, J. ., Auch, A. F., Klenk, H.-P., & Göker, M. (2013). Genome sequence-based species

delimitation with confidence intervals and improved distance functions. *BMC Bioinformatic*, *14*, 60. https://doi.org/10.1186/1471-2105-14-60

Melhem, H., Regan-komito, D., & Niess, J. (2021). Mucins Dynamics in Physiological and Pathological Conditions. *Int J Mol Sci*, *22*(24), 13642. https://doi.org/10.3390/ijms222413642

Meng, C., Bai, C., Brown, T. D., Hood, L. E., & Tian, Q. (2018). Human Gut Microbiota and Gastrointestinal Cancer. *Genomics, Proteomics and Bioinformatics*, *16*(1), 33–49. https://doi.org/10.1016/j.gpb.2017.06.002

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arboleya Montes, S., Mancabelli, L., Lugli, G. A., Rodriguez, J. M., Bode, L., de Vos, W., Gueimonde, M., Margolles, A., van Sinderen, D., & Ventura, M. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and Molecular Biology Reviews*, *81*(4), 1–67. https://doi.org/10.1128/mmbr.00036-17

Milano, V., Biehle, L., Patel, S., & Hammer, J. (2019). Clostridium tertium bacteremia and hepatic abscess in a non-neutropenic patient. *IDCases*, e00510. https://doi.org/10.1016/j.idcr.2019.e00510

Miller, D. L., Brazer, S., Murdoch, D., Reller, L. B., & Corey, G. R. (2001). Significance of Clostridium tertium bacteremia in neutropenic and nonneutropenic patients: Review of 32 cases. *Clinical Infectious Diseases*, *32*(6), 975–978. https://doi.org/10.1086/319346

Miller, R. S., & Hoskins, L. C. (1981). Mucin degradation in human colon ecosystems. Fecal population densities of mucin-degrading bacteria estimated by a "most probable number" method. *Gastroenterology*, *81*(4), 759–765. https://doi.org/10.1016/0016-5085(81)90503-5

Mishra, A., Lai, G. C., Yao, L. J., Aung, T. T., Shental, N., Rotter-Maskowitz, A., Shepherdson, E., Singh, G. S. N., Pai, R., Shanti, A., Wong, R. M. M., Lee, A., Khyriem, C., Dutertre, C. A., Chakarov, S., Srinivasan, K. G., Shadan, N. B., Zhang, X.-M., Khalilnezhad, S., … Ginhoux, F. (2021). Microbial exposure during early human development primes fetal immune cells. *Cell*, *184*(13), 3394-3409.e20. https://doi.org/10.1016/j.cell.2021.04.039

Moncada, D. M., Kammanadiminti, S. J., & Chadee, K. (2003). Mucin and Toll-like receptors in host defense against intestinal parasites. *Trends Parasitol*, *19*(7), 305–311. https://doi.org/10.1016/S1471-4922(03)00122-3

Moran-Ramos, S., Lopez-Contreras, B. E., Villarruel-Vazquez, R., Ocampo-Medina, E., Macias-Kauffer, L., Martinez-Medina, J. N., Villamil-Ramirez, H., León-Mimila, P., Del Rio-Navarro, B. E., Ibarra-Gonzalez, I., Vela-Amieva, M., Gomez-Perez, F. J., Velazquez-Cruz, R., Salmeron, J., Reyes-Castillo, Z., Aguilar-Salinas, C., & Canizales-Quinteros, S. (2020). Environmental and

intrinsic factors shaping gut microbiota composition and diversity and its relation to metabolic health in children and early adolescents: A population-based study. *Gut Microbes*, *11*(4), 900–917. https://doi.org/10.1080/19490976.2020.1712985

Morvan, C., Folgosa, F., Kint, N., Teixeira, M., & Martin-Verstraete, I. (2021). Responses of Clostridia to oxygen: from detoxification to adaptive strategies. *Environmental Microbiology*, *23*(8), 4112–4125. https://doi.org/10.1111/1462-2920.15665

Mukherjee, A. (2020). Gut microbes from the phylogenetically diverse genus Eubacterium and their various contributions to gut health. *Gut Microbes*, *12*(1), 1–28. https://doi.org/10.1080/19490976.2020.1802866

Nagano, Y., & Itoh, K. (2012). The induction of Treg cells by gut-indigenous Clostridium. *Current Opinion in Immunology*, *24*(4), 392–397. https://doi.org/10.1016/j.coi.2012.05.007

Narushima, S., Sugiura, Y., Oshima, K., Atarashi, K., Hattori, M., Suematsu, M., & Honda, K. (2014). Characterization of the 17 strains of regulatory T cell-inducing human-derived Clostridia. *Gut Microbes*, *5*(3), 333–339. https://doi.org/10.4161/gmic.28572

Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., & Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. *Nature*, *568*(7753), 505–510. https://doi.org/10.1038/s41586-019-1058-x

Newman, T. M., Shively, C. A., Register, T. C., Appt, S. E., Yadav, H., Colwell, R. R., Fanelli, B., Dadlani, M., Graubics, K., Nguyen, U. T., Ramamoorthy, S., Uberseder, B., Clear, K. Y. J., Wilson, A. S., Reeves, K. D., Chappell, M. C., & Tooze, J. A. (2021). Diet, obesity, and the gut microbiome as determinants modulating metabolic outcomes in a non-human primate model. *Microbiome*, *9*(1), 1–17. https://doi.org/10.1186/s40168-021-01069-y

Nishiyama, K., Nagai, A., Uribayashi, K., Yamamoto, Y., & Mukai, T. (2018). Two extracellular sialidases from Bifi dobacterium bifidum promote the degradation of sialyl-oligosaccharides and support the growth of Bifidobacterium breve. *Anaerobe*, *52*, 22–28. https://doi.org/10.1016/j.anaerobe.2018.05.007

Nye, T., Liò, P., & Gilks, W. (2006). A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics*, *22*, 117–119. https://doi.org/10.1093/bioinformatics/bti720

O'Hara, A. M., & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, *7*(7), 688–693. https://doi.org/10.1038/sj.embor.7400731

Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J. Z., Abe, F., & Osawa, R. (2016). Age-related changes in gut microbiota composition from newborn to centenarian: A cross-sectional study. *BMC Microbiology*, *16*(1), 1–12. https://doi.org/10.1186/s12866-016-

0708-5

Ottman, N., Davids, M., Suarez-Diez, M., Boeren, S., Schaap, P. J., Martins Dos Santos, V. A. P., Smidt, H., Belzer, C., & de Vos, W. M. (2017). Genome-Scale Model and Omics Analysis of Metabolic Capacities of Akkermansia muciniphila Reveal a Preferential Mucin-Degrading Lifestyle. *Appl Environ Microbiol.*, *83*(18), 1–15. https://doi.org/10.1128/AEM.01014-17

Ottman, N., Smidt, H., Vos, W. M. De, & Belzer, C. (2012). The function of our microbiota : who is out there and what do they do ? *Ront Cell Infect Microbiol*, *2*(August), 1–11. https://doi.org/10.3389/fcimb.2012.00104

Ouwerkerk, J. P., Vos, W. M. De, & Belzer, C. (2013). Glycobiome : Bacteria and mucus at the epithelial interface. *Best Pract Res Clin Gastroenterol*, *27*(1), 25–38. https://doi.org/10.1016/j.bpg.2013.03.001

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary : rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Paliy, O., Kenche, H., Abernathy, F., & Michail, S. (2009). High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. *Applied and Environmental Microbiology*, *75*(11), 3572–3579. https://doi.org/10.1128/AEM.02764-08

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biology*, *5*(7), 1556–1573. https://doi.org/10.1371/journal.pbio.0050177

Pannekoek, Y., Qi-Long, Q., Zhang, Y., & van der Ende, A. (2016). Genus delineation of Chlamydiales by analysis of the percentage of conserved proteins justifies the reunifying of the genera Chlamydia and Chlamydophila into one single genus Chlamydia. *Pathog Dis*, *74*(6). https://doi.org/10.1093/femspd/ftw071

Paone, P., & Cani, P. D. (2020). Mucus barrier , mucins and gut microbiota : the expected slimy partners ? *Gut*, *69*(12), 2232–2243. https://doi.org/10.1136/gutjnl-2020-322260

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM : assessing the quality of microbial genomes recovered from isolates , single cells , and metagenomes. *Genome Res*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114.Freely

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, *2*(11), 1533–1542. https://doi.org/10.1038/s41564-017-0012-7

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P.,

Tett, A., Ghensi, P., Collado, M., Rice, B., DuLong, C., Morgan, X., Golden, C., Quince, C., Huttenhower, C., & Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, *176*(3), 649–662. https://doi.org/10.1016/j.cell.2019.01.001

Passel, M. W. J. Van, Kant, R., Zoetendal, E. G., Plugge, C. M., Derrien, M., Malfatti, S. A., Chain, P. S. G., Woyke, T., Palva, A., & Vos, W. M. De. (2011). The Genome of Akkermansia muciniphila , a Dedicated Intestinal Mucin Degrader , and Its Use in Exploring Intestinal Metagenomes. *PLoS One*, *6*(3). https://doi.org/10.1371/journal.pone.0016876

Pelaseyed, T., Bergström, J., Gustafsson, J., Ermund, A., Birchenough, G., Schütte, A., van der Post, S., Svensson, F., Rodríguez-Piñeiro, A., Nyström, E., Wising, C., Johansson, M., & Hansson, G. (2014). The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev*, *260*(1), 8–20. https://doi.org/10.1111/imr.12182.The

Perez-Lopez, A., Behnsen, J., Nuccio, S. P., & Raffatellu, M. (2016). Mucosal immunity to pathogenic intestinal bacteria. *Nature Reviews Immunology*, *16*(3), 135–148. https://doi.org/10.1038/nri.2015.17

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., Bonazzi, V., McEwen, J. E., Wetterstrand, K. A., Deal, C., Baker, C. C., Di Francesco, V., Howcroft, T. K., Karp, R. W., Lunsford, R. D., Wellington, C. R., Belachew, T., Wright, M., Giblin, C., … Guyer, M. (2009). The NIH Human Microbiome Project. *Genome Research*, *19*(12), 2317–2323. https://doi.org/10.1101/gr.096651.109

Petersson, J., Schreiber, O., Hansson, G. C., Gendler, S. J., Velcich, A., Lundberg, J. O., Roos, S., Holm, L., & Phillipson, M. (2011). Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *Am. J. Physiol. Gastrointest. Liver Physiol*, *300*(2), G327-33. https://doi.org/10.1152/ajpgi.00422.2010

Pierro, F. Di. (2021). Gut Microbiota Parameters Potentially Useful in Clinical Perspective. *Microorganisms*, *9*(11), 2402. https://doi.org/10.3390/microorganisms9112402

Png, C. W., Lind, S. K., Gilshenan, K. S., Zoetendal, E. G., Mcsweeney, C. S., Sly, L. I., Mcguckin, M. A., & Florin, T. H. J. (2010). Mucolytic Bacteria With Increased Prevalence in IBD Mucosa Augment In Vitro Utilization of Mucin by Other Bacteria. *Am J Gastroenterol*, *105*(11), 2420–2428. https://doi.org/10.1038/ajg.2010.281

Pop, M. (2012). We are what we eat: How the diet of infants affects their gut microbiome. *Genome Biology*, *13*(4), 4–6. https://doi.org/10.1186/gb-2012-13-4-152

Popoff, M. R., & Bouvet, P. (2013). Genetic characteristics of toxigenic Clostridia and toxin gene

evolution. *Toxicon*, *75*, 63–89. https://doi.org/10.1016/j.toxicon.2013.05.003

Puértolas-Balint, F., & Schroeder, B. O. (2020). Does an Apple a Day Also Keep the Microbes Away? The Interplay Between Diet, Microbiota, and Host Defense Peptides at the Intestinal Mucosal Barrier. *Frontiers in Immunology*, *11*(June), 1–21. https://doi.org/10.3389/fimmu.2020.01164

Qin, Q., Xie, B., Zhang, X., Chen, X., Zhou, B., Zhou, J., Oren, A., & Zhang, Y. (2014). A Proposed Genus Boundary for the Prokaryotes Based on Genomic. *J Bacteriol*, *196*(12), 2210–2215. https://doi.org/10.1128/JB.01688-14

Quartieri, A., Simone, M., Gozzoli, C., Popovic, M., D'Auria, G., Amaretti, A., Raimondi, S., & Rossi, M. (2016). Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe*, *38*, 130–137. https://doi.org/10.1016/j.anaerobe.2015.10.006

Quercia, S., Candela, M., Giuliani, C., Turroni, S., Luiselli, D., & Rampelli, S. (2014). From lifetime to evolution : timescales of human gut microbiota adaptation. *Front Microbiol.*, *5*(5), 587. https://doi.org/10.3389/fmicb.2014.00587

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Rackaityte, E., Halkias, J., Fukui, E., Mendoza, V., Hayzelden, C., Crawford, E., Fujimura, K., Burt, T., & Lynch, S. (2020). Viable bacterial colonization is highly limited in the human intestine in utero. *Nat Med*, *26*(4), 599–607. https://doi.org/10.1038/s41591-020-0761-3

Raimondi, S., Luciani, R., Sirangelo, T. M., Amaretti, A., Leonardi, A., Ulrici, A., Foca, G., D'Auria, G., Moya, A., Zuliani, V., Seibert, T. M., Søltoft-Jensen, J., & Rossi, M. (2019). Microbiota of sliced cooked ham packaged in modified atmosphere throughout the shelf life: Microbiota of sliced cooked ham in MAP. *International Journal of Food Microbiology*, *289*(September 2018), 200–208. https://doi.org/10.1016/j.ijfoodmicro.2018.09.017

Raimondi, S., Musmeci, E., Candeliere, F., Amaretti, A., & Rossi, M. (2021). Identification of mucin degraders of the human gut microbiota. *Scientific Reports*, *0123456789*, 1–10. https://doi.org/10.1038/s41598-021-90553-4

Raimondi, S., Righini, L., Candeliere, F., Musmeci, E., Bonvicini, F., Gentilomi, G., Starˇ, M., Amaretti, A., & Rossi, M. (2019). Antibiotic Resistance , Virulence Factors , Phenotyping , and Genotyping of E . coli Isolated from the Feces of Healthy Subjects. *Microorganisms*, *7*(8), 251. https://doi.org/10.3390/microorganisms7080251

Rajilić-Stojanović, M., & de Vos, W. M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiology Reviews*, *38*(5), 996–1047.

https://doi.org/10.1111/1574-6976.12075

Ramasamy, D., Mishra, A. K., Lagier, J. C., Padhmanabhan, R., Rossi, M., Sentausa, E., Raoult, D., & Fournier, P. E. (2014). A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *International Journal of Systematic and Evolutionary Microbiology*, *64*(PART 2), 384–391. https://doi.org/10.1099/ijs.0.057091-0

Ravcheev, D. A., & Thiele, I. (2017). Comparative Genomic Analysis of the Human Gut Microbiome Reveals a Broad Distribution of Metabolic Pathways for the Degradation of Host-Synthetized Mucin Glycans and Utilization of Mucin-Derived Monosaccharides. *Front Genet*, *8*. https://doi.org/10.3389/fgene.2017.00111

Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K. P., Flint, H. J., & Louis, P. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME Journal*, *8*(6), 1323–1335. https://doi.org/10.1038/ismej.2014.14

Rescigno, M., Lopatin, U., & Chieppa, M. (2008). Interactions among dendritic cells, macrophages, and epithelial cells in the gut: implications for immune tolerance. *Curr Opin Immunol*, *20*(6), 669–675. https://doi.org/10.1016/j.coi.2008.09.007

Rey, F., Gonzalez, M., Cheng, J., Wu, M., Ahern, P., & Gordon, J. (2013). Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc Natl Acad Sci U S A*, *110*(33), 13582–13587. https://doi.org/10.1073/pnas.1312524110

Richter, M., & Rossello, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A*, *106*(45), 19126–19131.

Ringel-Kulka, T., Cheng, J., Ringel, Y., Salojärvi, J., Carroll, I., Palva, A., de Vos, W. M., & Satokari, R. (2013). Intestinal Microbiota in Healthy U.S. Young Children and Adults-A High Throughput Microarray Analysis. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0064315

Ringot-destrez, B., Alessandro, Z. D., Lacroix, J., Mercier-bonin, M., Renaud, L., & Robbe-masselot, C. (2018). A Sensitive and Rapid Method to Determine the Adhesion Capacity of Probiotics and Pathogenic Microorganisms to Human Gastrointestinal Mucins. *Microorganisms*, *6*(2), 49. https://doi.org/10.3390/microorganisms6020049

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W. T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., … Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, *499*(7459), 431–437. https://doi.org/10.1038/nature12352

Robbe, C., Capon, C., Coddeville, B., & Michalski, J. C. (2004). Structural diversity and specific

distribution of O-glycans in normal human mucins along the intestinal tract. *Biochemical Journal*, *384*(2), 307–316. https://doi.org/10.1042/BJ20040605

Robbe, C., Capon, C., Maes, E., Rousset, M., Zweibaum, A., Zanetta, J., Michalski, J., Upmc, I. U., & Biome, I. (2003). Evidence of Regio-specific Glycosylation in Human. *Journal of Biological Chemistry*, *278*(47), 46337–46348. https://doi.org/10.1074/jbc.M302529200

Rodriguez-R, LM Konstantinidis, K. (2014). Bypassing cultivation to identify bacterial species. *Microbe*, *9*(3), 111–118. https://doi.org/10.1128/microbe.9.111.1

Rodriguez-R, L., & Konstantinidis, K. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *Per J Preprints.*, *4*(e1900v1). https://doi.org/10.7287/peerj.preprints.1900v1

Rodriguez-R, L. M., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R., & Konstantinidis, K. T. (2020). Classifying prokaryotic genomes using the Microbial Genomes Atlas (MiGA) webserver. *Bergey's Manual of Systematics of Archaea and Bacteria*. https://doi.org/10.1002/9781118960608.bm00042

Rooks, M. G., & Garrett, W. S. (2016). Gut microbiota, metabolites ad host immunity. *Nat Rev Immunol*, *16*(6), 341–352. https://doi.org/10.1038/nri.2016.42

Rosen, M. ., Callahan, B. ., Fisher, D. ., & Holmes, S. P. (2012). Denoising PCR-amplified metagenome data. *BMC Bioinformatic*, *13*, 283. https://doi.org/10.1186/1471-2105-13-283

Rossi, M., Martínez-Martínez, D., Amaretti, A., Ulrici, A., Raimondi, S., & Moya, A. (2016). Mining metagenomic whole genome sequences revealed subdominant but constant Lactobacillus population in the human gut microbiota. *Environ Microbiol Rep*, *8*(3), 399–406. https://doi.org/10.1111/1758-2229.12405

Ruan, W., Engevik, M. A., Spinler, J. K., & Versalovic, J. (2020). Healthy Human Gastrointestinal Microbiome : Composition and Function After a Decade of Exploration. *Digestive Diseases and Sciences*, *65*(3), 695–705. https://doi.org/10.1007/s10620-020-06118-4

Sadiq, F. A., Wenwei, L., Heyndrickx, M., Flint, S., Wei, C., Jianxin, Z., & Zhang, H. (2021). Synergistic interactions prevail in multispecies biofilms formed by the human gut microbiota on mucin. *FEMS Microbiol Ecol*, *97*(8). https://doi.org/10.1093/femsec/fiab096

Sankar, R. ., Biswas, R., Raja, S., Sistla, S., Gopalakrishnan, M., & SK, S. (2018). Brain abscess and cervical lymphadenitis due to Paraclostridium bifermentans: A report of two cases. *Anaerobe*, *51*, 8–11. https://doi.org/10.1016/j.anaerobe.2018.03.006.

Sansonetti, P. J. (2004). WAR AND PEACE AT MUCOSAL SURFACES. *Nat Rev Immunol*, *4*(12), 953–964. https://doi.org/10.1038/nri1499

Scher, J. U., Sczesnak, A., Longman, R. S., Segata, N., Ubeda, C., Bielski, C., Rostron, T., Cerundolo,

V., Pamer, E. G., Abramson, S. B., Huttenhower, C., & Littman, D. R. (2013). Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *ELIFE*, *2*, e01202. https://doi.org/10.7554/eLife.01202

Scott, K. P., Gratz, S. W., Sheridan, P. O., Flint, H. J., & Duncan, S. H. (2013). The influence of diet on the gut microbiota. *Pharmacological Research*, *69*(1), 52–60. https://doi.org/10.1016/j.phrs.2012.10.020

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., Jian, R., & Doré, J. (2003). Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut*, *52*(2), 237–242. https://doi.org/10.1136/gut.52.2.237

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), 1–14. https://doi.org/10.1371/journal.pbio.1002533

Senn, V., Bassler, D., Choudhury, R., Scholkmann, F., Righini-Grunder, F., Vuille-dit-Bile, R. N., & Restin, T. (2020). Microbial Colonization From the Fetus to Early Childhood—A Comprehensive Review. *Frontiers in Cellular and Infection Microbiology*, *10*(October), 1–12. https://doi.org/10.3389/fcimb.2020.573735

Seo, D. O., & Holtzman, D. M. (2020). Gut microbiota: From the forgotten organ to a potential key player in the pathology of Alzheimer's disease. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *75*(7), 1232–1241. https://doi.org/10.1093/gerona/glz262

Shah, S., Hankenson, J., Pabbathi, S., Greene, J., & Nanjappa, S. (2016). Clostridium tertium in neutropenic patients: case series at a cancer institute. *International Journal of Infectious Diseases*, *51*, 44–46. https://doi.org/10.1016/j.ijid.2016.08.013

Shi, C., Cheng, M., Yang, X., Lu, Y., Yin, H., Zeng, Y., Wang, R., Jiang, Y., Yang, W., Wang, J., Zhao, D., Huang, H., Ye, L., Cao, X., Yang, G., & Wang, C. (2020). Probiotic Lactobacillus rhamnosus GG Promotes Mouse Gut Microbiota Diversity and T Cell Differentiation. *Front Microbiol*, *11*, 607735. https://doi.org/10.3389/fmicb.2020.607735

Shi, H., Wang, Q., Zheng, M., Hao, S., Lum, J. S., Chen, X., Huang, X. F., Yu, Y., & Zheng, K. (2020). Supplement of microbiota-accessible carbohydrates prevents neuroinflammation and cognitive decline by improving the gut microbiota-brain axis in diet-induced obese mice. *Journal of Neuroinflammation*, *17*(1), 1–21. https://doi.org/10.1186/s12974-020-01760-1

Shi, N., Li, N., Duan, X., & Niu, H. (2017). Interaction between the gut microbiome and mucosal immune system. *Military Medical Research*, *4*(1), 1–7. https://doi.org/10.1186/s40779-017-0122-9

Sicard, J. F., Vogeleer, P., Bihan, G. Le, Olivera, Y. R., Beaudry, F., Jacques, M., & Harel, J. (2018). N - Acetyl - glucosamine influences the biofilm formation of Escherichia coli. *Gut Pathogens*, 1–10. https://doi.org/10.1186/s13099-018-0252-y

Sievers, F., Wilm, A., Dineen, D., Gibson, T., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J., & Higgins, D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, *7*, 539. https://doi.org/10.1038/msb.2011.75

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., & Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research*, *34*(Database issue), 32–36. https://doi.org/10.1093/nar/gkj014

Silverstein, R. B., & Mysorekar, I. U. (2021). Group therapy on in utero colonization : seeking common truths and a way forward. *Microbiome*, *9*(1), 1–7. https://doi.org/10.1186/s40168-020-00968-w

Smith, E., & Macfarlane, G. (1998). Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. *FEMS Microbiol. Ecol*, 355–368. https://doi.org/10.1111/j.1574-6941.1998.tb00487.x

Smith, M. (2020). Information theoretic generalized Robinson-Foulds metrics for comparing phylogenetic trees. *Bioinformatics*, *36*, 5007–5013. https://doi.org/10.1093/bioinformatics/btaa614

Smith, P. M., Howitt, M. R., Panikov, N., Michaud, M., Ann, C., Bohlooly-y, M., Glickman, J. N., & Garrett, W. S. (2013). The microbial metabolites, short chain fatty acids, regulate colonic Treg cell homeostasis. *Science*, *341*(6145), 569–573. https://doi.org/10.1126/science.1241165

Sorbara, M. T., & Pamer, E. G. (2022). Microbiome-based therapeutics. *Nat Rev Microbiol*. https://doi.org/10.1038/s41579-021-00667-9

Stackebrandt, E., & Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today*, *33*, 152–155.

Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology*, *44*(4), 846–849. https://doi.org/10.1099/00207713-44-4-846

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sundin, O. H., Mendoza-ladd, A., Zeng, M., Diaz-arévalo, D., Morales, E., Fagan, B. M., Ordoñez, J., Velez, P., Antony, N., & Mccallum, R. W. (2017). The human jejunum has an endogenous

microbiota that differs from those in the oral cavity and colon. *BMC Microbiol*, *17*(1), 160. https://doi.org/10.1186/s12866-017-1059-6

Tailford, L. E., Crost, E. H., Kavanaugh, D., & Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. *Frontiers in Genetics*, *5*(FEB). https://doi.org/10.3389/fgene.2015.00081

Tappe, D., Dirks, J., Müller, R., Brederlau, J., Abele-Horn, M., Suerbaum, S., & Kurzai, O. (2005). Fatal Clostridium tertium septicemia in a nonneutropenic patient. *J Infect*, *50*(1), 76–80. https://doi.org/10.1016/j.jinf.2004.04.005

Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, *11*(5), 472–477. https://doi.org/10.1016/j.mib.2008.09.006

Theodoropoulos, G., & Carraway, K. L. (2007). Molecular signaling in the regulation of mucins. *Journal of Cellular Biochemistry*, *102*(5), 1103–1116. https://doi.org/10.1002/jcb.21539

Thevaranjan, N., Puchta, A., Schulz, C., Naidoo, A., Szamosi, J., Verschoor, C., Loukov, D., Schenck, L., Jury, J., Foley, K., Schertzer, J., Larché, M., Davidson, D., Verdú, E., Surette, M., & Bowdish, D. (2017). Age-Associated Microbial Dysbiosis Promotes Intestinal Permeability, Systemic Inflammation, and Macrophage Dysfunction. *Cell Host Microbe*, *21*(4), 455–466. https://doi.org/10.1016/j.chom.2017.03.002

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, *474*(11), 1823–1836. https://doi.org/10.1042/BCJ20160510

Toward, R. E., Montandon, S. L., Walton, G. E., & Gibson, G. R. (2012). Effect of prebiotics on the human gut microbiota of elderly persons. *Gut Microbes*, *3*(1), 57–60. https://doi.org/10.4161/gmic.19411

Truong, D., Franzosa, E., Tickle, T., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., & Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods*, *12*(10), 902–903. https://doi.org/10.1038/nmeth.3589

Tu, Q. V, Mcguckin, M. A., & Mendz, G. L. (2008). Campylobacter jejuni response to human mucin MUC2 : modulation of colonization and pathogenicity determinants Printed in Great Britain. *J Med Microbiol*, *2008*, 795–802. https://doi.org/10.1099/jmm.0.47752-0

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-liggett, C., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, *449*(7164), 804–810. https://doi.org/10.1038/nature06244.The

Turner, J. R. (2009). Intestinal mucosal barrier function in health and disease. *Nature Reviews Immunology*, *9*(11), 799–809. https://doi.org/10.1038/nri2653

Vaiserman, A., Romanenko, M., Piven, L., Moseiko, V., Lushchak, O., Kryzhanovska, N., Guryanov,

V., & Koliada, A. (2020). Differences in the gut Firmicutes to Bacteroidetes ratio across age groups in healthy Ukrainian population. *BMC Microbiology*, *10*(1), 221. https://doi.org/10.1186/s12866-020-01903-7

van der Hee, B., & Wells, J. (2021). Microbial Regulation of Host Physiology by Short-chain Fatty Acids. *Trends Microbiol*, *29*(8), 700–712. https://doi.org/10.1016/j.tim.2021.02.001

Van Heel, A. J., De Jong, A., Song, C., Viel, J. H., Kok, J., & Kuipers, O. P. (2018). BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research*, *46*(W1), W278–W281. https://doi.org/10.1093/nar/gky383

van Muijlwijk, G. H., van Mierlo, G., Jansen, P. W. T. C., Vermeulen, M., Bleumink-Pluym, N. M. C., Palm, N. W., van Putten, J. P. M., & de Zoete, M. R. (2021). Identification of Allobaculum mucolyticum as a novel human intestinal mucin degrader. *Gut Microbes*, *13*(1), 1–18. https://doi.org/10.1080/19490976.2021.1966278

Vancamelbeke, M., & Vermeire, S. (2017). The intestinal barrier: a fundamental role in health and disease. *Expert Review of Gastroenterology and Hepatology*, *11*(9), 821–834. https://doi.org/10.1080/17474124.2017.1343143

Vanderhofstadt, M., André, M., Lonchay, C., Levecque, P., Holemans, X., Canon, J. L., & D'Hondt, L. (2010). Clostridium tertium bacteremia: Contamination or true pathogen? A report of two cases and a review of the literature. *International Journal of Infectious Diseases*, *14*(SUPPL. 3), 335–337. https://doi.org/10.1016/j.ijid.2010.03.004

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C., & Pati, A. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Research*, *43*(14), 6761–6771. https://doi.org/10.1093/nar/gkv657

Vasapolli, R., Schütte, K., Schulz, C., Vital, M., Schomburg, D., Pieper, D. H., Vilchez-vargas, R., & Malfertheiner, P. (2019). BASIC AND TRANSLATIONAL — ALIMENTARY TRACT Analysis of Transcriptionally Active Bacteria Throughout the Gastrointestinal Tract of Healthy Individuals. *Gastroenterology*, *157*(4), 1081-1092.e3. https://doi.org/10.1053/j.gastro.2019.05.068

Ventura, M., O'Toole, P. W., de Vos, W. M., & van Sinderen, D. (2018). Selected aspects of the human gut microbiota. *Cellular and Molecular Life Sciences*, *75*(1), 81–82. https://doi.org/10.1007/s00018-017-2669-8

Wagner, C. E., Wheeler, K. M., & Ribbeck, K. (2018). Mucins and Their Role in Shaping the Functions of Mucus Barriers. *Annual Review of Cell and Developmental Biology*, *34*, 189–215. https://doi.org/10.1146/annurev-cellbio-100617-0628181

Wang, B. X., Wu, C. M., & Ribbeck, K. (2021). Home, sweet home: how mucus accommodates our

microbiota. *FEBS Journal*, *288*(6), 1789–1799. https://doi.org/10.1111/febs.15504

Wazir, M., Jain, A. G., Nadeem, M., Ur Rahman, A., & Everett, G. (2019). Clostridium tertium bacteremia in a non-neutropenic patient with liver cirrhosis. *Cureus*, *11*(4), 4–8. https://doi.org/10.7759/cureus.4432

Wei, B., Cha, S. Y., Zhang, J. F., Shang, K., Park, H. C., Kang, J., Lee, K. J., Kang, M., & Jang, H. K. (2020). Antimicrobial susceptibility and association with toxin determinants in Clostridium perfringens isolates from chickens. *Microorganisms*, *8*(11), 1–12. https://doi.org/10.3390/microorganisms8111825

Weinstock, G. M. (2012). Genomic approaches to studying the human microbiota. *Nature*, *71*(2), 233–236. https://doi.org/10.1038/nature11553.Genomic

Wexler, A. G., & Goodman, A. L. (2017). An insider's perspective: Bacteroides as a window into the microbiome. *Nat Microbiol*, *2*, 17026. https://doi.org/10.1038/nmicrobiol.2017.26

Wong, Y. M., Juan, J. C., Gan, H. M., & Austin, C. M. (2014). Draft genome sequence of Clostridium bifermentans strain WYM, a promising biohydrogen producer isolated from landfill leachate sludge. *Genome Announcements*, *2*(2), 1–2. https://doi.org/10.1128/genomeA.00077-14

Wopereis, H., Oozeer, R., Knipping, K., Belzer, C., & Knol, J. (2014). The first thousand days – intestinal microbiology of early life : establishing a symbiosis. *Pediatr Allergy Immunol*, *25*(13), 428–438. https://doi.org/10.1111/pai.12232

Wu, H., Rebello, O., Crost, E. H., Owen, C. D., Walpole, S., Bennati, C., Didier, G., Serena, N., & Anna, H. (2021). Fucosidases from the human gut symbiont Ruminococcus gnavus. *Cellular and Molecular Life Sciences*, *78*(2), 675–693. https://doi.org/10.1007/s00018-020-03514-x

Wu, S., Rhee, K., Albesiano, E., Rabizadeh, S., Yen, H., Huso, D. L., Brancati, F. L., Wick, E., Mcallister, F., Housseau, F., Pardoll, D. M., & Sears, C. L. (2009). A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med*, *15*(9), 1016–1022. https://doi.org/10.1038/nm.2015

Wu, W.-J. H., Kim, M., Chang, L.-C., Assie, A., Saldana-Morales, F. B., Zegarra-Ruiz, D. F., Norwood, K., Samuel, B. S., & Diehl, G. E. (2022). Interleukin-1β secretion induced by mucosa-associated gut commensal bacteria promotes intestinal barrier repair. *Gut Microbes*, *14*(1). https://doi.org/10.1080/19490976.2021.2014772

Yang, J., Yang, Y., Ishii, M., Nagata, M., Aw, W., Obana, N., Tomita, M., Nomura, N., & Fukuda, S. (2020). Does the gut microbiota modulate host physiology through polymicrobial biofilms? *Microbes and Environments*, *35*(3), 1–13. https://doi.org/10.1264/jsme2.ME20037

Yang, S. (2021). Role of Goblet Cells in Intestinal Barrier and Mucosal Immunity. *J Inflamm Res*, 3171–3183. https://doi.org/10.2147/JIR.S318327

Yarandi, S. S., Peterson, D. A., Treisman, G. J., Moran, T. H., Pasricha, P. J., & Words, K. (2016). Modulatory Effects of Gut Microbiota on the Central Nervous System: How Gut Could Play a Role in Neuropsychiatric Health and Diseases. *J Neurogastroenterol Motil.*, *22*(2), 201–212. https://doi.org/10.5056/jnm15146

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., Whitman, W. B., Euzéby, J., Amann, R., & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, *12*(9), 635–645. https://doi.org/10.1038/nrmicro3330

Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A., Härkönen, T., Ryhänen, S., Franzosa, E., Vlamakis, H., Huttenhower, C., Gevers, D., Lander, E., Knip, M., & DIABIMMUNE Study Group, X. R. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatments on strain-level diversity and stability Moran. *Science Translational Medicine*, *8*(343), 343ra81. https://doi.org/10.1126/scitranslmed.aad0917.Natural

Yu, Z., Chen, C., Kling, D. E., Liu, B., Mccoy, J. M., Merighi, M., Heidtman, M., & Newburg, D. S. (2013). The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. *Glycobiology*, *23*(2), 169–177. https://doi.org/10.1093/glycob/cws138

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., & Yin, Y. (2018). dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *46*(W1), W95–W101. https://doi.org/10.1093/nar/gky418

Zhang, Y., Wang, L., Ocansey, D. K. W., Wang, B., Wang, L., & Xu, Z. (2021). Mucin-Type O-Glycans: Barrier, Microbiota, and Immune Anchors in Inflammatory Bowel Disease. *Journal of Inflammation Research*, *Volume 14*, 5939–5953. https://doi.org/10.2147/jir.s327609

Zhou, Y., Chen, H., He, H., Du, Y., Hu, J., Li, Y., Li, Y., Zhou, Y., Wang, H., Chen, Y., & Nie, Y. (2016). Increased Enterococcus faecalis infection is associated with clinically active Crohn disease. *Medicine (United States)*, *95*(39). https://doi.org/10.1097/MD.0000000000005019

Publications

- Candeliere, F., Raimondi, S., Ranieri, R., **Musmeci, E.**, Zambon, A., Amaretti, A., Rossi, M. (**2022**). β-glucuronidase pattern predicted from gut metagenomes indicates potentially diversified pharmacomicrobiomics. ***Frontiers in Microbiology.*** ahead of print

- **Musmeci, E.**, Candeliere, F., Amaretti, A., Rossi, M., Raimondi, S. (**2021**). Draft Genome Sequence of the Mucin Degrader *Clostridium tertium* WC0709**.** ***Microbiology Resource Announcements***, 10(32):e0064221. doi: 10.1128/MRA.00642-21.

- Raimondi, S., **Musmeci, E**., Candeliere, F., Amaretti, A., Rossi, M., (**2021**). Identification of mucin degraders of the human gut microbiota. ***Scientific Reports***, 11(1), 11094. doi: 10.1038/s41598-021-90553-4.

- Amaretti, A., Righini, L., Candeliere, F., **Musmeci, E**., Bonvicini, F., Gentilomi, G. A., Rossi, M., Raimondi, S. (**2020**). Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of Non-Escherichia coli Enterobacterales from the Gut Microbiota of Healthy Subjects. ***International Journal of Molecular Sciences***, 21(5), 1847. doi.org/10.3390/ijms21051847

- Raimondi, S., Righini, L., Candeliere, F., **Musmeci, E.,** Bonvicini, F., Gentilomi, G., Starčič Erjavec M., Amaretti A., Rossi, M. (**2019**). Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of E. coli Isolated from the Feces of Healthy Subjects. ***Microorganisms***, *7*(8), 251. doi.org/10.3390/microorganisms7080251

Report of activities

1st Year

PARTICIPATION TO THE SCHOOL COURSES

Date November 12 and 13, 2018
Title/subject "Medicinal plants as sources of potential pharmaceuticals products" and "Cameroonian wild mushrooms from the Bamoun Region: Nutritional and Medicinal properties"; prof. Moundipa.

Date January 14,15,28 and 29, 2019
Title/subject "Scientific English"; prof. Adrian Wallwork.

Date May 29, 2019 – October 2,4 and 9, 2019
Title/subject "Biological Models"

Date September 18, 2019
Title/subject "Microbial Biotechnologies for Biorefineries"; dr. Alberto Amaretti.

Date September 18 and 19, 2019
Title/subject "Food bioactive compounds"; dr. Davide Tagliazucchi.

PARTICIPATION TO SYMPOSIA
Date September 25-27, 2019
Location Catania
Title/subject 5th International Conference on Microbial Diversity

2nd Year

PARTICIPATION TO THE SCHOOL COURSES
Date: June 10,12, and 16, 2020
Title/subject: "Applications of multivariate analysis in the agri-food context"; prof. Alessandro Ulrici

Date: September 12,23 and 30, 2020
Title/subject: "Introduction to MATLAB environment"; Dr. Rosalba Calvini

PARTICIPATION TO SYMPOSIA
Date February 5-7, 2020
Location Milan
Title MicrobiotaMI 2020

3rd Year

PARTICIPATION TO THE SCHOOL COURSES

Date: February 5 and 12, 2021
Title/subject: Microbial Ecology- Prof.ssa Rossi Maddalena

ß-glucuronidase pattern predicted from gut metagenomes indicates potentially diversified pharmacomicrobiomics

Francesco Candeliere, Stefano Raimondi, Raffaella Ranieri, Eliana Musmeci, Alfonso Zambon, Alberto Amaretti and Maddalena Rossi*

Original Research, Front. Microbiol. - Evolutionary and Genomic Microbiology

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of the Mucin Degrader *Clostridium tertium* WC0709

Eliana Musmeci,[a] Francesco Candeliere,[a] Alberto Amaretti,[a,b] Maddalena Rossi,[a,b] Stefano Raimondi[a]

[a]Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy
[b]BIOGEST-SITEIA, University of Modena and Reggio Emilia, Modena, Italy

ABSTRACT The draft genome sequence of *Clostridium tertium* WC0709, a gut bacterium able to use mucin in pure culture as the sole carbon and nitrogen source, is presented here. The genome sequence of *C. tertium* will provide valuable references for comparative genome analysis and for studying the relationship with the host.

Clostridium tertium was first described by Henry in 1918 as *Bacillus tertius* (1) and was reclassified as *Clostridium tertium* in 1923 (2). It is a Gram-positive, anaerobic but aerotolerant, nontoxigenic (3) bacterium that forms spores only under anaerobic conditions. It can be isolated from soil, animals, and the human gastrointestinal tract (4). *C. tertium* is an uncommon human pathogen responsible for clinically significant bacteremia (5–7).

The strain *C. tertium* WC0709 was identified as a mucinolytic bacterium of the human gut in a previous study approved by the local institutional review board (reference number 125-15, Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy) (8). Medium containing mucin as the sole carbon and nitrogen source (MM) was utilized in enrichment cultures inoculated with fresh feces of healthy adults. The isolated strains were taxonomically assigned by 16S rRNA gene sequencing. *C. tertium* WC0709 grew in pure culture utilizing mucin as the sole carbon and nitrogen source, without the need of any cross-feeding interaction with other intestinal bacteria that make possible the use of this complex glycoprotein.

The strain was cultivated in minimal medium (MM) at 37°C for 48 h under strictly anaerobic conditions. Biomass was collected by centrifugation, and the genomic DNA was extracted with a DNeasy blood and tissue kit (Qiagen GmbH, Düsseldorf, Germany). Before DNA purification, the pretreatment for Gram-positive bacteria was followed by incubation for 2 h at 37°C with 2-fold the volume of enzymatic lysis buffer and for 1 h at 56°C with 2-fold the volume of proteinase K and buffer AL. The DNA concentration was quantified with a Qubit 3.0 fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA). The library was generated with the NEBNext Ultra II FS DNA library prep kit (New England BioLabs, Ipswich, MA, USA) and sequenced with an Illumina NovaSeq 6000 device by Eurofins Genomics (Ebersberg, Germany).

The sequencing run produced 15,916,139 paired-end reads that were 150 bp long. The raw reads were checked for quality with FastQC v0.11.8 (9). To trim Illumina adapters and remove reads with a quality score lower than 20, Cutadapt v1.16 was used with the following parameters: overlap = 15, minimum length = 30, and quality cutoff = 20 (10). The trimmed reads were verified again with FastQC and then assembled with SPAdes v3.13.0 with the parameters -careful, -cov-cutoff auto, and -k auto (11). The quality of the assembly was evaluated using QUAST v5.0.2 (12). Trimming, quality checking, and assembly were performed on the Galaxy platform (usegalaxy.eu) (13). The annotation was carried out with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.2 with the methods "best-placed reference protein set" and "GeneMarkS-2+" (14). The taxonomy attribution was confirmed with SpeciesFinder (https://cge.cbs.dtu.dk/services/SpeciesFinder) (15). For all the tools, default parameters were applied unless otherwise specified.

Citation Musmeci E, Candeliere F, Amaretti A, Rossi M, Raimondi S. 2021. Draft genome sequence of the mucin degrader *Clostridium tertium* WC0709. Microbiol Resour Announc 10:e00642-21. https://doi.org/10.1128/MRA.00642-21.

Editor David Rasko, University of Maryland School of Medicine

Address correspondence to Stefano Raimondi, stefano.raimondi@unimore.it.

Received 23 June 2021
Accepted 19 July 2021
Published 12 August 2021

The assembly produced a draft genome sequence encompassing 42 contigs (≥200 nucleotides [nt], 102 in total). The N_{50} length is 1,159,315 bp, while the L_{50} count is 2. The estimated genome size is 3,895,757 bp with a 27.72% G+C content and 1,230× coverage. A total of 3,578 coding sequences were annotated, comprising 10 rRNA genes (three 5S, four 16S, and three 23S) and 69 tRNAs.

Data availability. The sequence data were deposited in GenBank under BioProject accession number PRJNA737738 with BioSample accession number SAMN19707909 and GenBank accession number JAHLZG000000000. The raw reads have been deposited in the Sequence Read Archive (SRA) under accession number SRR14829814.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.

REFERENCES

1. Henry CH. 1918. An investigation of the cultural reactions of certain anaerobes found in wounds. J Pathol 21:344–385. https://doi.org/10.1002/path.1700210304.
2. Bergey DH, Harrison FC, Breed RS, Hammer BW, Huntoon FM. 1923. Bergey's manual of determinative bacteriology, 1st ed. The Williams and Wilkins Co, Baltimore, MD.
3. Milano V, Biehle L, Patel S, Hammer J. 2019. *Clostridium tertium* bacteremia and hepatic abscess in a non-neutropenic patient. IDCases 15:e00510. https://doi.org/10.1016/j.idcr.2019.e00510.
4. Vanderhofstadt M, André M, Lonchay C, Levecque P, Holemans X, Canon JL, D'Hondt L. 2010. *Clostridium tertium* bacteremia: contamination or true pathogen? A report of two cases and a review of the literature. Int J Infect Dis 14 (Suppl 3):e335–e337. https://doi.org/10.1016/j.ijid.2010.03.004.
5. Speirs G, Warren RE, Rampling A. 1988. *Clostridium tertium* septicemia in patients with neutropenia. J Infect Dis 158:1336–1340. https://doi.org/10.1093/infdis/158.6.1336.
6. Steyaert S, Peleman R, Vaneechoutte M, De Baere T, Claeys G, Verschraegen G. 1999. Septicemia in neutropenic patients infected with *Clostridium tertium* resistant to cefepime and other expanded-spectrum cephalosporins. J Clin Microbiol 37:3778–3779. https://doi.org/10.1128/JCM.37.11.3778-3779.1999.
7. Onderdonk AB, Allen SD. 1995. *Clostridium*, p 574–586. *In* Murray PR, Baron EJ, Pfaller MA, Tenover FC, Yolken RH (ed), Manual of clinical microbiology, 6th ed. American Society for Microbiology, Washington, DC.
8. Raimondi S, Musmeci E, Candeliere F, Amaretti A, Rossi M. 2021. Identification of mucin degraders of the human gut microbiota. Sci Rep 11:11094. https://doi.org/10.1038/s41598-021-90553-4.
9. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
10. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pyshkin AV. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
12. Gurevich A, Saveliev V, Vyahh N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
13. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
14. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
15. Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Pontén T, Aarestrup FM, Ussery DW, Lund O. 2014. Benchmarking of methods for genomic taxonomy. J Clin Microbiol 52:1529–1539. https://doi.org/10.1128/JCM.02981-13.

scientific reports

Check for updates

OPEN

Identification of mucin degraders of the human gut microbiota

Stefano Raimondi[1,3], Eliana Musmeci[1,3], Francesco Candeliere[1], Alberto Amaretti[1,2] & Maddalena Rossi[1,2]✉

Mucins are large glycoproteins consisting of approximately 80% of hetero-oligosaccharides. Gut mucin degraders of healthy subjects were investigated, through a culture dependent and independent approach. The faeces of five healthy adults were subjected to three steps of anaerobic enrichment in a medium with sole mucins as carbon and nitrogen sources. The bacterial community was compared before and after the enrichment by 16S rRNA gene profiling. Bacteria capable of fermenting sugars, such as *Anaerotruncus*, *Holdemania*, and Enterococcaceae likely took advantage of the carbohydrate chains. *Escherichia coli* and Enterobacteriaceae, Peptococcales, the Coriobacteriale *Eggerthella*, and a variety of Clostridia such as Oscillospiraceae, *Anaerotruncus*, and *Lachnoclostridium*, significantly increased and likely participated to the degradation of the protein backbone of mucin. The affinity of *E. coli* and Enterobacteriaceae for mucin may facilitate the access to the gut mucosa, promoting gut barrier damage and triggering systemic inflammatory responses. Only three species of strict anaerobes able to grow on mucin were isolated from the enrichments of five different microbiota: *Clostridium disporicum*, *Clostridium tertium*, and *Paraclostridium benzoelyticum*. The limited number of species isolated confirms that in the gut the degradation of these glycoproteins results from cooperation and cross-feeding among several species exhibiting different metabolic capabilities.

Mucus is a complex gel barrier that covers the wet epithelial surfaces throughout the body, including the gastrointestinal tract, and offers protection against exogenous and endogenous aggressive agents[1,2]. It exerts a variety of functions such as lubrification, hydration, chemical protection, sensing, nutrient reservoir, and barrier against pathogen invasion. A continuous turnover, consisting in a dynamic balance between biosynthesis, secretion, and degradation of its structural components, is crucial for these functions[3].

Mucins are major component of mucus, made up by glycoproteins with high level of glycosylation. Residues of galactose, N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), fucose, and sialic acid, with relatively small amounts of mannose, constitute the oligosaccharides which represent approx. 80% of mucin mass[4]. The protein core is mainly organized in tandem repeated regions rich in serine, threonine, and proline. Serine and threonine are the site where an O-glycosidic bound links the protein core to GalNAc, the first moiety of the glycan chain[5,6].

Among all the sites of human body inhabited by a resident microbial community, the colon hosts the most complex and concentrated microbiota[7]. In this site, mucins behave as decoy molecules that avoid the interaction of bacterial adhesins with receptors of the colonic epithelium. In fact, the gut lumen is coated by a two-layered mucus system, with an inner dense stratus firmly attached to the epithelium and an outer one, looser and unattached[8]. The inner layer is thick and stratified and prevents gut bacteria from reaching the epithelial cell surface. It is progressively converted into a laxer and expanded coating through the lytic action of proteases and glycosidases of both the host and the commensal bacteria. The outer mucus layer is the colonization site of the resident commensal microbiota, which differs from that of the digesta-associated and fecal content in terms of relative abundance of the different taxa[9].

Mucus layers are the frontline of the interaction between the host and the gut bacteria, thus a balanced and symbiotic relationship which benefits both the former and the latter greatly relies on mucus structure[2]. Mucin plays a pivotal role in the selection of bacteria colonizing the mucus by supplying carbon and nitrogen substrates and exposing O-glycan chains that serve as attachment sites for colonization[4,8]. In turn, commensal bacteria feeding on and adhering to mucin limit the penetration of pathogens in the outer layer of the mucus[10]. These bacteria tightly interact with the host and modulate mucin gene expression, glycosylation, and secretion, thus affecting mucus and epithelial homeostasis[3,11].

[1]Department of Life Sciences, University of Modena and Reggio Emilia, 41125 Modena, Italy. [2]Biogest-Siteia, University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy. [3]These authors contributed equally: Stefano Raimondi and Eliana Musmeci. ✉email: maddalena.rossi@unimore.it

Sample	Anaerobic mucinolytic bacteria Log (cfu/mL)	Enterobacteriaceae Log (cfu/mL)	Isolated strains		Ref_seq ID
EC 1	7.8 ± 0.9	7.2 ± 0.5	*Clostridium disporicum*	WC0700	NR_026491.1
			Clostridium disporicum	WC0701	NR_026491.1
EC 2	8.4 ± 0.7	7.3 ± 0.4	*Clostridium disporicum*	WC0702	NR_026491.1
			Clostridium disporicum	WC0703	NR_026491.1
EC 3	8.5 ± 0.5	7.9 ± 0.5	*Clostridium disporicum*	WC0704	NR_026491.1
EC 4	8.3 ± 0.6	7.6 ± 0.6	*Paraclostridium benzoelyticum*	WC0705	NR_148815.1
			Clostridium disporicum	WC0706	NR_026491.1
EC 5	8.4 ± 0.5	8.0 ± 0.3	*Paraclostridium benzoelyticum*	WC0707	NR_148815.1
			Paraclostridium benzoelyticum	WC0708	NR_148815.1
			Clostridium tertium	WC0709	NR_113325.1

Table 1. Cultivable mucin degraders in EC samples.The viable counts of anaerobic mucinolytic bacteria and Enterobacteriaceae, respectively counted onto plates of MM medium and MacConkey, and the taxonomic attribution of the predominant anaerobic is reported. Counts are reported as mean Log (cfu/mL) ± standard deviation, n = 3. For both the media, no significant difference was observed among samples (ANOVA, $P > 0.05$). Cultivable anaerobic mucinolytic bacteria were significantly more numerous than Enterobacteriaceae (paired samples t-test, $P < 0.05$). Taxonomic attribution was obtained by alignment of the V1–V4 regions of 16S rRNA gene with the NCBI refseq_rna database using BLAST. The accession number of best hits and the sequence identity (%) obtained with MUSCLE are reported.

This study aimed to investigate mucin degraders of the gut, to unveil how mucins influence selection of resident bacteria and how specific bacteria shape the mucus layer. A medium containing mucin as sole carbon and nitrogen source (hereinafter referred to as MM medium) was utilized in strictly anaerobic enrichment cultures of human gut microbiota, inoculated with fresh feces of 5 healthy adults. The taxa that took advantage of mucins as growth substrate were identified with a 16S rRNA gene metataxonomic analysis, comparing the bacterial community of the fecal slurries (hereinafter referred to as FS) and the cultures after three enrichment steps (hereinafter referred to as EC). The bacteria able to grow using mucin as sole carbon and nitrogen source were isolated and taxonomically characterized from EC samples.

Results

Cultivable bacteria enriched on mucin. Intestinal mucin degrading bacteria were enumerated and isolated from EC samples, i.e. after three enrichment steps in MM broth. To discriminate anaerobic mucin degraders from mucinolytic Enterobacteriaceae, the enumeration was carried out with MM plates in anaerobiosis and MacConkey plates in aerobiosis (Table 1), and the isolates from MM plates were confirmed to be strict anaerobes. Strictly anaerobic mucin degraders were found in the magnitude of 10^8 cfu/mL (mean ± SD = 8.3 ± 0.3 Log_{10} cfu/mL), significantly more abundant ($P < 0.05$) than mucinolytic Enterobacteriaceae, that ranged between 10^7 and 10^8 cfu/mL (mean ± SD = 7.6 ± 0.4 Log_{10} cfu/mL). The viable counts of both anaerobic mucin degraders and mucinolytic Enterobacteriaceae did not differ among the EC samples (ANOVA, $P > 0.05$).

Based on RAPD-PCR fingerprinting, the anaerobic isolates clustered in 1–3 distinct biotypes *per* sample, for a total of 10 strains (Table 1). The partial sequences of 16S rRNA genes showed identity > 98.6% to the following species: *Clostridium disporicum* (6 biotypes), *Clostridium tertium* (1), and *Paraclostridium benzoelyticum* (3). These strains were not able to grow in MM plates without mucin.

Microbiota of enrichment cultures grown on mucin. The metataxonomic survey of 16S rRNA gene in FS and EC samples yielded a total 415,051 sequences (31,153–55,162 *per* sample). The reads were dereplicated into 522 ASVs hitting a reference sequence in Silva database, and collapsed at the 7th level of taxonomic annotation into 150 OTUs (Supplementary Table 1).

The analysis of microbiota alpha-diversity revealed that the richness (Chao1) and the evenness (Shannon and Pielou) tended to decrease with the enrichment but did not reach statistical significance ($P > 0.05$) (Supplementary Fig. 1). According to Weighted UniFrac computation of beta-diversity, FS and EC samples clustered in distinct groups (ANOSIM, PERMANOVA, $P < 0.05$) in a PCoA plot, with the former laying at lower values of PCo1 than the latter (Fig. 1A). *Escherichia-Shigella* presented the most positive contribution to PCo1, followed by ASVs of *Bacteroides* and of Lachnospiraceae, while other ASVs of *Bacteroides*, Prevotellaceae NK3B31, and *Roseburia* negatively weighted (Fig. 1B). Unweighted UniFrac yielded more dispersed groups (ANOSIM, PERMANOVA, $P > 0.05$), with EC samples laying at lower PCo2 compared to the corresponding FS ones (Supplementary Fig. 2A). The ASVs that negatively contributed to PCo2 included *Clostridium* sensu stricto *I* and other *Bacteroides*, while Prevotellaceae NK3B31 and *Eubacterium coprostanoligenes* were positive contributors (Supplementary Fig. 2B). Weighted UniFrac computation takes into account abundances of taxa, unlike Unweighted Unifrac that considers the mere presence or absence. Thus, the taxa that emerged from the former presented relevant changes in terms of relative abundance, while those that emerged from the latter likely appeared or disappeared over the enrichment steps.

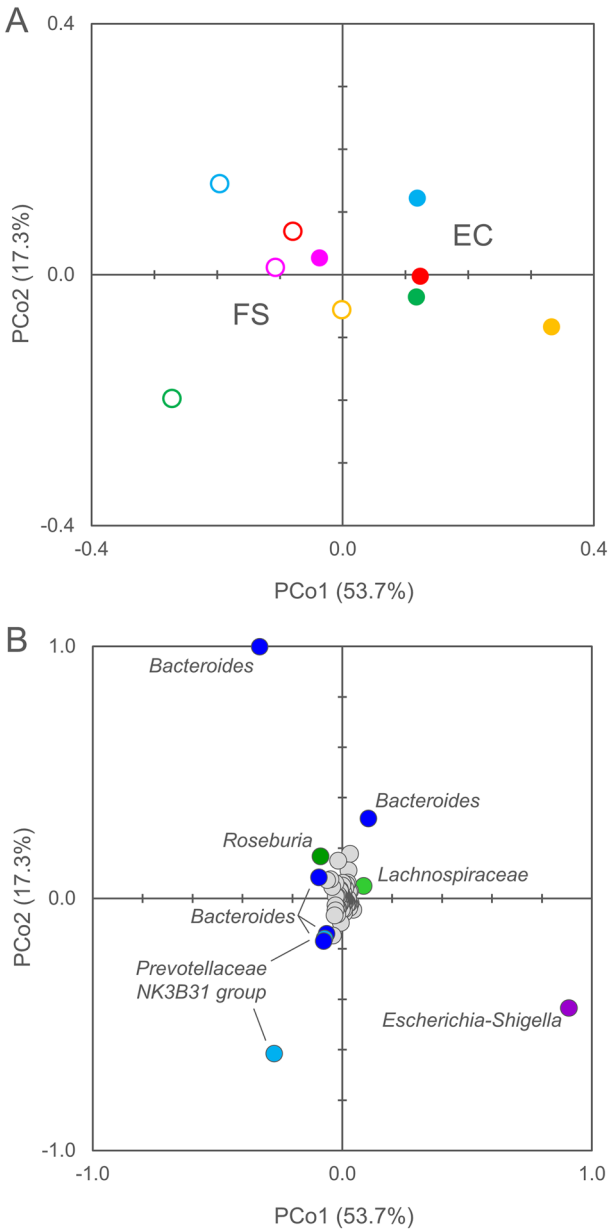


Figure 1. Beta diversity analysis of the microbiota in FS and EC samples. (**A**) PCo1-PCo2 visualization of distances computed with Weighted Unifrac, that considers phylogenetic relationships among bacteria. Symbols: FS, empty circle; EC, full circle; different colors correspond to different subjects (1, fuchsia; 2, cyan; 3, green; 4, red; 5, yellow). (**B**) PCo1-PCo2 visualization of the contribution of each ASV. The 10 ASVs exerting the highest effect on PCo1 are labelled and colored according to their phylum (Firmicutes, green; Bacteroidetes, blue; Proteobacteria, purple).

The microbiota of FS samples encompassed 125 of the 150 identified OTUs. It was dominated by Bacteroidota (59.5%), with remarkably high amounts of *Bacteroides* (39.3%) and Prevotellaceae (8.4%), followed by Firmicutes (35.3%), especially Lachnospiraceae (15.4%), and Proteobacteria (2.2%; Fig. 2). The microbiota of EC samples encompassed 101 OTUs, 25 of which became detectable because of the enrichment steps, as they lay below the limit of detection in FS samples. EC were dominated by Bacteroidota (39.3%) and Proteobacteria (30.8%). *Escherichia-Shigella* (with Enterobacteriaceae, Enterobacteriales, Gammaproteobacteria, and Proteobacteria) presented a significant positive differential abundance in EC compared to FS (Fig. 3) and represented one of the dominant genera in EC samples (mean = 30.5%), with remarkably high abundance in sample EC-5 (56.7%). Several taxa of Firmicutes, such as *Enterococcus* and many Clostridia (e.g., Oscillospiraceae, Peptococcaceae, Ruminococcaceae UBA1819, *Soleaferrea*, *Lachnoclostridium*, and *Anaerotruncus*), presented a significant increase in EC samples, but generally remained below a mean abundance of 1%.

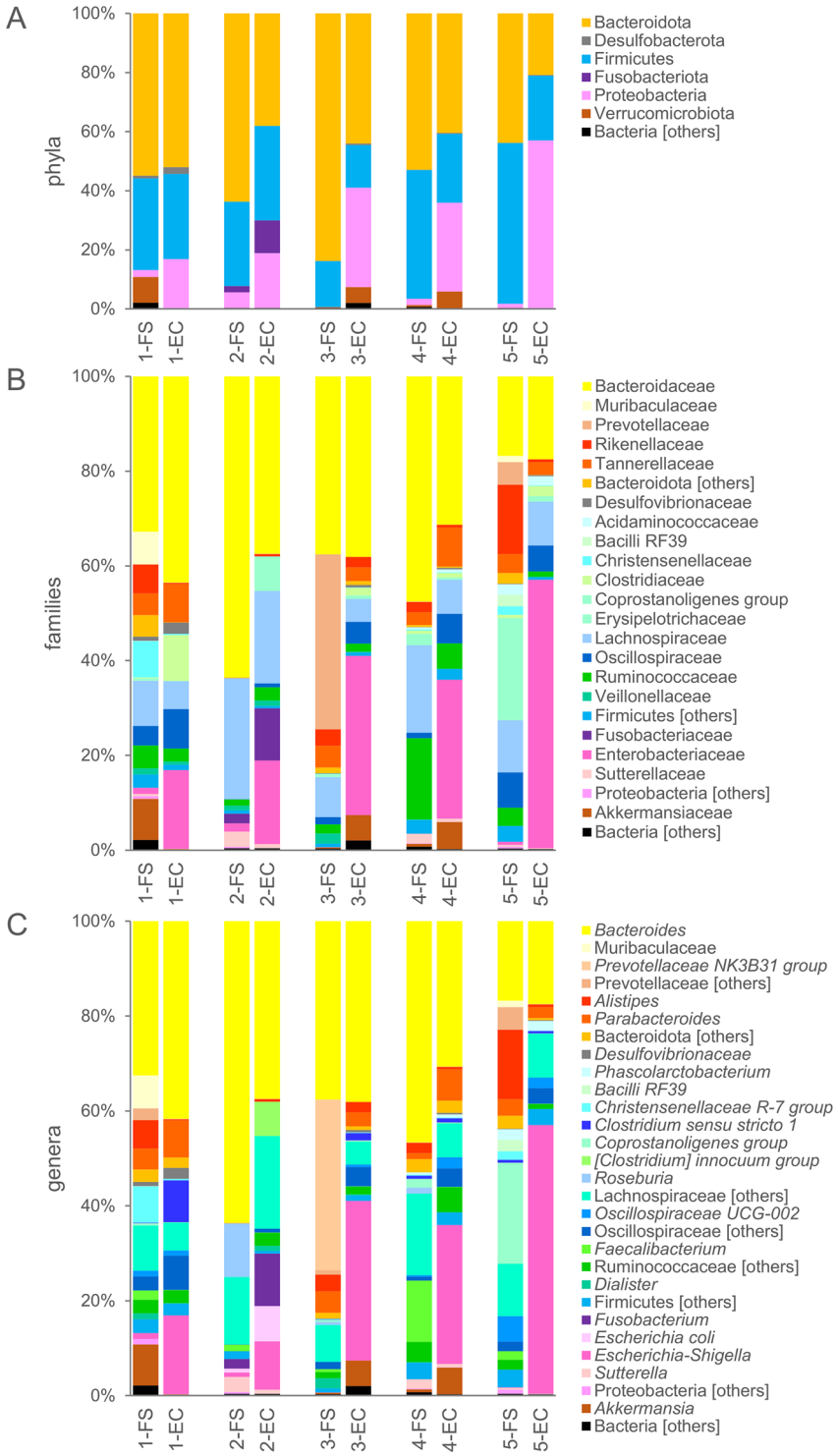


Figure 2. Stacked bar-plot representation of microbiota composition before and after the enrichment in MM, with taxonomic features collapsed at the level of phyla (**A**), families (**B**), and genera (**C**). The taxa that remained unclassified at the deeper level or that never occurred with abundance higher than 0.5% are grouped and marked with "[others]".

A number of taxa were not detected in any FS sample but appeared after the enrichment: *Eggerthella* occurred in 4 EC samples; *Clostridium innocuum*, *Enterococcus durans*, *Erysipelatoclostridium*, and *Soleaferrea* in 3; *Angelakisiella* and *Eubacterium nodatum* in 2 EC samples. These taxa were generally a minor portion of the microbiota (< 1%), except for *Clostridium innocuum* that reached the 7.3% in EC-2. *Enterorhabdus*, *Gordonibacter*, *Dielma*, *Enterococcus faecalis* and other *Enterococcus* spp., *Catabacter, Ruminiclostridium*, Lachnospiraceae FCS020 group,

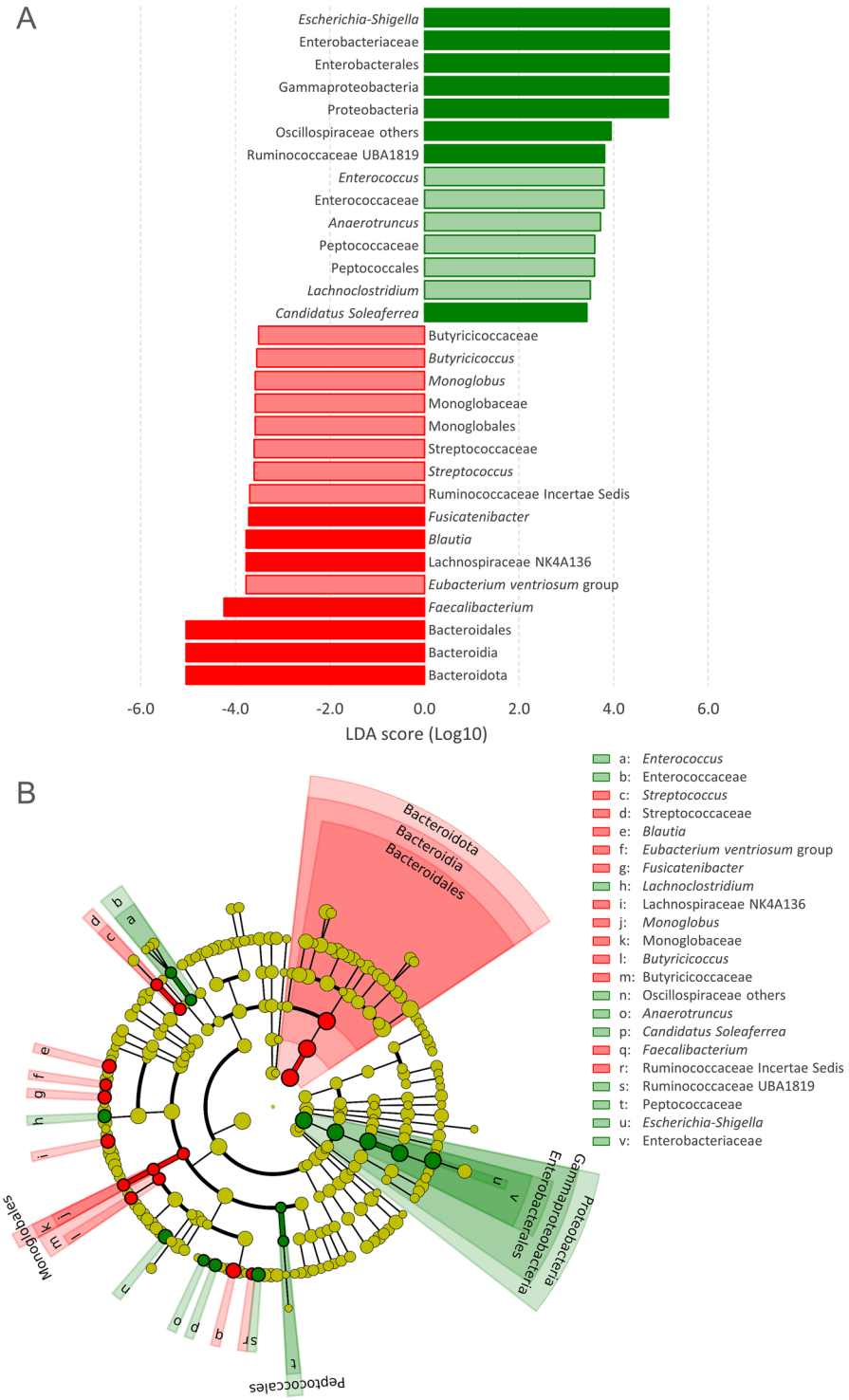


Figure 3. Effect of mucin enrichment on faecal microbiota. The taxonomic biomarkers characterizing FS and EC according to LEfSe are colored in red and green, respectively. (**A**) Logarithmic LDA scores of the taxa exhibiting significant differential abundance ($P < 0.05$, logarithmic LDA score ≥ 2.0) between FS and EC; taxa that were > 1% in at least one sample are reported with bars in dark shades. (**B**) Cladogram visualization of the taxonomic biomarkers. The figure was created with the software LEfSe, Galaxy version 1.0, available at https://huttenhower.sph.harvard.edu/galaxy/.

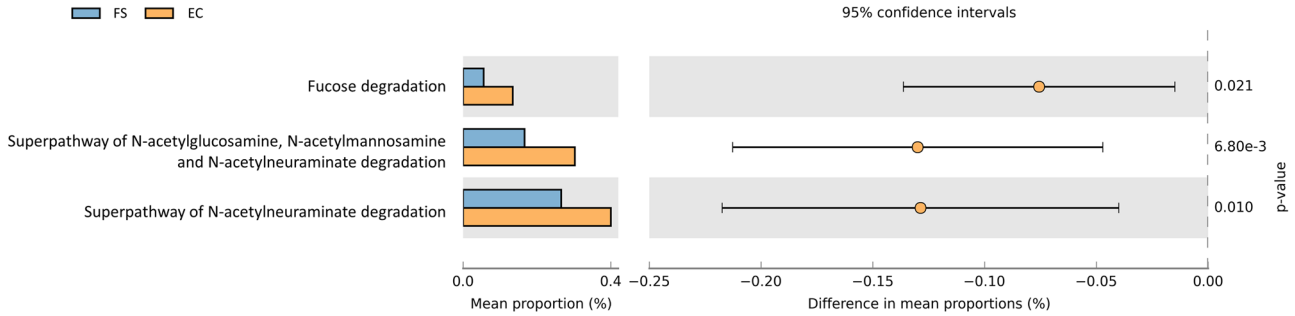


Figure 4. Pathways related to mucin degradation reconstructed from 16S rRNA profile with PICRUSt2 and presenting a significant ($P < 0.05$) differential abundance between FS and EC. The figure was created with STAMP v2.1.3, available at https://beikolab.cs.dal.ca/software/STAMP.

Hydrogenoanaerobacterium, *Caproiciproducens*, Ruminococcaceae DTU089, *Peptostreptococcus, Peptoniphilus, Tissierella, Sedimentibacter, Citrobacter*, and *Cloacibacillus* were detected in a single EC sample, generally in very low concentration, except for *Dielma* (1%) and *Cloacibacillus* (1.5%).

Bacteroidota, Bacteroidia, and Bacteroidales, significantly decreased as consequence of mucin enrichment, albeit the genus *Bacteroides* remained the most abundant in 4 out of 5 EC samples. The genus *Streptococcus* also diminished, leading to significantly lower abundances also of Streptococcaceae. Numerous taxa of Clostridia decreased during the enrichment steps, such as *Eubacterium ventriosum*, the genera *Faecalibacterium, Fusicatenibacter, Blautia, Monoglobus* (plus Monoglobaceae and Monoglobales), and *Butyricicoccus* (and Butyricicoccaceae). Furthermore, the abundance of a clade of Lachnospiraceae (NK4A136) and Ruminococcaceae *incertae sedis* significantly diminished. Other taxa detected only in some FS samples decreased below the limit of detection in the EC samples, such as Coriobacteriales *incertae sedis,* other Eggerthellaceae, *Paraprevotella*, Defluviitaleaceae_UCG-011, *Eubacterium siraeum* group, *Veillonella*, *Butyricimonas*, Erysipelotrichaceae_UCG-003, *Turicibacter*, *Coprococcus*, *Lachnospira*, *Roseburia*, Anaerovoracaceae Family XIII UCG-001, and *Haemophilus*.

The functions of the fecal cultures' metagenomes were predicted with PICRUSt2. A significant increase (t-test, $p < 0.05$) of pathways related to mucin metabolism was detected, i.e. fucose degradation, superpathway of N-acetylneuraminate degradation, and superpathway of GlcNAc, N-acetylmannosamine, and N-acetylneuraminate degradation (Fig. 4). In particular, genes encoding N-acetylglucosamine kinase (EC:2.7.1.59), N-acetylglucosamine-6-phosphate deacetylase (EC:3.5.1.25), N-acetylneuraminate epimerase (EC:5.1.3.24), N-acylglucosamine-6-phosphate 2-epimerase (EC:5.1.3.9), N-acylmannosamine kinase (EC:2.7.1.60) showed a statistically significant increase in the enriched cultures.

Discussion

This study aimed to identify and isolate the intestinal bacteria that can have a role in mucin degradation, utilizing three steps of enrichment where the intestinal microbiota of healthy subjects was inoculated in a medium containing mucin as the sole carbon and nitrogen source. Mucin can support the growth of gut microbes with different metabolic capabilities. Both saccharolytic and proteolytic activities are involved in its breakdown, providing the bacterial community with carbon and nitrogen supply and sustaining growth of both sugar and amino acid fermenters. Based on the increase of genes and pathways involved in mucin degradation, the cultures successfully selected bacteria performing mucin breakdown and fermentation.

The 16S rRNA gene profiling highlighted the taxa that proliferated in mucin, likely because of cooperation and crossfeeding. Proteobacteria and taxa related to Enterobacteriaceae encompassed the main biomarkers of growth, in agreement with their major role as protein degraders in gut microbiota. As opportunistic commensals, they may persist in the gut contributing to the mucus homeostasis without interacting with epithelial cells or they may degrade mucins, reach the epithelium, and initiate disease[12–14]. The high affinity of Enterobacteriaceae for mucins confirms the possibility that they may get in contact with the intestinal epithelial barrier and exert this pathogenic action. In fact, these bacteria exhibit an intrinsic resistance to the oxygen released from the host tissue and may concur to disrupt the integrity of the epithelium, promoting a leaky gut and triggering systemic inflammatory response[15–17].

Enterobacteriaceae, Peptococcales/Peptococcaceae, a variety of Clostridia (such as Oscillospiraceae, *Lachnoclostridium*, and *Anaerotruncus*), and the Coriobacteriale *Eggerthella,* are protein degraders that may have taken part to the breakdown of protein backbone of mucin[18]. The genus *Anaerotruncus* includes species (*Anaerotruncus colihominis* and *Anaerotruncus massiliensis*) that ferment some amino acids and also a small set of carbohydrates, including GlcNAc, and may have participated to the degradation of both the protein and saccharidic portions of mucins[19,20]. *Holdemania*, a Firmicute belonging to the family Erysipelotrichaceae, and Enterococcaceae (Lactobacillales) are saccharolytic bacteria, thus their contribution to mucin degradation seems limited to the carbohydrate chains. Several other taxa, which are still poorly characterized at the physiological level, emerged as participating in mucin catabolism, for some species representing one of the first sources of functional information. Interestingly, the list of bacteria that got enriched in EC includes taxa that are associated to colorectal adenoma and cancer (e.g. *Lachnoclostridium*) and are emerging as uncommon human pathogens (*Eggerthella* and some Enterococcaceae)[21–23].

Even though the 16S rRNA gene profiling disclosed a complex population of bacteria taking advantage of mucin, only three species of strict anaerobic bacteria (*C. disporicum*, *C. tertium*, and *P. benzoelyticum*) were isolated from the cultures enriched with 5 different microbiotas. *C. disporicum* is a saccharolytic bacterium that presents epimerase activity resulting in the production of ursodeoxycholic acid[24,25]. *C. tertium* is an uncommon pathogen, albeit associated with cases of severe diseases, mostly in immunocompromised and neutropenic patients[26]. *P. benzoelyticum* can ferment some amino acids, including threonine, a major component of mucin backbone[27]. Interestingly, the isolates of *C. disporicum* and *C. tertium* are all designated as *Clostridium* sensu strictu 1 according to SILVA database. The 16S rRNA gene profiling revealed that the OTU *Clostridium* sensu strictu 1 increased in all the samples where it was present, although it did not reach a significant differential abundance, and was composed of four ASVs contributing to FS and EC separation along PCo2 in the plot of Unweighted UniFrac beta-diversity.

Albeit *Clostridium* species harbor few glycosyl hydrolases (GH) involved in mucin degradation and, for instance, genome of *C. sporogenes,* a closely related species to *C. tertium*, has no mucin-related GHs, based on preliminary genomic analysis (data not shown), the genome of *C. tertium* WC 0709 has at least one exo-α-sialidase (EC 3.2.1.18) belonging to the GH33 family, involved in the cleavage of the terminal sialic acid, and a GH3 family β-hexosaminidase (EC 3.2.1.52) that can release GalNac. On the other side *P. bifermentans*, closely related to *P. benzoelyticum*, harbors only the galactosidases GH2 and GH20. Generally, commensal clostridia do not appear to harbor GH33 or other mucin-related GHs, whereas two pathogenic clostridia degrade mucin glycans: *C. perfingens* and *C. septicum*[28]. The capability of *C. disporicum*, *C. tertium*, and *P. benzoelyticum* to dismantle complex mucin glycans deserves deeper investigations and opens new perspectives on the relationship of these species with the host, that can go beyond the commensalism. Incidentally, *C. tertium* is described as an uncommon pathogen associated with various cases of severe diseases[26,29]. The glycans of intestinal mucins, enriched in diverse monosaccharides with different positional and stereochemical linkages and branching, are hydrolyzed by a number of diverse sialidases, hexosaminidases, galactosidases, and fucosidases. The genetic and functional characterization of the enzymes that in *C. disporicum*, *C. tertium*, and *P. benzoelyticum* play key roles in disassembling mucin oligosaccharides is a great challenge.

In the gut, a complex community of saccharolytic and proteolytic bacteria is expected to participate to mucin breakdown, with linkage-specific glycosidases and proteases of primary degraders providing other bacteria with simpler oligosaccharides and peptides, and eventually with their monomeric moieties. It is plausible that several taxa detected by the 16S rRNA gene analysis could have grown on mucins only when co-cultured within a bacterial community where each species contributed with different hydrolytic activities, while they could not grow in pure cultures where they did not benefit from crossfeeding relationships. Only *C. disporicum*, *C. tertium*, and *P. benzoelyticum* gained carbon, energy, and nitrogen from mucin in pure cultures. The isolates of these species deserve deeper investigation, including the genome sequencing and physiology studies, aiming to unveil the key functions that enabled growth on mucin, in particular their set of glycosyl hydrolases and proteases and their amino acid biosynthetic capability.

To date, *Bacteroides* and *Akkermansia muciniphila* have been recognized as main mucin degraders, with novel insights on endo-acting O-glycanases involved in the first steps of glycan breakdown[30]. In this study, the relative abundance of *Bacteroides* did not increase significantly, even if some ASVs of *Bacteroides* heavily contributed to separate samples along both PCo1 or PCo2 in the Weighted UniFrac analysis, confirming the presence of some species greatly associated to mucin degradation. It is plausible that different populations of *Bacteroides*, which could not be discriminated into species by the partial 16S rRNA gene sequencing, behaved differently with respect to mucin breakdown, exerting contrary effects and balancing the counts of the genus. Indeed, a wide number of ASVs attributed to *Bacteroides* presented opposite contributions along the principal coordinates, in agreement with the fact that the genus *Bacteroides* comprises tens of species that supply different and complementary functions to the intestinal community[31].

One of the key members of the colonic mucus-associated microbiota is *A. muciniphila*. It releases monosaccharides and amino acids during mucin degradation, providing nutrients to other bacteria of the gut, and exerts beneficial effects in various metabolic disorders, being generally negatively correlated with inflammation and metabolic disorders[32]. In this study, a sole ASV ascribed to *Akkermansia* was detected in 3 out of 5 set of samples. In subject 1, it was remarkably abundant in the founding microbiota (8.7%) but was negligible (0.2%) after the enrichment steps. In samples 3 and 4, *Akkermansia* was initially < 0.6% and increased by one magnitude over the enrichment. It is plausible that in sample 1 the high initial load of *Akkermansia* prevented a net increase of the abundance, that was further squeezed by major increase in some other groups.

The absence of *Bacteroides* and *Akkermansia* also within the isolates is noteworthy. These bacteria likely encountered some substrate limitation that hampered growth. In fact, the medium did not contain nitrogen sources other than mucin in order to be as much stringent as possible. Thus, all the amino acids had to be obtained from the protein backbone of mucin, and this might have been too challenging, in the absence of a cooperative crossfeeding. The composition of the medium and the choice of the pH are major drivers of microbial selection and a mucin-based medium with different composition would likely result in the isolation of different bacterial species. Another drawback of the present study was the use of mucin from porcine stomach, that mainly consists of MUC5AC and MUC6 mucins and differs in purity and glycosylation profile from the colonic one MUC2, but is a cheap and easily available substrate, commonly utilized to culture mucin degraders[5,33–35].

This study shed light on the bacterial taxa thriving in cultures of human microbiota with mucin as primary substrate for growth. It provided the first information on *C. disporicum*, *C. tertium*, and *P. benzoelyticum* as bacteria able to grow in pure cultures utilizing mucin as sole carbon and nitrogen source, complementing the recognized role of *Bacteroides* and *Akkermansia* in mucin degradation and turnover. Sequencing of the 10 strains is in progress, as well as the formulation of a new mucin-based medium allowing a faster growth with higher yields, in order to better address physiology and metabolism studies. Proteobacteria, including Enterobacteriaceae

and *E. coli*, are among the microbial group that most benefit from mucin as substrate, increasing the risk of the flourishing of opportunistic pathogens or bacteria associated to gut inflammation. Furthermore, metataxonomic analysis identified *Lachnoclostridium*, *Eggerthella*, *Anaerotruncus*, *Enterococcaceae*, and *Peptococcaceae* as taxa that took advantage, likely in a cross-feeding relationship, in the mucin-based medium. The limited number of isolated species able to grow on mucin in pure culture confirms that the degradation of these glycoproteins in the gut results from cooperation and cross-feeding among several species exhibiting different metabolic capabilities.

Materials and methods

Chemicals and media. All the chemicals were purchased from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany), unless otherwise stated. Mucin was purified from porcine stomach type II mucin (Sigma-Aldrich), according to the protocol by Miller and Hoskins[36]. Briefly, mucin was precipitated with ice-cold ethanol at the concentration of 60% (v/v) from a clarified solution, obtained by centrifugation (10,000×*g* for 10 min at 4 °C) of a 25 g/L mucin suspension in 0.1 M NaCl containing 0.02 M phosphate buffer, pH 7.6. The precipitate was dissolved in 0.1 M NaCl, re-precipitated with ethanol (60% v/v), and dissolved in water. The mucin solution was dialyzed with (14 kDa cut-off) against water and finally freeze dried. According to the producer, mucin had a molecular mass of 4000–5500 kDa, with a carbohydrate portion of 80–84%, and a protein portion of 16–20%. Proline, serine, and threonine residues in the peptide backbone comprise approx. 43% of the amino acid composition.

MM medium contained 3.0 g/L purified mucin, 2.0 g/L KH_2PO_4, 4.5 g/L NaCl, 0.5 g/L $MgSO_4{\cdot}7H_2O$, 0.045 g/L $CaCl_2{\cdot}2H_2O$, 0.005 g/L $FeSO_4{\cdot}7H_2O$, 0.01 g/L hemin, 0.05 g/L bile salts (Oxgall, BD Difco, Sparks, MD, USA), 0.6 mg/L resazurin, 2.0 mL/L minerals solution (0.5 g/L EDTA, 0.010 g/L $ZnSO_4{\cdot}7H_2O$, 0.003 g/L $MnCl_2{\cdot}7H_2O$, 0.03 g/L H_3BO_3, 0.02 g/L $CoCl_2{\cdot}6H_20$, 0.001 g/L $CuCl_2{\cdot}2H_2O$, 0.002 g/L $NiCl_2{\cdot}6H_2O$, and 0.003 g/L $NaMoO_4{\cdot}2H_2O$), 1.4 mL/L vitamins solution (1.0 g/L menadione, 2.0 g/L biotin, 2.0 g/L calcium pantothenate, 10 g/L nicotinamide, 0.5 g/L cyanocobalamin, 0.5 g/L folic acid, 4 g/L thiamine, and 5 g/L PABA), and 40 mL/L reducing solution (12.5 g/L L-cysteine·HCl and 80 g/L $NaHCO_3$). The basal medium was autoclaved at 121 °C for 20 min and supplemented with filter sterilized (0.22 µm) minerals, vitamins, and reducing solutions. Soft-agar MM plates, containing 8 g/L agar (BD Difco), were used to count, and purify mucin degraders. MacConkey agar plates (BD Difco) were utilized to count *Escherichia coli* and Enterobacteriaceae and to discriminate them from strictly anaerobic isolates.

Enrichment cultures. This study was carried out under the recommendations of the protocol approved with ref. no. 125-15 by the local research ethics committee (Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy), with written informed consent from all subjects following the Declaration of Helsinki. Five healthy adults (3 men and 2 women, aged 25–50 years) who had not taken prebiotics and/or probiotics in the previous 2 weeks or antibiotics for at least 3 months were enrolled for feces collection. Fecal samples were collected fresh, sealed in anaerobic plastic bags (AnaeroGen, Oxoid, Basingstoke, UK), transferred in an anaerobic cabinet (Concept Plus, Ruskinn Technology, Ltd., Bridgend, UK) with an atmosphere of 85% N_2, 10% CO_2, and 5% H_2, and homogenized (10%, w/v) in sterile MM.

The fecal slurries (FS) were inoculated with a syringe (10% v/v) into butyl-rubber stoppered bottles containing 50 mL of sterile anaerobic MM. After an incubation of 72 h at 37 °C, these cultures were utilized to seed (10% v/v) the next passage in MM, repeated to accomplish three enrichment steps. Once grown, the third enrichment culture (EC) was utilized for isolation of mucinolytic bacteria. The viable counts of anaerobic mucinolytic bacteria and Enterobacteriaceae in EC samples were determined in triplicate onto agar plates of MM and MacConkey, respectively. For both the media, one-way ANOVA followed by Tukey's *post-hoc* analysis was utilized to compare the counts of the different subjects. Paired samples t-test was utilized to compare MM and MacConkey counts.

Isolation of anaerobic mucin degraders. All the steps of the isolation and purification procedure were carried out in the anaerobic cabinet. The EC was serially diluted (1:10) in anaerobic MM, then 1 mL of each dilution was poured into plates and covered with 20 mL of melted soft-agar MM. The plates were allowed to cool and incubated in anaerobiosis at 37 °C for 72–96 h. Inclusion colonies were picked and checked for aerobic growth at 37 °C in MacConkey agar plates. Colonies unable to grow in aerobiosis were further purified at least three times by serial dilution and isolation in soft-agar MM.

Approximately 10 pure clones per sample were fingerprinted by RAPD-PCR according to Quartieri et al.[37] and clustered into biotypes with a similarity level of 75% using the Pearson correlation coefficient. Single biotypes from each sample were taxonomically characterized by partial sequencing of the 16S rRNA gene (Eurofins Genomics, Ebersberg, Germany), utilizing primers targeting the V1–V4 regions[38]. Sequences were compared to the NCBI refseq_rna database with the alignment tools BLAST and MUSCLE (https://www.ncbi.nlm.nih.gov/; https://www.ebi.ac.uk/Tools/msa/muscle/).

16S rRNA gene profiling. Total DNA was extracted from 2 mL of FS and EC samples with the QiAmp PowerFecal DNA kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. The DNA was normalized to 5 ng/µL after quantification with a Qubit 3.0 fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA).

Partial 16S rRNA gene sequences were amplified using Probio_Uni/Probio_Rev primers, which targeted the V3 region of the 16S rRNA gene. Amplicons were sequenced using a MiSeq platform (Illumina Inc., San Diego, CA, USA) according to Milani et al.[39]. Raw sequences were analyzed with the QIIME 2.0 pipeline, version 2019.10[40], with appropriate plugins for trimming (CUTADAPT) and denoising (DADA2) into amplicon sequence variants (ASVs)[41,42]. Taxonomy assignment was carried out with the feature classifier VSEARCH[43],

with SILVA SSU database release 138 (https://www.arb-silva.de/download/arb-files/) as reference and the the similarity threshold set at 0.97. The appropriate QIIME2 plugins were utilized to compute the alpha- (observed taxa, Chao1, Shannon, and Pielou's evenness) and beta-diversity (Unweighted Normalized UniFrac, and Weighted Normalized UniFrac) and to compare them within and between FS and EC samples (i.e., the Kruskal–Wallis test for alpha diversity; ANOSIM and PERMANOVA for beta diversity). Differences were considered significant for $P < 0.05$. Principal Coordinate Analysis (PCoA) was computed with QIIME2, based on the beta-diversity distance matrices.

Linear discriminant analysis Effect Size (LEfSe, http://huttenhower.sph.harvard.edu/galaxy) algorithm was utilized to discover distinctive taxonomic features characterizing FS and EC samples[44]. Taxa presenting a significant differential abundance ($P < 0.05$) and logarithmic LDA (linear discriminant analysis) score > 2 were considered as microbial biomarkers of FS or EC samples.

The dataset was analyzed with PICRUSt2 to predict metagenomes' functions and investigate the presence of any difference between the FS and EC samples in terms of enzymes and pathways related to mucin metabolism[45]. The tool predicted the enzymes and the relative MetaCyc pathways abundances[46]. STAMP was used to visualize the results and conduct statistical analysis[47].

The 16S rRNA gene sequences datasets generated and analyzed during the current study are available in the NCBI repository with the BioProject ID: PRJNA649741 (https://www.ncbi.nlm.nih.gov/).

Received: 26 January 2021; Accepted: 7 May 2021
Published online: 27 May 2021

References

1. Bansil, R. & Turner, B. S. The biology of mucus: Composition, synthesis and organization. *Adv. Drug Deliv. Rev.* **124**, 3–15. https://doi.org/10.1016/j.addr.2017.09.023 (2018).
2. Alemao, C. A. *et al.* Impact of diet and the bacterial microbiome on the mucous barrier and immune disorders. *Allergy* https://doi.org/10.1111/all.14548 (2020) **(Online ahead of print)**.
3. Wagner, C. E., Wheeler, K. M. & Ribbeck, K. Mucins and their role in shaping the functions of mucus barriers. *Annu. Rev. Cell Dev. Biol.* **34**, 189–215. https://doi.org/10.1146/annurev-cellbio-100617-062818 (2018).
4. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6**, 81. https://doi.org/10.3389/fgene.2015.00081 (2015).
5. Corfield, A. P. The interaction of the gut microbiota with the mucus barrier in health and disease in human. *Microorganisms.* **6**, 78. https://doi.org/10.3390/microorganisms6030078 (2018).
6. Bergstrom, K. S. & Xia, L. Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* **23**, 1026–1037. https://doi.org/10.1093/glycob/cwt045 (2013).
7. Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* **14**, e1002533. https://doi.org/10.1371/journal.pbio.1002533 (2016).
8. Johansson, M. E., Larsson, J. M. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc. Natl. Acad. Sci.* **108**, 4659–4665. https://doi.org/10.1073/pnas.1006451107 (2011).
9. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32. https://doi.org/10.1038/nrmicro3552 (2016).
10. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. Mucins in the mucosal barrier to infection. *Mucosal Immunol.* **1**, 183–197. https://doi.org/10.1038/mi.2008.5 (2008).
11. Cornick, S., Tawiah, A. & Chadee, K. Roles and regulation of the mucus barrier in the gut. *Tissue Barriers.* **3**, e982426. https://doi.org/10.4161/21688370.2014.982426 (2015).
12. Troge, A. *et al.* More than a marine propeller–the flagellum of the probiotic *Escherichia coli* strain Nissle 1917 is the major adhesin mediating binding to human mucus. *Int. J. Med. Microbiol.* **302**, 304–314. https://doi.org/10.1016/j.ijmm.2012.09.004 (2012).
13. Raimondi, S. *et al.* Antibiotic resistance, virulence factors, phenotyping, and genotyping of *E. coli* isolated from the feces of healthy subjects. *Microorganisms.* **7**, 251. https://doi.org/10.3390/microorganisms7080251 (2019).
14. Shawki, A. & McCole, D. F. Mechanisms of intestinal epithelial barrier dysfunction by adherent-invasive *Escherichia coli*. *Cell. Mol. Gastroenterol. Hepatol.* **3**, 41–50. https://doi.org/10.1016/j.jcmgh.2016.10.004 (2016).
15. Ugalde-Silva, P., Gonzalez-Lugo, O. & Navarro-Garcia, F. Tight junction disruption induced by type 3 secretion system effectors injected by enteropathogenic and enterohemorrhagic *Escherichia coli*. *Front. Cell. Infect. Microbiol.* **6**, 87. https://doi.org/10.3389/fcimb.2016.00087 (2016).
16. Edwards, L. A., Bajaj-Elliott, M., Klein, N. J., Murch, S. H. & Phillips, A. D. Bacterial-epithelial contact is a key determinant of host innate immune responses to enteropathogenic and enteroaggregative *Escherichia coli*. *PLoS ONE* **6**, e27030. https://doi.org/10.1371/journal.pone.0027030 (2011).
17. Litvak, Y., Byndloss, M. X. & Bäumler, A. J. Colonocyte metabolism shapes the gut microbiota. *Science* **362**, eaat9076. https://doi.org/10.1126/science.aat9076 (2018).
18. Amaretti, A. *et al.* Profiling of protein degraders in cultures of human gut microbiota. *Front. Microbiol.* **10**, 2614. https://doi.org/10.3389/fmicb.2019.02614 (2019).
19. Lau, S. K. *et al.* Bacteraemia caused by *Anaerotruncus colihominis* and emended description of the species. *J. Clin. Pathol.* **59**, 748–752. https://doi.org/10.1136/jcp.2005.031773 (2006).
20. Togo, A. H. *et al. Anaerotruncus massiliensis* sp. nov., a succinate-producing bacterium isolated from human stool from an obese patient after bariatric surgery. *New Microbes New Infect.* **29**, 100508. https://doi.org/10.1016/j.nmni.2019.01.004 (2019).
21. Liang, J. Q. *et al.* Fecal microbial DNA markers serve for screening colorectal neoplasm in asymptomatic subjects. *J. Gastroenterol. Hepatol.* https://doi.org/10.1111/jgh.15171 (2020) **(Online ahead of print)**.
22. Bo, J. *et al. Eggerthella lenta* bloodstream infections: Two cases and review of the literature. *Future Microbiol.* **15**, 981–985. https://doi.org/10.2217/fmb-2019-0338 (2020).
23. García-Solache, M. & Rice, L. B. The *Enterococcus*: A model of adaptability to its environment. *Clin. Microbiol. Rev.* **32**, e00058-e118. https://doi.org/10.1128/CMR.00058-18 (2019).
24. Horn, N. *Clostridium disporicum* sp. nov., a saccharolytic species able to form two spores per cell, isolated from a rat cecum free. *Int. J. Syst. Bacteriol.* **37**, 398–401. https://doi.org/10.1099/00207713-37-4-398 (1987).
25. Tawthep, S. *et al.* Isolation of six novel 7-oxo- or urso-type secondary bile acid-producing bacteria from rat cecal contents. *J. Biosci. Bioeng.* **124**, 514–522. https://doi.org/10.1016/j.jbiosc.2017.06.002 (2017).

26. Muñoz, M. *et al.* Comparative genomics identifies potential virulence factors in *Clostridium tertium* and *C. paraputrificum*. *Virulence.* **10**, 657–676. https://doi.org/10.1080/21505594.2019.1637699 (2019).
27. Sasi Jyothsna, T. S., Tushar, L., Sasikala, C. & Ramana, C. V. *Paraclostridium benzoelyticum* gen. nov., sp. nov., isolated from marine sediment and reclassification of *Clostridium bifermentans* as *Paraclostridium bifermentans* comb. nov. Proposal of a new genus *Paeniclostridium* gen. nov. to accommodate *Clostridium sordellii* and *Clostridium ghonii*. *Int. J. Syst. Evol. Microbiol.* **66**, 1268–1274. https://doi.org/10.1099/ijsem.0.000874 (2016).
28. Low, K. E., Smith, S. P., Abbott, D. W. & Boraston, A. B. The glycoconjugate-degrading enzymes of *Clostridium perfringens*: Tailored catalysts for breaching the intestinal mucus barrier. *Glycobiology* https://doi.org/10.1093/glycob/cwaa050 (2020).
29. Kiu, R. *et al.* Preterm infant-associated *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* strains: Genomic and evolutionary insights. *Genome Biol. Evol.* **9**, 2707–2714. https://doi.org/10.1093/gbe/evx210 (2017).
30. Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105**, 2420–2428. https://doi.org/10.1038/ajg.2010.281 (2010).
31. Wexler, H. M. Bacteroides: The good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621. https://doi.org/10.1128/CMR.00008-07 (2007).
32. Zhou, J. C. & Zhang, X. W. *Akkermansia muciniphila*: A promising target for the therapy of metabolic syndrome and related diseases. *Chin. J. Nat. Med.* **17**, 835–841. https://doi.org/10.1016/S1875-5364(19)30101-3 (2019).
33. Holmén Larsson, J. M., Thomsson, K. A., Rodríguez-Piñeiro, A. M., Karlsson, H. & Hansson, G. C. Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *Am. J. Physiol. Gastrointest. Liver Physiol.* **305**, G357–G363. https://doi.org/10.1152/ajpgi.00048.2013 (2013).
34. Van den Abbeele, P. *et al.* Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J.* **7**, 949–961. https://doi.org/10.1038/ismej.2012.158 (2013).
35. Marczynski, M. *et al.* Structural alterations of mucins are associated with losses in functionality. *Biomacromol* **22**, 1600–1613. https://doi.org/10.1021/acs.biomac.1c00073 (2021).
36. Miller, R. S. & Hoskins, L. C. Mucin degradation in human colon ecosystems. Fecal population densities of mucin-degrading bacteria estimated by a "most probable number" method. *Gastroenterology* **81**, 759–765 (1981).
37. Quartieri, A. *et al.* Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe* **38**, 130–137. https://doi.org/10.1016/j.anaerobe.2015.10.006 (2016).
38. Raimondi, S. *et al.* Microbiota of sliced cooked ham packaged in modified atmosphere throughout the shelf life: Microbiota of sliced cooked ham in MAP. *Int. J. Food Microbiol.* **289**, 200–208. https://doi.org/10.1016/j.ijfoodmicro.2018.09.017 (2019).
39. Milani, C. *et al.* Assessing the fecal microbiota: An optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS ONE* **8**, e68739. https://doi.org/10.1371/journal.pone.0068739 (2013).
40. Bolyen, E. *et al.* Author correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 1091. https://doi.org/10.1038/s41587-019-0252-6 (2019).
41. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–17. https://doi.org/10.14806/ej.17.1.200 (2011).
42. Callahan, B. J. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods.* **13**, 581–583. https://doi.org/10.1038/nmeth.3869 (2016).
43. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **4**, e2584. https://doi.org/10.7717/peerj.2584 (2016).
44. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60. https://doi.org/10.1186/gb-2011-12-6-r60 (2011).
45. Douglas, G. M. *et al.* PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* **38**, 685–688. https://doi.org/10.1038/s41587-020-0548-6 (2020).
46. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes—A 2019 update. *Nucleic Acids Res.* **48**, D445–D453. https://doi.org/10.1093/nar/gkz862 (2020).
47. Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123–3124. https://doi.org/10.1093/bioinformatics/btu494 (2014).

Author contributions

S.R. and M.R. conceived the study and supervised experiments; S.R. and E.M carried out enrichment cultures and isolation; F.C. and A.A. carried out metagenome analysis, bioinformatics, and data presentation; A.A. and M.R. wrote the main manuscript text; all authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-90553-4.

Correspondence and requests for materials should be addressed to M.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

International Journal of
Molecular Sciences

MDPI

Article

Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of Non-*Escherichia coli* Enterobacterales from the Gut Microbiota of Healthy Subjects

Alberto Amaretti [1,2,†], Lucia Righini [1,†], Francesco Candeliere [1], Eliana Musmeci [1], Francesca Bonvicini [3], Giovanna Angela Gentilomi [3,4], Maddalena Rossi [1,2] and Stefano Raimondi [1,*]

1 Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 103, 41125 Modena, Italy; alberto.amaretti@unimore.it (A.A.); lucia.righini@unimore.it (L.R.); francesco.candeliere@unimore.it (F.C.); eliana.musmeci@unimore.it (E.M.); maddalena.rossi@unimore.it (M.R.)
2 Biogest-Siteia, University of Modena and Reggio Emilia, Modena, Viale Amendola 2, 42122 Reggio Emilia, Italy
3 Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, Via Massarenti 9, 40138 Bologna, Italy; francesca.bonvicini4@unibo.it (F.B.); giovanna.gentilomi@unibo.it (G.A.G.)
4 Unit of Microbiology, Alma Mater Studiorum-University of Bologna, S. Orsola-Malpighi Hospital, Via Massarenti 9, 40138 Bologna, Italy
* Correspondence: stefano.raimondi@unimore.it; Tel.: +39-059-205-8595
† These authors contributed equally to the work.

Received: 2 March 2020; Accepted: 5 March 2020; Published: 7 March 2020

Abstract: Non-*Escherichia coli* Enterobacterales (NECE) can colonize the human gut and may present virulence determinants and phenotypes that represent severe heath concerns. Most information is available for virulent NECE strains, isolated from patients with an ongoing infection, while the commensal NECE population of healthy subjects is understudied. In this study, 32 NECE strains were isolated from the feces of 20 healthy adults. 16S rRNA gene sequencing and mass spectrometry attributed the isolates to *Klebsiella pneumoniae*, *Klebsiella oxytoca*, *Enterobacter cloacae*, *Enterobacter aerogenes*, *Enterobacter kobei*, *Citrobacter freundii*, *Citrobacter amalonaticus*, *Cronobacter* sp., and *Hafnia alvei*, *Morganella morganii*, and *Serratia liquefaciens*. Multiplex PCR revealed that *K. pneumoniae* harbored virulence genes for adhesins (*mrkD*, *ycfM*, and *kpn*) and enterobactin (*entB*) and, in one case, also for yersiniabactin (*ybtS*, *irp1*, *irp2*, and *fyuA*). Virulence genes were less numerous in the other NECE species. Biofilm formation was spread across all the species, while curli and cellulose were mainly produced by *Citrobacter* and *Enterobacter*. Among the most common antibiotics, amoxicillin-clavulanic acid was the sole against which resistance was observed, only *Klebsiella* strains being susceptible. The NECE inhabiting the intestine of healthy subjects have traits that may pose a health threat, taking into account the possibility of horizontal gene transfer.

Keywords: Enterobacterales; *Klebsiella*; *Enterobacter*; *Citrobacter*; virulence; antibiotic resistance; biofilm

1. Introduction

Important enteric pathogens belong to Enterobacterales, a bacterial order within the phylum Proteobacteria. This order encompasses permanent colonizers of the human gut that, in healthy conditions, constitute minor bacterial components of the microbiota. Opportunistic Enterobacterales can persist as gut commensals without inducing any infections, as long as the microbiota is balanced and

the complex and dense bacterial community prevents their overgrowth. A bloom of Enterobacterales may occur as a result of disturbance of the microbiota, yielding pathogen-mediated infections and triggering inflammatory host responses.

Escherichia coli is the most studied among the Enterobacterales with regards to the traits that differentiate commensalism and pathogenicity. It normally colonizes the intestine but comprises both harmless commensals and different pathogenic variants that may instigate infections in the gut or in other tissues [1,2]. Virulent strains of *E. coli* isolated from infected patients attracted most research interest [3,4], but also fecal isolates from healthy subjects and environmental strains are the target of increasing attention, aiming to determine the pathogenic potential of a wider biodiversity reservoir [5,6].

Many non-*E. coli* Enterobacterales (hereinafter referred to as NECE) that can colonize the gut (e.g., *Klebsiella*, *Enterobacter*, *Citrobacter*, and *Serratia*) also present traits that can confer them virulence and pathogenicity or phenotypes that may result in severe heath concern, such as multidrug resistance [7–10]. The greatest efforts have been carried out to describe virulent strains, generally isolated from patients with an ongoing infection, while the pathogenic potential of NECE inhabiting the gut of healthy subjects has not been thoroughly investigated with genetic and phenotypic analysis, except for some genera [9]. The research herein presented aims to fill this gap, providing a genotypic and phenotypic description of the NECE population isolated from the feces of 20 healthy adults, and to complement a previous study that described the *E. coli* population of the same cohort of subjects [5]. A set of 32 NECE strains was isolated, taxonomically classified, subjected to PFGE genotyping, and described in terms of the genotypic determinants and phenotypic traits that may confer on them potential pathogenicity or invasiveness.

Information on the genes associated to virulence is detailed especially for *Klebsiella* spp., where a number of genes associated to harmful traits were identified, such as those encoding adhesins, siderophores (e.g., enterobactin, aerobactin, yersiniabactin), protectines, or invasins (responsible for mucoid phenotype and invasiveness), and involved in allantoin metabolism [7,11]. For other NECE genera, such as *Enterobacter*, *Cronobacter*, and *Citrobacter*, the knowledge of the genetic determinants associated to virulence and invasiveness is less comprehensive and, with few exceptions (e.g., *Citrobacter koseri*), mainly acquired from better characterized pathogens [12–15].

In addition to fimbrial and afimbrial adhesins, the production of surface cellulose structures and curli favors the adhesion of Enterobacterales and can exert a significant role in enteric biofilm-related infections [16,17]. Although not directly involved in pathogenic mechanisms, the acquisition of multiple antibiotic resistances favors the success of opportunistic Enterobacterales pathogens in invasion, survival, and spread, severely complicating the containment and treatment of infections [9,18]. Therefore, the occurrence of drug resistant bacteria within a commensal population and the possibility to exchange genetic material by horizontal gene transfer may represent a major health concern.

In the present study, multiplex PCR assays were utilized to screen the NECE, isolated from the feces of healthy subjects, for the presence of 17 main virulence genes associated to those of *Klebsiella* and *E. coli* [19–21]. From the point of view of the phenotype, the isolates were characterized for the ability to form biofilm and to yield curli and surface cellulose, were screened for the susceptibility to the most common antibiotics and for the ability to act as recipients in conjugation experiments, and biochemical tests were performed to compare the metabolic profile.

2. Results

2.1. Counting and Isolation of NECE

The selective differential medium HiCrome Coliform Agar (HCCA) was utilized to count and isolate *E. coli* [5] and NECE from fecal samples of 20 healthy adults. Total counts in HCCA ranged from 4.6×10^5 to 2.2×10^8 cfu/g (Figure 1). Blue colonies attributed to *E. coli* overcounted the pink ones attributed to NECE in all the samples except V11 (Figure 1; Supplementary Figure S1). NECE ranged between $< 10^4$ and 1.3×10^8 cfu/g and, except in V11, accounted for a minority of total Enterobacterales

(NECE + *E. coli*), the 75th percentile being the 5.7% (Figure 1). In some cases, NECE were not recovered, being outnumbered by *E. coli*. Spearman's rank correlation analysis excluded any significant correlation between NECE and *E. coli* counts.

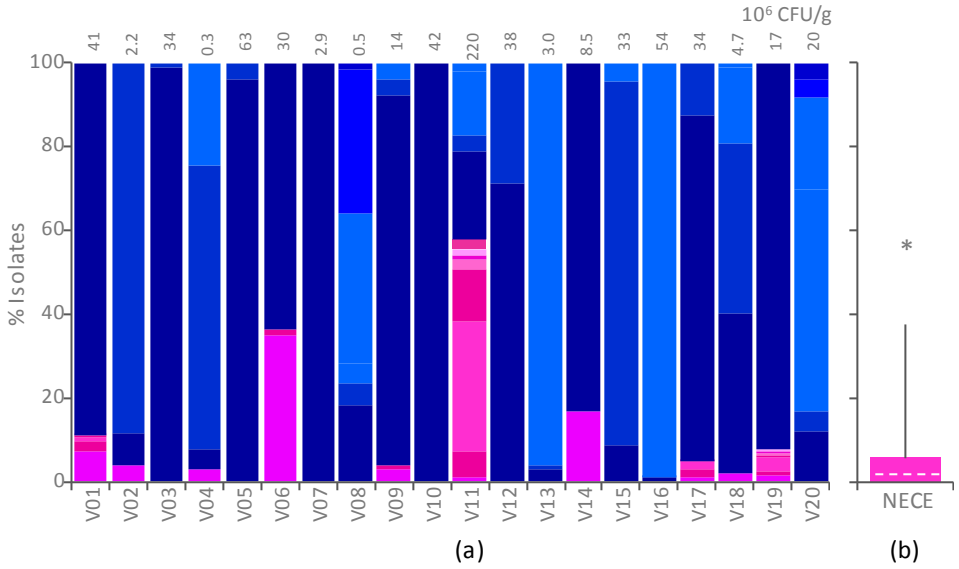


Figure 1. Counts of *Escherichia coli* and non-*Escherichia coli* Enterobacterales (NECE), enumerated onto HiCrome Coliform Agar (HCCA) plates. (**a**) Percentage of colonies attributed to *E. coli* (blue shades) and NECE (pink shades) in the feces of 20 subjects. For each subject, different shades indicate different biotypes according to enterobacterial repetitive intergenic consensus-PCR (ERIC-PCR) and random amplification of polymorphic DNA-PCR (RAPD-PCR) fingerprinting. The total count of Enterobacterales (*E. coli* + NECE) is reported in the top margin. (**b**) Distribution of the percentage of NECE colonies. The median (dashed line), the 25th and 75th percentiles (colored box), the 10th and 90th percentiles (whiskers), and outliers (*) are indicated.

2.2. Taxonomic Attribution and PFGE Genotyping

The isolates putatively attributed to NECE were clustered in 32 different biotypes utilizing ERIC-PCR (enterobacterial repetitive intergenic consensus-PCR) and RAPD-PCR (random amplification of polymorphic DNA-PCR) fingerprinting. A representative isolate of each biotype was assigned a taxonomic designation utilizing 16S rRNA gene sequencing and MALDI-TOF MS (Supplementary Table S1). The genus *Klebsiella* was the most represented (14/32), with the species *K. pneumoniae* (10 strains) and *Klebsiella oxytoca* (4) found in seven and three fecal samples, respectively. The genus *Enterobacter* was represented by eight strains belonging to *Enterobacter cloacae* (6), *Enterobacter aerogenes*, and *Enterobacter kobei* (1 strain each). Other strains belonged to *Citrobacter* (4 to *Citrobacter freundii* and 1 to *Citrobacter amalonaticus*), *Cronobacter* sp. (2), and to *Hafnia alvei*, *Morganella morganii*, and *Serratia liquefaciens* (1 strain each).

PFGE highlighted a wide diversity of the NECE isolates, which did not cluster according to the taxonomic attribution (Figure 2).

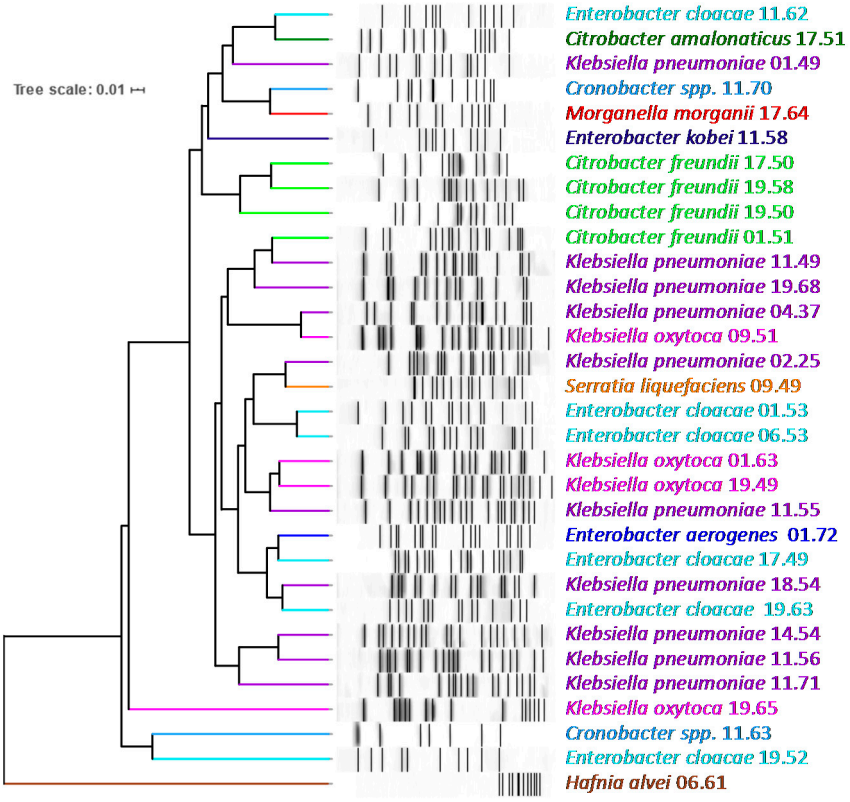


Figure 2. *Xba*I-PFGE pattern of NECE strains: unweighted pair group method with arithmetic means (UPGMA) dendrogram derived from Dice's coefficients, calculated based on the band profile. Strains are colored based on their MALDI-TOF MS taxonomic attribution.

2.3. Virulence Genotyping

PCR was used to investigate 17 virulence genes encoding adhesins (*fimH1*, *mrkD*, *kpn*, and *ycfM*), siderophores (enterobactin, *entB*; aerobactin, *iutA*; yersiniabactin, *irp-1*, *irp2*, *ybtS*, *fyuA*; catechols receptor, *iroN*; and other, *kfu*), protectines or invasins (*K2*, *magA*, *rmpA*, and *traT*), and involved in allantoin metabolism (*allS*).

Most strains of *K. pneumoniae* harbored the *mrkD*, *ycfM*, and *kpn* encoding adhesins and *entB* encoding enterobactin (Figure 3). Only *K. pneumoniae* 11.55 was positive to the main virulence genes involved in the synthesis of *Yersinia* siderophore, including *ybtS* (encoding for the synthase), *irp1* and *irp2* (for regulatory proteins), and *fyuA* (for the siderophore receptor). *irp2* was also detected in most of the other strains of *K. pneumoniae* although they lacked the counterpart *ybtS*. All the *K. pneumoniae* strains were negative to the genes *K2*, *magA*, and *rmpA* associated with hypermucoid phenotype and invasivity, except for *K. pneumoniae* 01.49 that was positive to *K2*. Similarly, other virulence genes, such as *allS*, *kfu*, and *iutA* occurred only once among the tested strains.

The strains of *K. oxytoca* harbored *entB* (three out of four isolates) but were negative to most of the other virulence genes. A sole strain harbored *ytbS*. Most of *Cronobacter* and *Enterobacter* isolates were characterized by the presence of the gene *irp2* but never harbored *ybtS* or other *Yersinia* siderophore genes. A few strains were positive to *mrkD* or *entB*.

The strains ascribed to *Citrobacter*, *H. alvei*, and *M. morganii* were negative to those virulence genes whose presence could not be excluded by primer-blast search. The strain of *S. liquefaciens* was positive to *ytbS*.

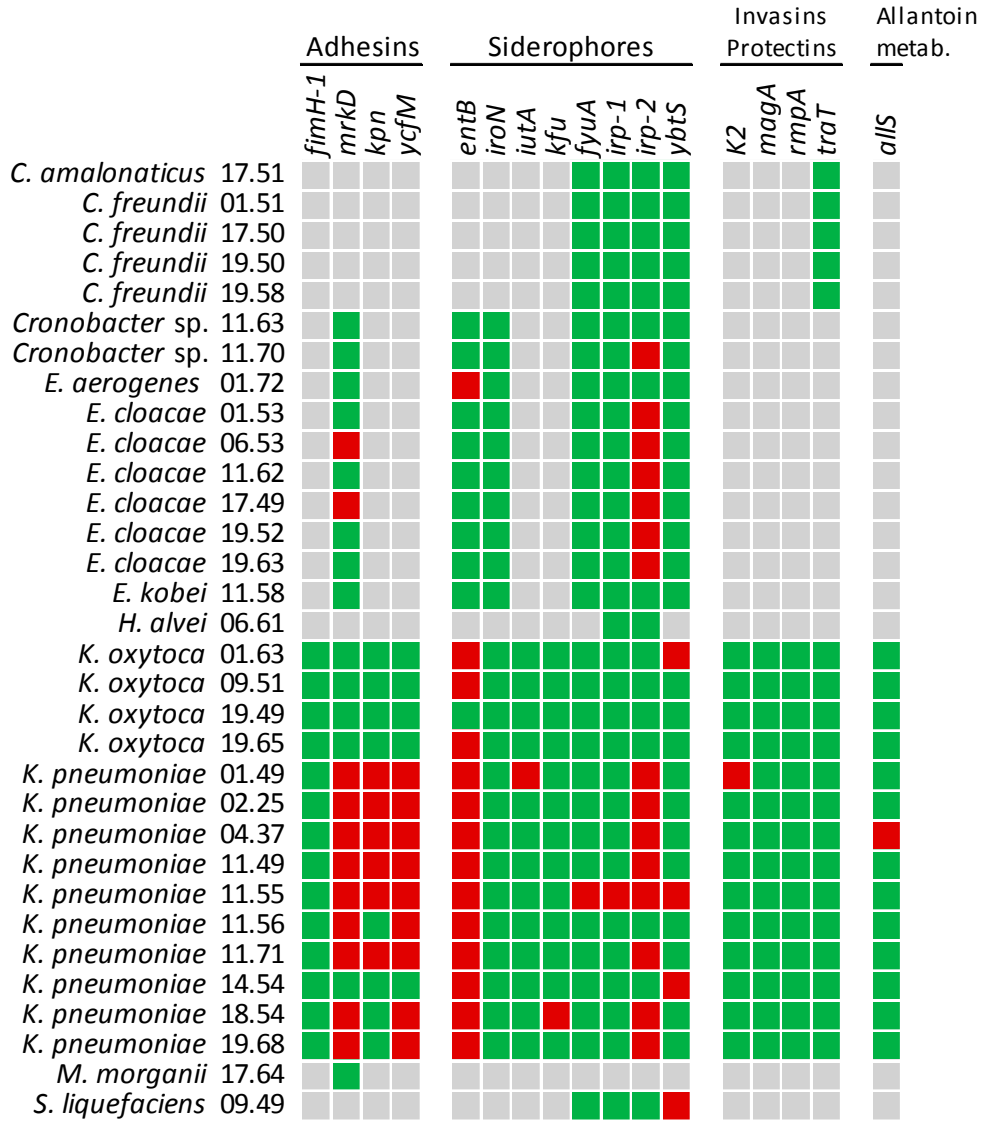


Figure 3. PCR assay of the NECE isolates for the presence of virulence genes. Colors: red, positive amplification; green, negative amplification; grey, PCR analysis not performed since the gene was putatively absent based on a primer-blast search.

2.4. Biofilm Formation and Production of Curli and Cellulose

NECE strains were tested for biofilm formation in minimal and rich media (M9 and LBWS, respectively; Supplementary Figure S2). The vast majority of the strains (26 out of 32), belonging to all the species except *E. aerogenes*, *H. alvei*, and *M. morganii*, formed biofilm in M9 (Figure 4; Supplementary Figure S2). Biofilm formation was less frequent in LBWS, being observed only in 10 strains of *K. oxytoca*, *K. pneumoniae,* and *S. liquefaciens*. The extent of biofilm production was always less abundant in the rich medium compared to M9 ($p < 0.05$).

Extracellular cellulose was detected in most of *Citrobacter* and *Enterobacter* strains, in five strains of *K. pneumoniae* and two out of four of *K. oxytoca*, in *H. alvei* and in *S. liquefaciens*. Curli were produced by nearly all *Citrobacter*, *Cronobacter*, and *Enterobacter* strains and by one isolate belonging to *K. oxytoca*. The isolates of *H. alvei* and *M. morganii* were also positive to curli. The strains belonging to *Citrobacter*, *E. cloacae*, *H. alvei,* and *K. oxytoca* 19.49 produced both cellulose and curli.

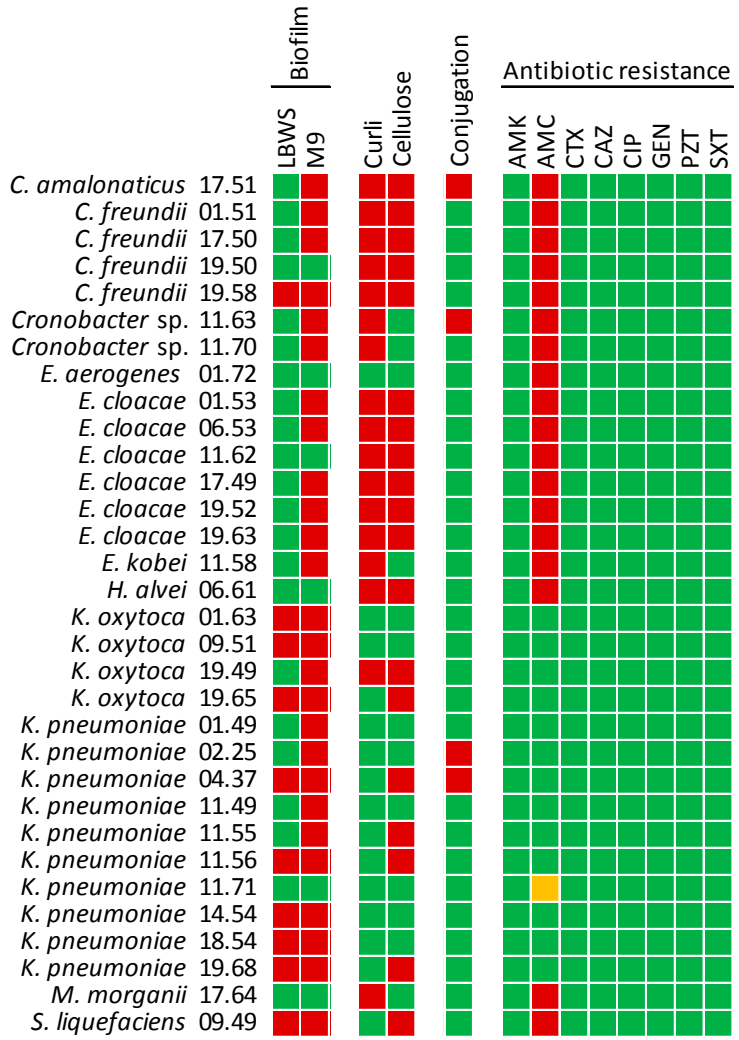


Figure 4. Phenotypic characterization of NECE isolates: biofilm formation in LBWS and M9 media, curli and cellulose production, conjugation, and antibiotic resistance. Colors: red, positive; green, negative. For antibiotics: red, resistant; green, susceptible, yellow, intermediate. Antibiotics: amikacin (AMK), amoxicillin–clavulanic acid (AMC), cefotaxime (CTX), ceftazidime (CAZ), ciprofloxacin (CIP), gentamicin (GEN), piperacillin-tazobactam (PZT), and trimethoprim-sulfamethoxazole (SXT).

2.5. Conjugation

The strains were challenged as conjugation recipients for receiving pOX38: Cm plasmid from *E. coli* N4i. Only two strains of *K. pneumoniae*, and single strains of *Cronobacter* and *Citrobacter amalonaticus* succeeded in plasmid acquisition (Figure 4).

2.6. Antibiotic Resistance

Phenotypic susceptibility to amikacin, amoxicillin–clavulanic acid, cefotaxime, ceftazidime, ciprofloxacin, gentamicin, piperacillin-tazobactam, and trimethoprim-sulfamethoxazole was assayed. Amoxicillin-clavulanic acid was the sole antibiotic against which few isolates presented some resistance, with all the strains of *Enterobacter*, *Citrobacter*, *Cronobacter*, *H. alvei*, *M. morganii*, and *S. liquefaciens* being resistant. All the 14 biotypes of *Klebsiella* spp. were sensitive to the whole set of tested antibiotics, with the exception of *K. pneumoniae* 11.71 that was partially resistant to amoxicillin–clavulanic acid, presenting a minimum inhibitory concentration (MIC) intermediate between resistance and susceptibility thresholds.

2.7. Biochemical Characterization

The fermentation of substrates and some distinctive enzymatic reactions and metabolic routes were assayed utilizing the API 20 E system (Figure 5). Generally, the NECE strains were positive to β-galactosidase. Most strains were capable of utilizing citrate, glucose, mannose, inositol, sorbitol, rhamnose, sucrose, melibiose, amygdalin, and arabinose. The main exceptions were *M. morganii* that could utilize only glucose, *H. alvei* that fermented a restricted number of sugars, and some *Citrobacter*, *Cronobacter*, and *Enterobacter* strains that exhibited specific substrate preferences. The majority of the isolates produced either lysine decarboxylase (*Klebsiella*) or ornithine decarboxylase (*Cronobacter* and *Enterobacter*). Urease was characteristic of *Klebsiella*, while arginine dihydrolase was found in most *Enterobacter*, *Citrobacter*, and *Cronobacter*. Acetoin was produced by *Cronobacter*, *Enterobacter*, *Hafnia*, and *Klebsiella*. Indole was produced by *K. oxytoca* and by few other strains, H_2S by two strains of *Citrobacter freundii*. Only *S. liquefaciens* was positive to gelatinase. All the strains except *M. morganii* and *S. liquefaciens* exhibited denitrifying activity, in most cases yielding nitrite. Nitrate reduction to N_2 was observed in *K. pneumoniae* and few other strains.

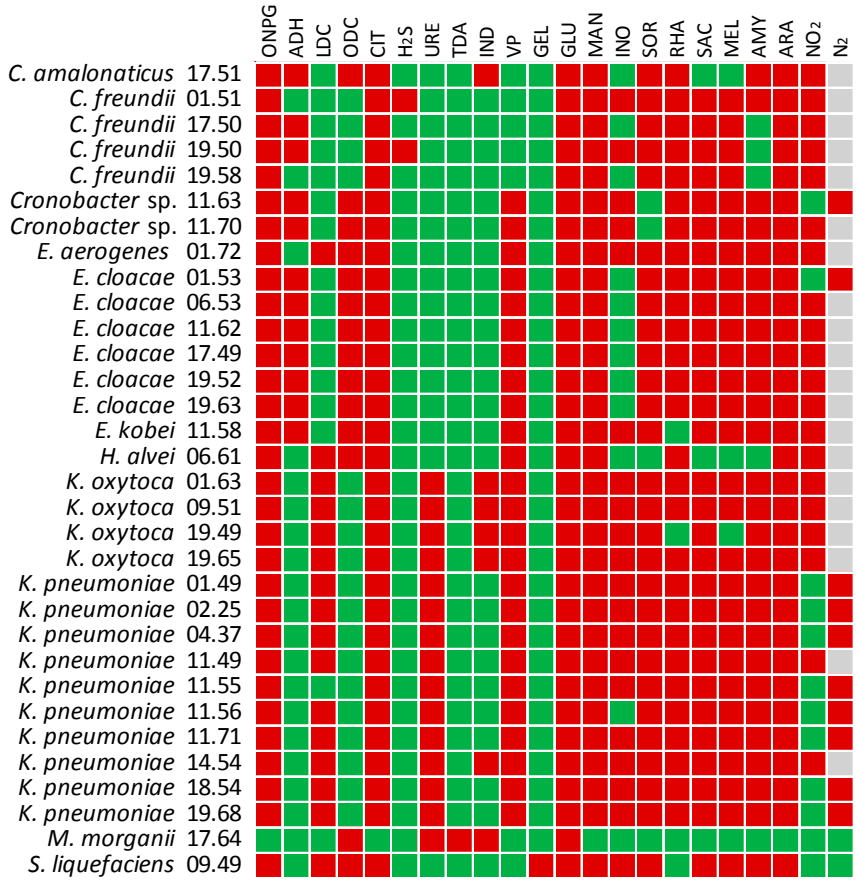


Figure 5. Biochemical reaction profiles of NECE isolates in the API 20 E assay: β-galactosidase (ONPG), arginine dihydrolase (ADH), lysine decarboxylase (LDC), ornithine decarboxylase (ODC), citrate utilization (CIT), production of hydrogen sulfide (H_2S), urease (URE), tryptophan deaminase (TDA), indole (Kovac's test, IND), acetoin (Voges-Proskauer test, VP), gelatinase (GEL), fermentation of glucose (GLU), mannitol (MAN), inositol (INO), sorbitol (SOR), rhamnose (RHA), sucrose (SAC), melibiose (MEL), amygdalin (AMY), arabinose (ARA), and reduction of nitrates to nitrites (N_2O) or nitrogen (N_2, tested only in case of negative N_2O). Colors: red, positive; green, negative.

3. Discussion

Thirty-two NECE strains were isolated from the feces of 20 healthy adults that did not present any dysbiosis, and thus as members of a relatively balanced gut microbiota. The load of Enterobacterales was in the order of millions or tens of millions per gram of feces, with a sole exception where they reached the magnitude of 10^8. NECE represented a small population of Enterobacterales, with *E. coli* being on average 20 times more abundant, and a minor component of the whole microbiota, being less than 0.1%. The genus *Klebsiella*, which is ubiquitous in nature, colonizing humans, animals, and plants, and frequently detected in waters, sewages, and soils, was the most represented, encompassing 14 of the 32 strains. *K. pneumoniae*, the most frequently isolated species (10 strains), was present in 7 out of 20 fecal samples, in accordance to the 35% of healthy adults from which it was isolated in a pioneering study [22].

K. pneumoniae encompasses opportunistic pathogens that can cause human infections in lungs, urinary tract, and bloodstreams, mostly to hospitalized and/or immunocompromised patients [23,24]. Virulence of *K. pneumoniae* is associated to the presence of capsule and pili, to the production of lipopolysaccharides and siderophores, to allantoin utilization, and to iron uptake systems, efflux pumps, and type VI secretion systems [7]. Surface molecules, such as capsular polysaccharides and lipopolysaccharides, are some of the major virulence factors that *Klebsiella* use to protect itself from the host innate immune apparatus. Furthermore, the capability to compete for iron has a pivotal role to the establishment of the infection. The genes involved in iron assimilation are generally clustered in pathogenicity islands, large chromosomal regions that were likely acquired by horizontal transfer. Virulence genetic determinants can be located both in the core or in the accessory genome [7], the former also including metabolic genes required for some species to cause disease.

Among the 10 *K. pneumoniae* isolates, all but one strain harbored *entB*, encoding a common catecholate siderophore located in the core genome, and *mrkD*, involved in the synthesis of type 3 pili that promote adherence to the surfaces [19]. The genes encoding other adhesins, such as *ycfM* and *kpn* were detected in nine and five strains, respectively. Only two *K. pneumoniae* strains were positive to the yersiniabactin gene *ybtS*, a common virulence factor associated with human infections [25,26], and one of the two also harbored *irp1*. These genes are involved in the synthesis of the siderophore yersiniabactin by virulent *Yersinia* strains, which harbor them within the high-pathogenicity island (HPI). HPI is widely distributed among members of the order Enterobacterales, including *E. coli*, *K. pneumoniae*, *Citrobacter* spp., *Salmonella enterica*, *Serratia liquefaciens*, and *Enterobacter* spp. [27,28]. In addition, *irp2*, another marker of HPI, was detected in the majority of the strains of *K. pneumonia* and *E. cloacae*, albeit they lacked the counterpart *ybtS*.

Although Enterobacterales are normally present as a low fraction of commensal bacteria in the healthy gut, their numbers can increase in the inflamed gut, and take advantage over other commensals [29]. Siderophores are major contributors of exploitative competition, since iron is an essential nutrient present in very low amounts in the gut and may play a role in virulence. Overall, *K. pneumoniae* 11.55 was the strain equipped with the broadest set of virulence genes involved in iron metabolism, including those encoding enterobactin (*entB*), yersiniabactin (*irp1*, *irp2*, and *ybtS)*, and, the sole strain among the 32 isolates, also the *Yersinia* siderophore receptor (*fyuA*).

In agreement with their commensal behavior, none of the *K. pneumoniae* strains were positive for the two genes associated with invasive infections, i.e., the mucoviscosity-associated gene *magA* and the regulator of mucoid phenotype *rmpA* [30,31], that are associated with a hypervirulent and hypermucoid phenotype. In our isolates, the presence of other genetic determinants of virulence was sporadic or quite rare. In the gut of the healthy host, the Enterobacterales resides in the outer loose mucus layer, separated from the epithelial mono-cell barrier by an inner dense mucin coat [32]. During the infection, they penetrate the mucus layer, interact with the epithelial cells, and may breach the mucosal barrier. In case these enterobacteria strains reach other tissues, biofilm formation may act as a fitness factor concurring to pathogenesis [33]. Adhesion factors and extracellular matrix components are involved in formation of biofilms [34]. All but one strain produced biofilm in minimal medium M9, whereas this

phenotype was less frequent when growth occurred in the rich medium LBWS. The higher extent of biofilm formation in M9 is consistent with the fact that this mineral medium is more challenging for these bacteria. Extracellular cellulose structures, determined in five strains, were consistent with the capability to form biofilm on both media. Curli were found in a sole strain that presented also cellulose structures (19.58 CA) and produced biofilm on both the substrates.

Based on genetic and functional features, some *K. pneumoniae* isolates present a higher potential to cause infections, albeit they are present at low charge in the microbiota of healthy hosts. The link between colonization and infection by *K. pneumoniae* in hospitalized patients has been demonstrated, with robust evidence that their own microbiota is the main source of the infective strains [35,36]. Thus, potentially more virulent *K. pneumoniae* strains may take advantage of critical conditions, becoming responsible for nosocomial infections [37].

The isolates of *K. oxytoca*, another prominent pathogen that may be involved in diseases and life-threatening infections [38,39], generally encoded the siderophore gene *entB*, but were negative to most of the other virulence genes. A sole strain harbored the gene encoding the yersiniabactin siderophore. Biofilm production was a general feature of the *K. oxytoca* strains, regardless of the presence of curli or cellulose structures.

Similar considerations are valid for the other NECE strains herein described, belonging to genera that may have clinical relevance, such as *Enterobacter*, *Citrobacter*, and *Cronobacter*. The isolates were generally capable of forming biofilm and producing curli and cellulose and were negative for most virulence genes. However, unlike *Klebsiella* that shares many virulence genes with *E. coli*, the genetic determinants of virulence of these genera have not been fully disclosed.

In this study emerged that NECE isolates from feces of healthy subjects are still quite susceptible to most of the antibiotics. This is important, since any treatment of opportunistic outbreaks of NECE (e.g., in case nosocomial infections) requires antibiotics, but resistance developments would seriously curb the therapeutic options [40]. All the isolates of *Klebsiella* were sensitive to the whole set of tested antibiotics. Amoxicillin-clavulanic acid was the sole antibiotic against which was detected some resistance: the strains ascribed to the genus *Enterobacter* and the isolates belonging to *Citrobacter*, *Cronobacter*, *Hafnia alvei*, *Morganella morganii*, and *Serratia liquefaciens* were all resistant to this combination of antibiotics. Interestingly, Enterobacterales presented the highest increase in terms of relative abundance in a short-term amoxicillin-clavulanic acid treatment in healthy adults [41]. Some genera belonging to this family, such as *Enterobacter* and *Citrobacter*, are recognized as intrinsically resistant [42], and may take advantage to this antibiotic treatment. In general, the profile of resistance was independent by the subject of the fecal sample, but two clusters encompassing sensitive or resistant strains were sharply differentiated by taxonomy.

PFGE genotyping was carried out to evaluate the genetic similarity among the bacterial isolates of this study and highlighted a wide dispersion of the strains, regardless their taxonomic attribution and phylogenetic relationships. The spreading of the strains regardless of species or genera may be attributed to the presence of plasmids, the horizontal acquisition of additional genes from diverse species of Enterobacterales, and to the exchange of mobile elements that rapidly integrate and promote DNA shuffling, in agreement with the capability of some strains to accept DNA by conjugation from *E. coli* as a donor. However, the biochemical profiling mostly clustered the strains in species (data not shown), confirming that energy production and conservation, and lipid, amino acid, and nucleotide metabolism are part of the conserved reactome, despite the genome plasticity.

The Enterobacterales encountered in this study are generally recognized as opportunistic pathogens, with some potential capability to cause disease, on the basis of predicted virulence factors. Except for *K. pneumoniae* hypervirulent strains, NECE virulence seems more associated to the host features than to the strain traits. According to the similar results obtained for the *E. coli* isolated from the same cohort of healthy subjects, the absence of antibiotic resistance for most of the tested antibiotics does not pose a serious challenge for infection control. This highlights the stratification

of antibiotic resistance distribution among healthy and hospitalized/diseased subjects, with NECE associated risk increasing with both illness and antibiotic therapy.

4. Materials and Methods

4.1. Isolation and Enumeration of Enterobacterales

Fresh fecal samples were collected from 20 healthy adult subjects who gave written informed consent regarding their participation in the study in accordance with the protocol approved by the local research ethics committee (reference number: 974/2019/SPER-UNIMO-ENTEROPOP; Comitato etico dell'Area Vasta Emilia Nord, Italy). The subjects—10 males and 10 females aged 35 to 45, following a western omnivore diet, and who had not been treated with prebiotics and/or probiotics for 1 month and antibiotics for 3 months—were enrolled among the employees of the University of Modena and Reggio Emilia and their relatives and were not in relationship with the researchers.

Feces were homogenized (10% *w/v*) in isotonic Buffered Peptone Water (Sigma, Steinheim, Germany), then serial dilutions were spread onto plates of HiCrome Coliform Agar (HCCA, Sigma) and incubated at 37 °C. The medium differentiates *E. coli* (blue colonies) from NECE (salmon to red). For each subject, up to 48 colonies of putative NECE were picked and clustered into biotypes with ERIC-PCR [43] and RAPD-PCR [44] fingerprint presenting Pearson's similarity > 75%.

4.2. PFGE Genotyping

PFGE was performed according to PulseNet protocol (http://www.cdc.gov/pulsenet/PDF/ecoli-shigella-salmonella-pfge-protocol-508c.pdf). The genomic DNA was digested with 50 U of *Xba*I at 37 °C for 3 h. Fragments were resolved in a CHEF-DRIII apparatus (Bio-Rad, Hercules, CA, USA) using counter-clamped homogeneous electric field electrophoresis (24 h at 6.0 V/cm; initial switch time, 2.2 s; final switch time, 54.2 s). The run was digitally captured and analyzed with GelCompare II 6.0 software (Applied Maths NV, Sint-Martens-Latem, Belgium). Dice coefficient was computed to evaluate similarity between band profiles (position tolerance, 1%; optimization, 1%) and to derive an UPGMA dendrogram (unweighted pair group method with arithmetic means). Strains were ascribed to the same pulsotype if PFGE profile possessed >85% similarity.

4.3. Taxonomic Attribution

A strain for each biotype was taxonomically characterized by partial sequencing of the 16S rRNA gene sequencing, utilizing primers targeting the V1-V3 portion. Primer sequences and PCR conditions were set up according to Raimondi et al. [45]. The sequences, obtained from a service provider (Eurofins Genomics, Ebersberg, Germany), were compared to those in SILVA SSU database utilizing SINA Aligner v1.2.11 (https://www.arb-silva.de/aligner/).

In addition, the MALDI-TOF MS-based biotyping was carried using the MALDI Biotyper 3.1 system (Bruker Daltonics, Bremen, Germany). Sample preparation for MALDI-TOF MS was performed as previously described with minor modifications [46]. Briefly, colonies of fresh overnight culture were placed on a MALDI sample slide (Bruker Daltonics) and dried at room temperature. The sample was then overlaid with 1 µL of matrix solution (α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 2.5% trifluoroacetic acid) and dried at room temperature. A MALDI-TOF MS measurement was performed with a Bruker MicroFlex MALDI-TOF MS (Bruker Daltonics) using FlexControl software and a *Escherichia coli* DH5α protein extract (Bruker Daltonics) was placed on the calibration spot of the sample slide for external calibration. Spectra collected in the positive-ion mode within a mass range of 2000–20,000 Da were analyzed using a Bruker Biotyper (Bruker Daltonics) automation control and the Bruker Biotyper 3.1 software and library (a database with 5627 entries). High confidence species identification was accepted, if the log(score) was ≥2.00, low confidence species identification log(score) values (≥1.70 and <2.00) were accepted if the three best matches showed the same species name. Any results with log(score)<1.70 were considered as an unacceptable identification.

4.4. Profiling of Virulence Genes

All the isolates were screened by multiplex-PCR for the genes associated to 17 virulence factors: *allS*, *entB*, *fimH*, *fyuA*, *iroN*, *irp1*, *irp2*, *iutA*, *K2*, *kfu*, *kpn*, *magA*, *mrkD*, *rmpA*, *traT*, *ybtS*, and *ycfM*. Primer sequences and amplification conditions were set up according to El Fertas-Aissani et al. [19], Compain et al. [20], and Johnson et al. [21]. In order to assess the possibility to obtain the amplicon in the different NECEs species, a search with the primer-blast tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was performed for all the set of primers developed for *K. pneumoniae* and *E. coli*. The result of PCR amplification was reported only for the species in which the annealing and the possibility to yield an amplicon were predicted.

4.5. Biofilm and Phenotype Assays

Biofilm formation was quantified with crystal violet in the microtiter assay described in [47]. Two growth media were compared: Luria Bertani without salt (LBWS) and M9 (BD Difco, Sparks, MD, USA) containing 4 g/L glucose and 0.25 g/L yeast extract. Strains exhibiting a specific biofilm formation (i.e., the ratio between crystal violet absorbance at 570 nm and culture turbidity at 600 nm) > 1. The data herein reported are the means of three independent experiments, each carried out in triplicate.

The strains were screened for curli and cellulose production utilizing LBWS agar plates supplemented with the appropriate stain [48]. Red colonies in Congo red-supplemented plates were considered positive to curli. Colonies in calcofluor white-supplemented plates that emitted fluorescence due to UV exposure (315–400 nm) were considered positive to cellulose.

4.6. Solid Mating Conjugation Experiments

The strains were screened as recipients in conjugation experiments with the donor *E. coli* N4i pOX38:Cm (N4i: EcN immE7 Gmr; pOX38:Cm: Tra+ RepFIA+ Cmr) [5]. The donor and the recipient strains were cultured and put in contact onto LB plates under the conditions described in [5]. HCCA and HCCA with 20 μg/mL chloramphenicol were utilized to differentiate recipient, transconjugant, and donor colonies [5].

4.7. Antibiotic Susceptibility

The strains were tested for antimicrobial susceptibility with a Vitek2 semi-automated system (bioMerieux, Marcy-l'Étoile, France). Minimum inhibitory concentrations (MICs) were interpreted according to EUCAST (European Committee on Antimicrobial Susceptibility Testing—www.eucast.org) and susceptibility (S) or resistance (R) were defined based on the following thresholds (mg/L): amikacin, S < 8 and R > 16; amoxicillin/clavulanic acid, S < 8 and R > 8; cefotaxime, S < 8 and R > 2; ceftazidime, S < 8 and R > 8; ciprofloxacin, S < 0.5 and R > 1; gentamicin, S < 2 and R > 4; piperacillin + tazobactam, S < 8 and R > 16; trimethoprim/sulfamethoxazole, S < 40 and R > 80.

4.8. Biochemical Characterization

The strains were tested for distinctive enzymatic reactions and metabolic routes utilizing API 20 E test system (bioMerieux, France), according to the manufacturer's instructions.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/21/5/1847/s1.

Author Contributions: Conceptualization, G.A.G., F.B., M.R, and S.R.; data curation, S.R., F.C., and A.A.; investigation, A.A., F.C., L.R., E.M., S.R., and F.B.; methodology, S.R., L.R., F.B., M.R., and G.A.G.; project administration, M.R.; resources, M.R. and G.A.G.; supervision, M.R.; visualization, A.A.; writing—original draft, F.B., S.R., A.A. and M.R.; writing—review & editing, G.A.G. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

NECE	non-*E. coli* Enterobacterales
HCCA	HiCrome Coliform Agar
MALDI-TOF MS	Matrix-assisted laser desorption ionization–time of flight mass spectrometry
ERIC	Enterobacterial repetitive intergenic consensus
RAPD	Random Amplification of Polymorphic DNA
UPGMA	Unweighted pair group method with arithmetic means

References

1. Leimbach, A.; Hacker, J.; Dobrindt, U. *E. coli* as an all-rounder: The thin line between commensalism and pathogenicity. *Curr. Top. Microbiol. Immunol.* **2013**, *358*, 3–32. [CrossRef] [PubMed]
2. Manges, A.R.; Geum, H.M.; Guo, A.; Edens, T.J.; Fibke, C.D.; Pitout, J.D.D. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin. Microbiol. Rev.* **2019**, *32*. [CrossRef] [PubMed]
3. Clermont, O.; Christenson, J.K.; Denamur, E.; Gordon, D.M. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **2013**, *5*, 58–65. [CrossRef] [PubMed]
4. Croxen, M.A.; Law, R.J.; Scholz, R.; Keeney, K.M.; Wlodarska, M.; Finlay, B.B. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* **2013**, *26*, 822–880. [CrossRef]
5. Raimondi, S.; Righini, L.; Candeliere, F.; Musmeci, E.; Bonvicini, F.; Gentilomi, G.; Starčič Erjavec, M.; Amaretti, A.; Rossi, M. Antibiotic resistance, virulence factors, phenotyping, and genotyping of *E. coli* isolated from the feces of healthy subjects. *Microorganisms* **2019**, *7*, 251. [CrossRef]
6. Sarowska, J.; Futoma-Koloch, B.; Jama-Kmiecik, A.; Frej-Madrzak, M.; Ksiazczyk, M.; Bugla-Ploskonska, G.; Choroszy-Krol, I. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: Recent reports. *Gut Pathog.* **2019**, *11*, 10. [CrossRef]
7. Martin, R.M.; Bachman, M.A. Colonization, Infection, and the accessory genome of *Klebsiella pneumoniae*. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 4. [CrossRef]
8. Davin-Regli, A.; Pagès, J.M. *Enterobacter aerogenes* and *Enterobacter cloacae*; versatile bacterial pathogens confronting antibiotic treatment. *Front. Microbiol.* **2015**, *6*, 392. [CrossRef]
9. Liu, L.; Chen, D.; Liu, L.; Lan, R.; Hao, S.; Jin, W.; Sun, H.; Wang, Y.; Liang, Y.; Xu, J. Genetic diversity, multidrug resistance, and virulence of *Citrobacter freundii* from diarrheal patients and healthy individuals. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 233. [CrossRef]
10. Gajdács, M.; Urbán, E. Resistance Trends and epidemiology of citrobacter-enterobacter-serratia in urinary tract infections of inpatients and outpatients (RECESUTI): A 10-year survey. *Medicina (Kaunas)* **2019**, *55*, 285. [CrossRef]
11. Lee, C.R.; Lee, J.H.; Park, K.S.; Jeon, J.H.; Kim, Y.B.; Cha, C.J.; Jeong, B.C.; Lee, S.H. Antimicrobial resistance of hypervirulent *Klebsiella pneumoniae*: Epidemiology, hypervirulence-associated determinants, and resistance mechanisms. *Front. Cell. Infect. Microbiol.* **2017**, *7*, 483. [CrossRef] [PubMed]
12. Azevedo, P.A.A.; Furlan, J.P.R.; Oliveira-Silva, M.; Nakamura-Silva, R.; Gomes, C.N.; Costa, K.R.C.; Stehling, E.G.; Pitondo-Silva, A. Detection of virulence and β-lactamase encoding genes in *Enterobacter aerogenes* and *Enterobacter cloacae* clinical isolates from Brazil. *Braz. J. Microbiol.* **2018**, *49* (Suppl. 1), 224–228. [CrossRef] [PubMed]
13. Mokracka, J.; Koczura, R.; Kaznowski, A. Yersiniabactin and other siderophores produced by clinical isolates of *Enterobacter* spp. and *Citrobacter* spp. *FEMS Immunol. Med. Microbiol.* **2004**, *40*, 51–55. [CrossRef]
14. Yuan, C.; Yin, Z.; Wang, J.; Qian, C.; Wei, Y.; Zhang, S.; Jiang, L.; Liu, B. Comparative genomic analysis of *Citrobacter* and key genes essential for the pathogenicity of *Citrobacter koseri*. *Front. Microbiol.* **2019**, *10*, 2774. [CrossRef]
15. Cruz, A.; Xicohtencatl-Cortes, J.; González-Pedrajo, B.; Bobadilla, M.; Eslava, C.; Rosas, I. Virulence traits in *Cronobacter* species isolated from different sources. *Can. J. Microbiol.* **2011**, *57*, 735–744. [CrossRef]
16. Zogaj, X.; Bokranz, W.; Nimtz, M.; Römling, U. Production of cellulose and curli fimbriae by members of the family Enterobacteriaceae isolated from the human gastrointestinal tract. *Infect. Immun.* **2003**, *71*, 4151–4158. [CrossRef]

17. Tursi, S.A.; Tükel, Ç. Curli-containing enteric biofilms inside and out: Matrix composition, immune recognition, and disease implications. *Microbiol. Mol. Biol. Rev.* **2018**, *82*. [CrossRef]
18. Partridge, S.R. Resistance mechanisms in Enterobacteriaceae. *Pathology* **2015**, *47*, 276–284. [CrossRef]
19. El Fertas-Aissani, R.; Messai, Y.; Alouache, S.; Bakour, R. Virulence profiles and antibiotic susceptibility patterns of *Klebsiella pneumoniae* strains isolated from different clinical specimens. *Pathol. Biol. (Paris)* **2013**, *61*, 209–216. [CrossRef]
20. Compain, F.; Babosan, A.; Brisse, S.; Genel, N.; Audo, J.; Ailloud, F.; Kassis-Chikhani, N.; Arlet, G.; Decré, D. Multiplex PCR for detection of seven virulence factors and K1/K2 capsular serotypes of *Klebsiella pneumoniae*. *J. Clin. Microbiol.* **2014**, *52*, 4377–4380. [CrossRef]
21. Johnson, J.R.; Porter, S.; Johnston, B.; Kuskowski, M.A.; Spurbeck, R.R.; Mobley, H.L.; Williamson, D.A. Host characteristics and bacterial traits predict experimental virulence for *Escherichia coli* bloodstream isolates from patients with urosepsis. *Open Forum Infect. Dis.* **2015**, *2*. [CrossRef] [PubMed]
22. Davis, T.J.; Matsen, J.M. Prevalence and characteristics of *Klebsiella* species: Relation to association with a hospital environment. *J. Infect. Dis.* **1974**, *130*, 402–405. [CrossRef] [PubMed]
23. Lau, H.Y.; Huffnagle, G.B.; Moore, T.A. Host and microbiota factors that control *Klebsiella pneumoniae* mucosal colonization in mice. *Microbes Infect.* **2008**, *10*, 1283–1290. [CrossRef] [PubMed]
24. Magill, S.S.; Edwards, J.R.; Bamberg, W.; Beldavs, Z.G.; Dumyati, G.; Kainer, M.A.; Lynfield, R.; Maloney, M.; McAllister-Hollod, L.; Nadle, J.; et al. Multistate point-prevalence survey of health care-associated infections. *N. Engl. J. Med.* **2014**, *370*, 1198–1208. [CrossRef]
25. Lan, Y.; Zhou, M.; Jian, Z.; Yan, Q.; Wang, S.; Liu, W. Prevalence of pks gene cluster and characteristics of *Klebsiella pneumoniae*-induced bloodstream infections. *J. Clin. Lab. Anal.* **2019**, *33*, e22838. [CrossRef]
26. Lawlor, M.S.; O'connor, C.; Miller, V.L. Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect. Immun.* **2007**, *75*, 1463–1472. [CrossRef]
27. Schubert, S.; Cuenca, S.; Fischer, D.; Heesemann, J. High-pathogenicity island of *Yersinia pestis* in Enterobacteriaceae isolated from blood cultures and urine samples: Prevalence and functional expression. *J. Infect. Dis.* **2000**, *182*, 1268–1271. [CrossRef]
28. Schubert, S.; Rakin, A.; Heesemann, J. The *Yersinia* high-pathogenicity island (HPI): Evolutionary and functional aspects. *Int. J. Med. Microbiol.* **2004**, *294*, 83–94. [CrossRef]
29. Raffatellu, M. Learning from bacterial competition in the host to develop antimicrobials. *Nat. Med.* **2018**, *24*, 1097–1103. [CrossRef]
30. Fang, C.T.; Chuang, Y.P.; Shun, C.T.; Chang, S.C.; Wang, J.T. A novel virulence gene in *Klebsiella pneumoniae* strains causing primary liver abscess and septic metastatic complications. *J. Exp. Med.* **2004**, *199*, 697–705. [CrossRef]
31. Yu, W.L.; Ko, W.C.; Cheng, K.C.; Lee, H.C.; Ke, D.S.; Lee, C.C.; Fung, C.P.; Chuang, Y.C. Association between *rmpA* and *magA* genes and clinical syndromes caused by *Klebsiella pneumoniae* in Taiwan. *Clin. Infect. Dis.* **2006**, *42*, 1351–1358. [CrossRef] [PubMed]
32. Johansson, M.E.; Sjovall, H.; Hansson, G.C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **2013**, *10*, 352–361. [CrossRef] [PubMed]
33. Flores-Mireles, A.L.; Walker, J.N.; Caparon, M.; Hultgren, S.J. Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **2015**, *13*, 269–284. [CrossRef] [PubMed]
34. Hobley, L.; Harkins, C.; MacPhee, C.E.; Stanley-Wall, N.R. Giving structure to the biofilm matrix: An overview of individual strategies and emerging common themes. *FEMS Microbiol. Rev.* **2015**, *39*, 649–669. [CrossRef] [PubMed]
35. Martin, R.M.; Cao, J.; Brisse, S.; Passet, V.; Wu, W.; Zhao, L.; Malani, P.N.; Rao, K.; Bachman, M.A. Molecular epidemiology of colonizing and infecting isolates of *Klebsiella pneumoniae*. *mSphere* **2016**, *1*. [CrossRef] [PubMed]
36. Gorrie, C.L.; Mirceta, M.; Wick, R.R.; Edwards, D.J.; Thomson, N.R.; Strugnell, R.A.; Pratt, N.F.; Garlick, J.S.; Watson, K.M.; Pilcher, D.V.; et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clin. Infect. Dis.* **2017**, *65*, 208–215. [CrossRef] [PubMed]
37. Gajdács, M.; Bátori, Z.; Ábrók, M.; Lázár, A.; Burián, K. Characterization of resistance in gram-negative urinary isolates using existing and novel indicators of clinical relevance: A 10-year data analysis. *Life (Basel)* **2020**, *10*, 16. [CrossRef]

38. Carrie, C.; Walewski, V.; Levy, C.; Alexandre, C.; Baleine, J.; Charreton, C.; Coche-Monier, B.; Caeymaex, L.; Lageix, F.; Lorrot, M.; et al. *Klebsiella pneumoniae* and *Klebsiella oxytoca* meningitis in infants. Epidemiological and clinical features. *Arch. Pediatr.* **2019**, *26*, 12–15. [CrossRef]
39. Pötgens, S.A.; Brossel, H.; Sboarina, M.; Catry, E.; Cani, P.D.; Neyrinck, A.M.; Delzenne, N.M.; Bindels, L.B. Klebsiella oxytoca expands in cancer cachexia and acts as a gut pathobiont contributing to intestinal dysfunction. *Sci. Rep.* **2018**, *8*, 12321. [CrossRef]
40. Gajdács, M.; Ábrók, M.; Lázár, A.; Burián, K. Comparative epidemiology and resistance trends of common urinary pathogens in a tertiary-care hospital: A 10-year surveillance study. *Medicina (Kaunas)* **2019**, *55*, 356. [CrossRef]
41. MacPherson, C.W.; Mathieu, O.; Tremblay, J.; Champagne, J.; Nantel, A.; Girard, S.A.; Tompkins, T.A. Gut bacterial microbiota and its resistome rapidly recover to basal state levels after short-term amoxicillin-clavulanic acid treatment in healthy adults. *Sci. Rep.* **2018**, *8*, 11192. [CrossRef] [PubMed]
42. Leclercq, R.; Cantón, R.; Brown, D.F.; Giske, C.G.; Heisig, P.; MacGowan, A.P.; Mouton, J.W.; Nordmann, P.; Rodloff, A.C.; Rossolini, G.M.; et al. EUCAST expert rules in antimicrobial susceptibility testing. *Clin. Microbiol. Infect.* **2013**, *19*, 141–160. [CrossRef] [PubMed]
43. Wilson, L.A.; Sharp, P.M. Enterobacterial repetitive intergenic consensus (ERIC) sequences in *Escherichia coli*: Evolution and implications for ERIC-PCR. *Mol. Biol. Evol.* **2006**, *23*, 1156–1168. [CrossRef] [PubMed]
44. Quartieri, A.; Simone, M.; Gozzoli, C.; Popovic, M.; D'Auria, G.; Amaretti, A.; Raimondi, S.; Rossi, M. Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe* **2016**, *38*, 130–137. [CrossRef]
45. Raimondi, S.; Luciani, R.; Sirangelo, T.M.; Amaretti, A.; Leonardi, A.; Ulrici, A.; Foca, G.; D'Auria, G.; Moya, A.; Zuliani, V.; et al. Microbiota of sliced cooked ham packaged in modified atmosphere throughout the shelf life: Microbiota of sliced cooked ham in MAP. *Int. J. Food Microbiol.* **2019**, *16*, 200–208. [CrossRef]
46. Mencacci, A.; Monari, C.; Leli, C.; Merlini, L.; De Carolis, E.; Vella, A.; Cacioni, M.; Buzi, S.; Nardelli, E.; Bistoni, F.; et al. Typing of nosocomial outbreaks of Acinetobacter baumannii by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* **2013**, *51*, 603–606. [CrossRef]
47. Sicard, J.F.; Vogeleer, P.; Le Bihan, G.; Rodriguez Olivera, Y.; Beaudry, F.; Jacques, M.; Harel, J. N-Acetyl-glucosamine influences the biofilm formation of *Escherichia coli*. *Gut Pathog.* **2018**, *22*, 10–26. [CrossRef]
48. Vogeleer, P.; Tremblay, Y.D.N.; Jubelin, G.; Jacques, M.; Harel, J. Biofilm-forming abilities of shiga toxin-producing *Escherichia coli* isolates associated with human infections. *Appl. Environ. Microbiol.* **2015**, *28*, 1448–1458. [CrossRef]

microorganisms

MDPI

Article

Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of *E. coli* Isolated from the Feces of Healthy Subjects

Stefano Raimondi [1], Lucia Righini [1], Francesco Candeliere [1], Eliana Musmeci [1], Francesca Bonvicini [2], Giovanna Gentilomi [2,3], Marjanca Starčič Erjavec [4], Alberto Amaretti [1] and Maddalena Rossi [1,*]

1 Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 103, 41125, Modena, Italy
2 Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, Via Massarenti 9, 40138 Bologna, Italy
3 Unit of Microbiology, Alma Mater Studiorum-University of Bologna, S. Orsola-Malpighi Hospital, Via Massarenti 9, 40138 Bologna, Italy
4 Department of Biology, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia
* Correspondence: maddalena.rossi@unimore.it; Tel.: +39-059-205-8589

Received: 19 July 2019; Accepted: 7 August 2019; Published: 10 August 2019

Abstract: *Escherichia coli* may innocuously colonize the intestine of healthy subjects or may instigate infections in the gut or in other districts. This study investigated intestinal *E. coli* isolated from 20 healthy adults. Fifty-one strains were genotyped by molecular fingerprinting and analyzed for genetic and phenotypic traits, encompassing the profile of antibiotic resistance, biofilm production, the presence of surface structures (such as curli and cellulose), and their performance as recipients in conjugation experiments. A phylogroup classification and analysis of 34 virulence determinants, together with genes associated to the *pks* island (polyketide-peptide genotoxin colibactin) and conjugative elements, was performed. Most of the strains belonged to the phylogroups B1 and B2. The different phylogroups were separated in a principal coordinate space, considering both genetic and functional features, but not considering pulsed-field gel electrophoresis. Within the B2 and F strains, 12 shared the pattern of virulence genes with potential uropathogens. Forty-nine strains were sensitive to all the tested antibiotics. Strains similar to the potential pathogens innocuously inhabited the gut of healthy subjects. However, they may potentially act as etiologic agents of extra-intestinal infections and are susceptible to a wide range of antibiotics. Nevertheless, there is still the possibility to control infections with antibiotic therapy.

Keywords: *Escherichia coli*; typing; gut microbiota; PFGE (pulsed-field gel electrophoresis); virulence; antibiotic resistance; conjugation; curli; co-occurrence

1. Introduction

E. coli are permanent colonizers of the human microbiota that may persist as gut commensals without inducing any intestinal or extraintestinal infections. On the other hand, certain *E. coli* strains may instigate infections not only in the gut but also in other districts, such as those caused by the extraintestinal pathogenic *E. coli* (ExPEC) [1]. The commensalism or virulence of *E. coli* derives from a complex balance between the whole status of the host and the presence and expression of virulence determinants. Commonly, ExPEC strains reside as harmless commensals in the gut of healthy subjects. However, these strains can cause an infection in compromised patients, in case they reach a usually sterile extraintestinal site, such as the urinary tract [2]. The gut microbiota, therefore, act as a powerful

reservoir of ExPEC strains potentially responsible for infections, with pathogenic and commensal *E. coli* generally differing in terms of their phylogenetic backgrounds and virulence attributes [3].

The strains that innocuously colonize the intestine of healthy subjects may differ from those that are prone to cause diseases, especially those in possession of accessory traits that confer fitness and competitiveness and shape a specific relationship with the host. The virulence capability of *E. coli* depends on adhesion, biofilm formation, attachment, acquirement of nutrients, competition with other bacteria, toxin production, and avoidance or subversion of host defense mechanisms [4].

Adherence to the epithelial cells is mediated by surface structures or molecules, like fimbrial and afimbrial adhesins, curli, and outer membrane proteins encoded by the *pap* cluster and other genes (*afa/draBC*, *fimH*, *focG*, *gafD*, *hra*, *iha*, *sfa/focDE*, *sfaS*, and *yfcV*) [5]. Furthermore, *E. coli* exploits several mechanisms of iron uptake that are associated with siderophores and other binding proteins encoded by *chuA*, *fyuA*, *ire*, *iroN*, and *iutA* [5,6]. Indeed, like other pathogens, needing iron for metabolism, *E. coli* must face the host's response to infection, which involves a reduction in the amount of iron available via a decrease of intestinal iron absorption, the synthesis of additional iron proteins, and shifting iron from the plasma pool into intracellular storage. *E. coli* virulence is also enhanced by the production of toxins (e.g., cytotoxic necrotizing factor 1, autotransporter toxins, and alpha hemolysin) that target the cell's skeleton, metabolism, or cytoplasmic membrane [5,7].

Genetic exchange increases the success of commensals in invasion, intracellular survival, and spread, providing them with increased fitness and versatility. Hence, the boundary between commensals and pathogens is made fainter by horizontal gene transfer. Mobile genetic elements, such as transposons, plasmids, and insertion sequences, contribute to the plasticity of the *E. coli* genome, resulting in an extremely large pangenome of more than 16,000 genes [8]. Moreover, horizontal gene transfer favors the diffusion of antimicrobial resistance (AR) among both *E. coli* and other commensals, thereby enlarging the spectrum of resistance and promoting epidemiological success, with a bloom in worldwide public health concern associated with the misuse of antibiotics. In particular, conjugation is one of the most important ways for genes to exchange in prokaryotes, leading to genetic variation within a species and the acquisition of new traits. This process requires complex circuits that regulate the transcription of conjugation genes, the assembly of conjugative pili, the formation of the mating pore connecting donor and recipient cells, and the enzymatic processing of plasmid DNA to be transferred [9].

Research interest has been manly focused on characterizing virulent clinical *E. coli.* [10], whereas strains isolated from healthy subjects have been mainly investigated in comparative studies with patients affected by specific diseases [11,12]. A few studies specifically describing the intestinal *E. coli* of healthy subjects mainly focused on antibiotic resistance, without performing a thorough genetic and phenotypic analysis of the strains [13–15]. The present study aimed to deeply characterize the population of *E. coli* isolated from the feces of 20 healthy adults in order to determine whether the relationship between PFGE genotyping, phylogroups, genetic determinants, and functional features can be established. A set of 51 strains was analyzed and characterized according to a generally accepted phylotype classification of *E. coli* that includes seven phylogroups of *E. coli sensu stricto* (A, B1, B2, C, D, E, and F) [16,17]. Thirty-four virulence determinants were searched, together with the *pks* island claimed as a concurrent cause in the development of human colorectal cancer [18], the gene *traD*, as an indicator of the presence of conjugative elements [19], the genes encoding E7 colicin (*colE7*), and the immunity protein (*immE7*) [20]. Phenotypical characterization of the isolates included the AR profile, the capability to form biofilms, and the presence of the surface structures involved in adherence.

2. Results

2.1. Quantitation, Phyloptyping, and Genotyping

E. coli were quantified in the feces of 20 healthy subjects by enumeration onto selective plates and through qPCR. These techniques yielded consistent values ($p = 0.34$), ranging from 3.4×10^5 to

9.3×10^7 cfu, or cells per g of feces (Figure 1; Table S1). The ratio between the *E. coli* and total bacteria ranged from 5.6×10^{-6} to 1.9×10^{-3}, with a mean of 2.8×10^{-5}.

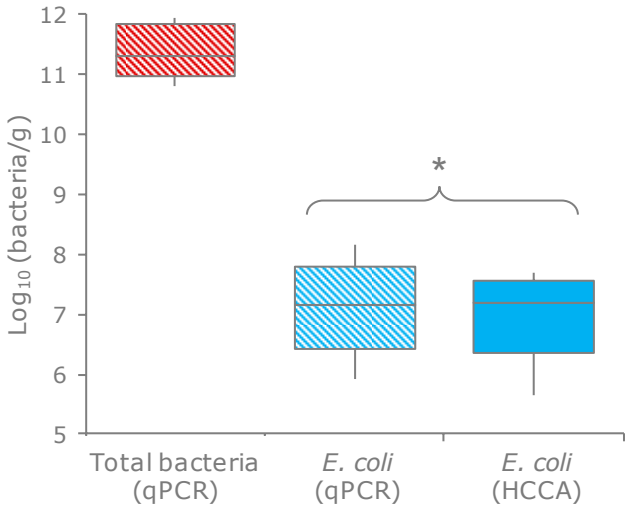


Figure 1. Counts of *E. coli* and total bacteria in the feces of 20 subjects. Boxes indicate the median and 25th and 75th percentiles; whiskers indicate the 10th and 90th percentiles. * indicates the means that did not significantly differ ($p > 0.05$, paired samples t-test). HCCA, HiCrome Coliform Agar (the selective medium for *E. coli*).

Ninety-six colonies per sample were randomly selected and clustered based on their ERIC-PCR and RAPD-PCR profiles (Table 1). When profiles were similar across different hosts, an isolate was picked from each of the subjects. A collection of 51 *E. coli* strains was obtained, also encompassing 10 β-glucuronidase-negative clones isolated from the same samples during a parallel study that aimed to characterize Enterobacteriaceae other than *E. coli* (which were ascribed to this species by a MALDI Biotyper matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) system (Bruker Daltonik GmbH, Bremen, Germany) [21]). Within each sample, one to five biotypes were recognized, with most of the samples (14 out of 20) characterized by a dominant biotype (>80%). According to the XbaI PFGE analysis, the 51 strains were ascribed to 44 different pulsotypes (Figure 2).

The strains were assigned to a phylogroup according to Clermont et al. [17] (Table 1). The most prevalent phylogroup was B2 (14 strains), followed by B1, F, A, D, E, and C (12, 7, 7, 5, 4, and 2 strains, respectively). The strains of phylogroup B2 presented the highest absolute charge and were the most abundant within the *E. coli* population, accounting for the all colonies recovered from several samples. B1 phylogroups were detected at lower concentrations and represented a minor subpopulation (Table 1).

Table 1. *E. coli* isolates from 20 healthy subjects. The strains were genotyped by ERIC-PCR, RAPD-PCR, and PFGE analysis. The relative abundance represents the ratio of each strain within the set of *E. coli* clones analyzed. The charge of the viable cells has been calculated by multiplying the total *E. coli* count and the relative amount. Phlylotyping was done according to Clermont et al. 2013 [17]. The presence of β-glucuronidase was confirmed by growth of blue colonies on the HCCA medium. The performance of isolates as recipients in conjugation experiments was assessed only in β-glucuronidase negative strains. Biofilm formation was tested on LBWS and M9glu media. Curli and cellulose production was determined.

E. coli Strain	Genotyping and Phyloptyping						Phenotype Assays					
	ERIC-PCR			RAPD-PCR Profile	*Xba*I-PFGE Profile	Phylo-Group	β-glu	Conju-Gation	Biofilm		Curli	Cellulose
	Profile	Relative Abundance	Log (cfu/g)						LBWS	M9		
01.01	E01	100%	7.56	R01	P01	B1	+		−	−	−	−
02.03	E02	8%	5.22	R02	P02	B2	+		−	−	−	−
02.16	E03	92%	6.28	R03	P03	F	+		+	−	+	+
03.25	E04	99%	7.52	R04	P04	B2	+		+	−	−	−
03.73	E05	1%	5.53	R03	P05	A	+		+	+	−	−
04.05	E06	5%	4.22	R05	P06	A	+		+	−	+	−
04.06	E07	70%	5.37	R06	P07	B2	+		−	−	−	−
04.10	E08	25%	4.92	R07	P08	E	+		−	−	+	+
05.20	E07	96%	7.78	R02	P09	B2	+		−	−	−	+
05.47	E09	4%	6.40	R08	P10	B1	−	+	−	−	−	−
06.16	E07	100%	7.28	R07	P11	B2	+		−	−	−	−
07.09	E10	100%	6.45	R04	P12	D	+		+	−	+	−
08.01	E11	18%	4.91	R09	P12	D	+		+	−	+	−
08.02	E12	5%	4.37	R10	P13	A	+		−	+	−	−
08.10	E13	5%	4.37	R11	P14	A	+		+	+	+	−
08.27	E14	36%	5.20	R11	P15	B1	−	+	−	+	−	−
08.57	E15	34%	5.18	R12	P16	F	−	+	−	−	−	−
08.109	E16	2%	3.86	R13	P17	F	−	+	−	−	−	−
09.13	E01	92%	7.09	R14	P18	C	+		+	−	+	+
09.21	E17	4%	5.73	R12	P19	F	+		−	−	−	−
09.42	E07	4%	5.73	R12	P09	B2	+		−	−	−	−
10.17	E04	100%	7.62	R12	P04	B2	+		+	−	−	−

Table 1. *Cont.*

E. coli Strain	Genotyping and Phyloptyping						Phenotype Assays				
	ERIC-PCR			RAPD-PCR Profile	*Xba*I-PFGE Profile	Phylo-Group	β-glu	Conju-Gation	Biofilm		Curli	Cellulose
	Profile	Relative Abundance	Log (cfu/g)						LBWS	M9		
11.02	E18	50%	7.67	R15	P20	A	+		−	−	−	−
11.04	E19	9%	6.91	R10	P21	A	+		−	−	−	−
11.06	E07	37%	7.54	R16	P22	B2	+		−	−	−	−
11.16	E20	4%	6.57	R17	P23	B2	+		−	−	−	−
12.04	E13	71%	7.43	R18	P24	B2	+		−	−	+	−
12.09	E11	29%	7.04	R18	P25	D	+		−	−	−	−
13.01	E01	3%	4.98	R19	P26	B1	+		−	−	+	−
13.03	E11	1%	4.47	R11	P27	D	+		−	−	+	+
13.20	E21	97%	6.46	R13	P28	F	−	−	−	−	−	−
14.03	E22	100%	6.85	R13	P29	E	+		−	−	+	−
15.01	E07	9%	6.45	R13	P30	B2	+		−	−	−	+
15.02	E01	87%	7.45	R11	P31	B1	+		+	−	+	+
15.13	E11	4%	6.15	R20	P32	D	+		−	−	+	−
16.04	E21	1%	5.74	R21	P33	F	+		−	−	−	−
16.15	E01	1%	5.74	R22	P34	B1	+		+	−	+	−
16.27	E21	98%	7.73	R23	P33	F	−	−	−	−	−	−
17.10	E23	87%	7.45	R23	P35	A	+		−	−	−	−
17.18	E07	13%	6.63	R24	P36	B2	+		+	−	−	−
18.03	E24	39%	6.25	R25	P37	B1	+		−	+	−	−
18.09	E13	41%	6.28	R25	P24	B2	+		−	−	+	−
18.18	E25	19%	5.93	R13	P38	E	+		−	−	−	+
18.50	E26	1%	4.70	R21	P39	B1	−	+	−	−	−	−
19.02	E07	100%	7.19	R14	P40	B2	+		+	−	−	−
20.04	E27	12%	6.39	R26	P41	B1	+		−	−	+	−
20.17	E28	5%	6.01	R12	P42	C	+		−	−	+	−
20.27	E29	53%	7.03	R27	P43	E	+		+	−	+	+
20.50	E30	22%	6.65	R04	P44	B1	−	+	−	−	−	−
20.51	E31	4%	5.91	R06	P44	B1	−	+	−	−	−	−
20.71	E09	4%	5.91	R06	P44	B1	−	+	−	−	−	−

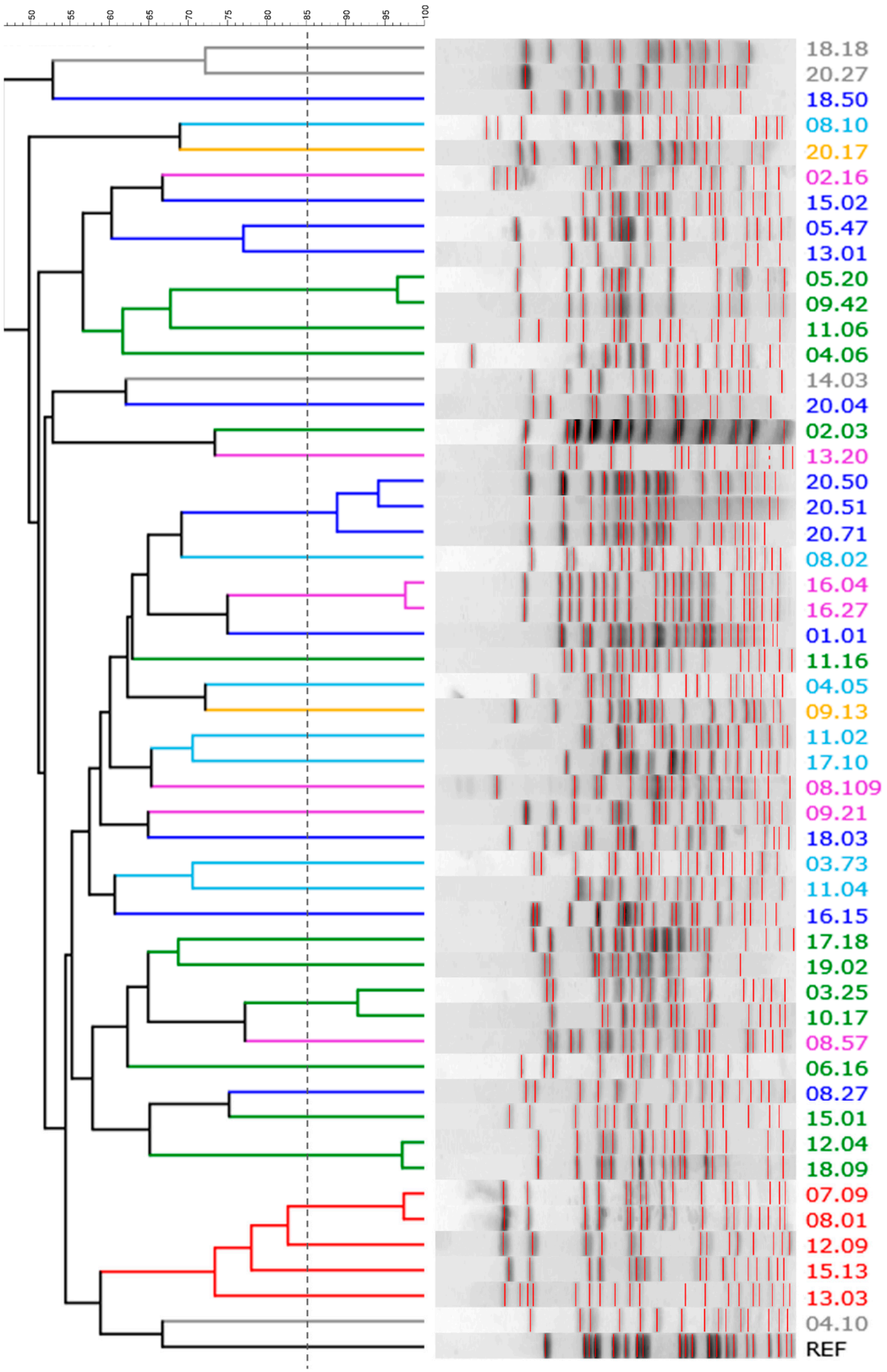


Figure 2. The *Xba*I-PFGE pattern of *E. coli* strains: a UPGMA dendrogram derived from Dice's coefficients, calculated based on their band profiles. Patterns with similarities >85% (dashed line) were ascribed to the same pulsotype. Strains are colored based on their phylogroup: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; F, pink.

2.2. Virulence Factors

Thirty-four genetic determinants encoding virulence factors that potentially enhance the risk of infections were screened by PCR. Most of the strains were positive for *fimH*, *fyuA*, *ompT*, *traT*, *chuA*, and *kpsMTII* (47, 36, 35, 35, 32, and 27 strains, respectively). On the other hand, *afa*/draBC, *cnf1*, *focG*, *hlyD*, *ibeA*, *rfc*, *sfa*/*focDE*, and *sfaS* recurred rarely (≤ 5 strains), while *cdtB*, *gafD*, *pic*, and *vat* were absent in all the strains (Figure 3). Based on virulence factors, all the strains belonging to the phylotypes B2, D (with an exception), E, and F clustered together (Figure 4). Strains ascribed to the phylotypes C were grouped in another cluster that also included the majority of B1 and most of the A isolates. A single B1 strain lay outside both clusters.

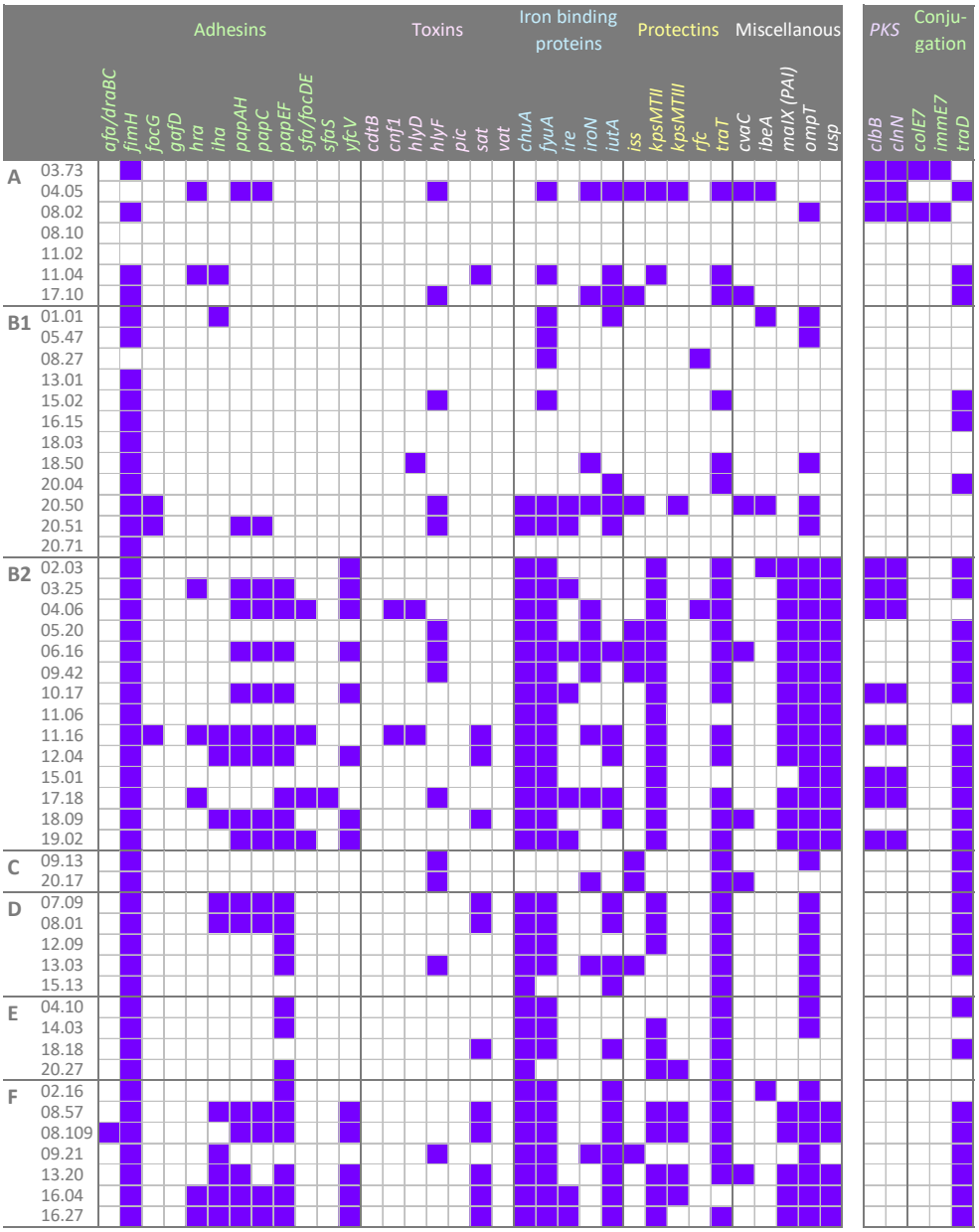


Figure 3. The pattern of genetic determinants in the 51 *E. coli* strains, ordered on the basis of their phylotypes. Virulence genes coding for adhesins, toxins, iron binding proteins, and protectins were screened. The presence of the *pks* pathogenicity island and other genes involved in the conjugation machinery were searched. Purple squares, presence; white squares, absence.

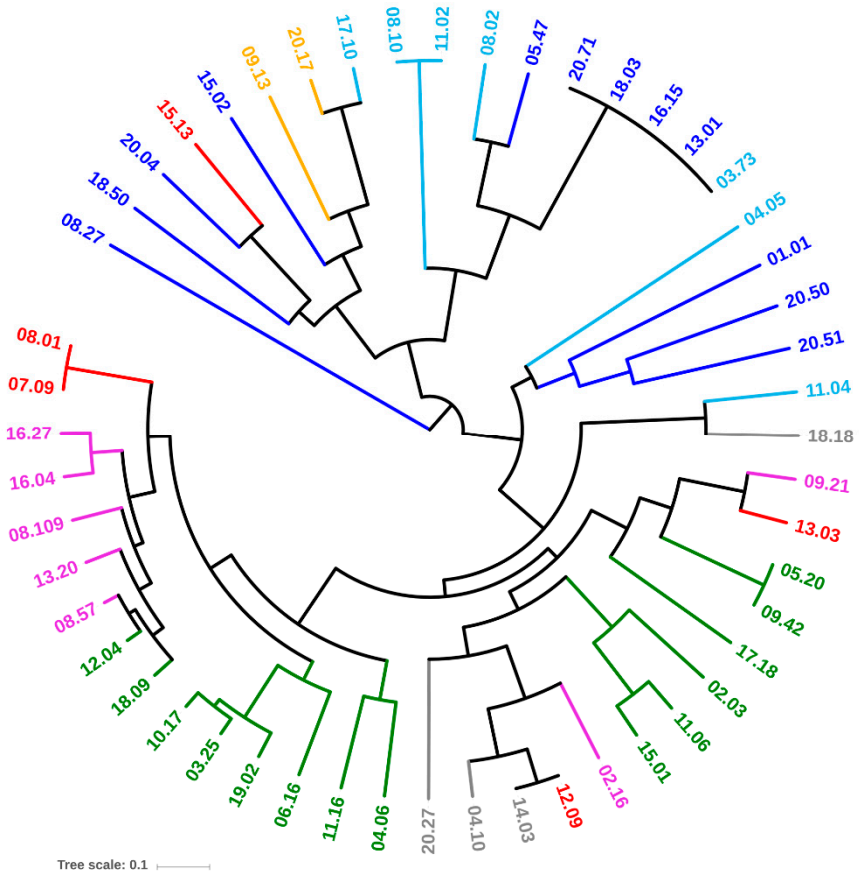


Figure 4. UPGMA dendrogram of *E. coli* strains, computed from Dice's distance matrix of the virulence determinants. Strains are colored based on their phylogroups: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; F, pink.

2.3. Other Genetic Determinants Affecting Fitness and Pathogenicity

E. coli strains were screened for the presence of the *pks* pathogenicity island (Figure 3). The genes *clbB* and *clbN*, utilized as markers, were found in 11 out of 51 strains, 3 ascribed to phylotype A and 8 to B2.

The presence of *traD* was investigated to establish the potential in conjugative DNA exchange. Most isolates were positive to *traD* (32/51), including the the majority of pks-positive strains (Figure 3). *traD* was distributed differently among the phylogroups: most B2, C, D, and F strains harbored the gene, while only a minority within A and B1 did the same. When the same subject harbored more biotypes, some isolates bore *traD* while others did not.

The genes *colE7* and *immE7*, respectively encoding the bactericidal nuclease Colicin E7 and the corresponding immunity protein, were searched. Two strains harbored both the genes.

2.4. Antibiotic Susceptibility

Phenotypic antibiotic susceptibility tests were carried out on the 51 *E. coli* strains (Table S1). Nearly all the isolates were extensively sensitive to the whole set of tested antibiotics (49/51). Only 2 strains recovered from the same fecal sample, belonging to the phylogroups B2 (17.18) and A (17.10), were resistant to gentamycin. *E. coli* 17.10 was also resistant to ciprofloxacin.

2.5. Conjugation

The 10 β-glucuronidase-negative *E. coli* strains, belonging to phylotypes B1 (6) and F (4), were challenged as conjugation recipients for receiving the pOX38: Cm plasmid from *E. coli* N4i (Table 1).

The plasmid transfer to most of the strains succeeded with high efficiency: all the B1 strains and 2 out of the the 4 F strains acquired the plasmid.

2.6. Production of Curli and Cellulose and Biofilm Formation

The production of curli was assessed by observing the colonies grown in CR-containing LB agar plates. B2 and F strains presented the lowest ability to bind CR dye (2/14 and 1/7, respectively), whereas C, D, and E strains seemed to be inclined to produce curli (Table 1). The occurrence of cellulose-like extracellular components was analyzed by growing the isolates on LB plates supplemented with CF. The ability of the strains to bind CF was less frequent (9/51) than their ability to bind CR.

E. coli strains were assayed for biofilm formation in LBWS and M9glu. Most of the strains did not form a biofilm (Figure 5). Formation occurred more frequently in LBWS than in M9glu (13/51 and 5/51 strains, respectively). Only two strains, both ascribed to phylotype A (03.73 and 08.10), produced a biofilm in LBWS and M9glu. For five other strains, the score of the biofilm formation ranged between 0.6 and 0.8. All but one of these strains presented a very high standard deviation (>0.5; Figure 5), suggesting that biofilm formation was sensitive to environmental conditions that were hardly controlled by the operator.

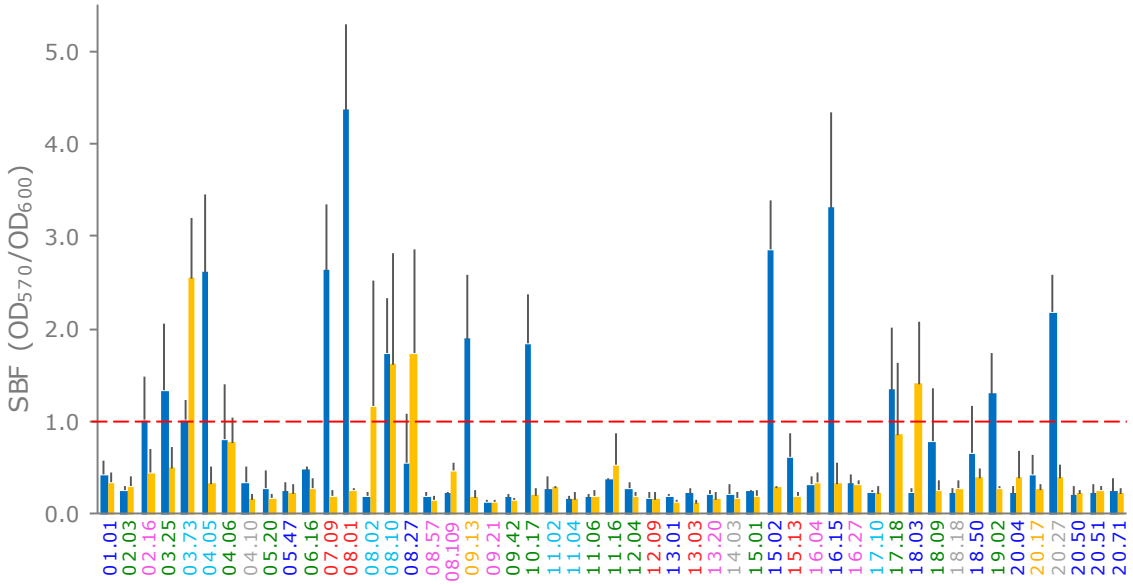


Figure 5. formation of *E. coli* strains on LBWS (blue bars) and M9glu (yellow bars). The specific biofilm formation index (SBF) is calculated as the ratio between the crystal violet absorbance at 570 nm and the culture turbidity at 600 nm, setting a threshold of 1 for biofilm producers (red dashed line). The reported data are means ± standard deviations of at least three independent experiments, each carried out in triplicate. The names of the strains are colored based on their phylogroups: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; and F, pink (hereafter referred to as LBWS and M9glu, respectively).

2.7. Co-Occurrence and Principal Coordinate Analysis (PCoA) Analysis of Genetic and Functional Features

A co-occurrence analysis was applied to the whole set of genetic and function observations (Figure 6). The presence of pks islands and several virulence determinants was significantly associated ($p < 0.05$ or < 0.01) to certain genotypic classifications (such as the ERIC profile, the pulsotype, and the phylogroup). In particular, the genes encoding some adhesins (*iha*, *papAH*, *papC*, *papF*, and *sat*), iron binding proteins (*chuA*, *fyuA*, and *iutA*), protectins (*kpsMTII* and *traT*), and *malX*, *ompT*, *usp*, and *traD* were significantly associated to phylogroups and presented a significantly positive tendency to co-occur within the same strains ($p < 0.05$ or 0.01).

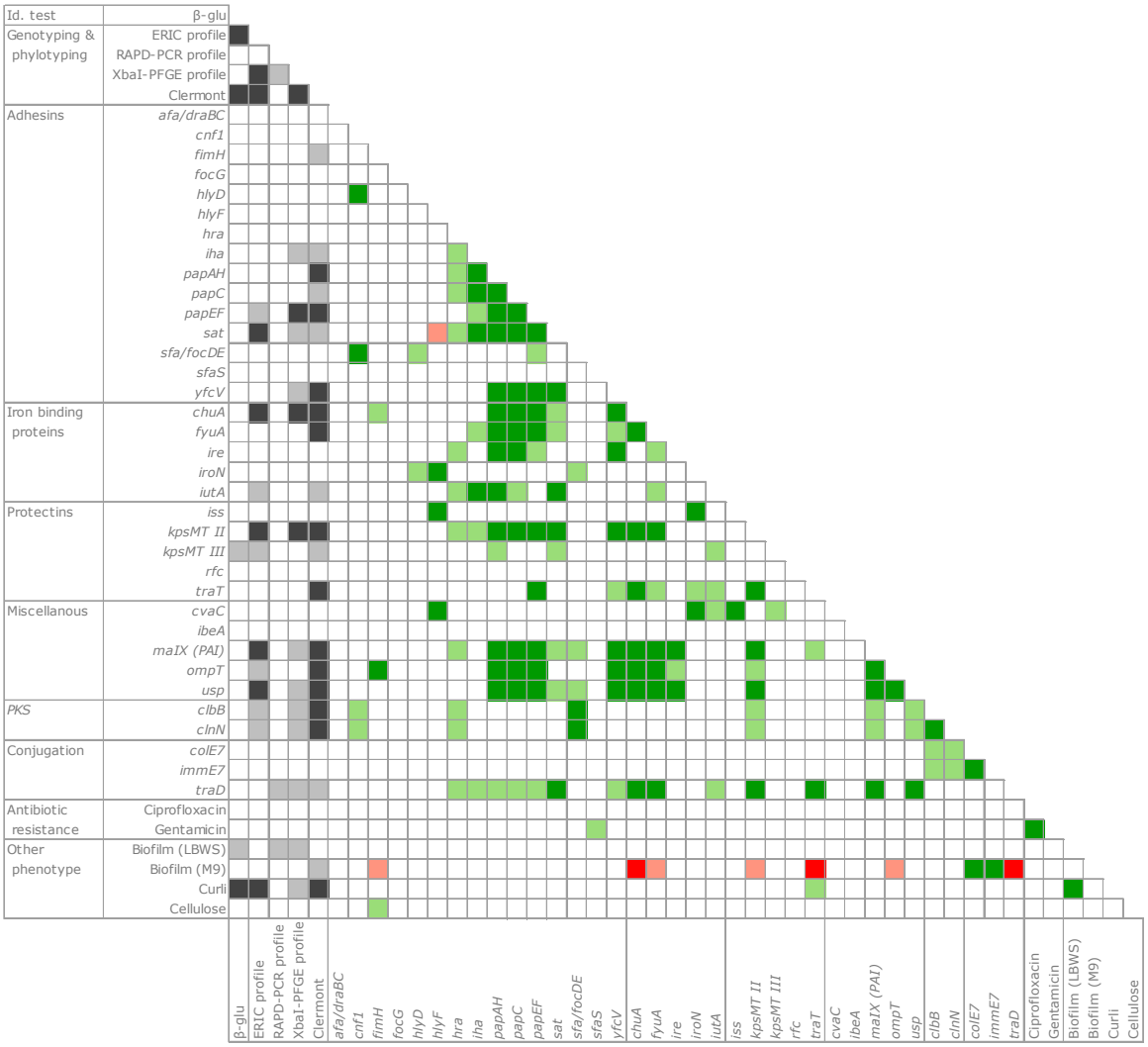


Figure 6. Genetic determinants and phenotypic properties evaluated by Cramér's V metrics. Association with genotype/phylotype clusters, grey; positive co-occurrence, green; negative co-occurrence, red. Dark and light shades indicate the levels of statistical significance ($p < 0.01$ and $p < 0.05$, respectively).

With regards to the phenotypic features involved in bacterial adhesion, the production curli were significantly associated with the phylogroup ($p < 0.01$). Although biofilm formation was not frequent, curli presented a significant co-occurrence with the biofilm in the LBWS medium.

A Principal Coordinate Analysis (PCoA) was applied to assist in the identification of clusters, taking into account both their genetic and functional features (Figure 7). The first two coordinates were the most informative, accounting for approximately 28.9% and 10.2% of the total observed variances. In the PCoA plot, PCo1 separated B2 and F strains from B1, C, and E ones (negative and positive PCo1, respectively). F strains were grouped in two clusters. With respect to PCo2, most of the A, C, D, and E strains lie in the positive quadrants.

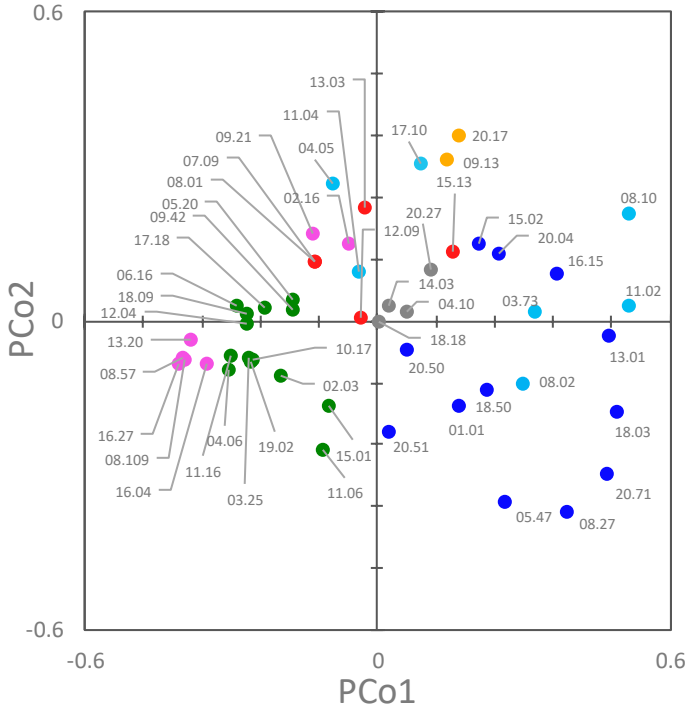


Figure 7. 2D Principal Coordinate Analysis (PCoA) visualization of *E. coli* strains, computed from the Dice's distance matrix of genetic and phenotypic determinants. Strains are colored based on their phylogroup: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; F, pink.

3. Discussion

Fifty-one *E. coli* strains were isolated from the feces of 20 healthy adults and genotypically and phenotypically characterized. Most of the strains belonged to phylogroups B1 (12) and B2 (14). Several B2 strains presented a greater attitude to colonization, with higher charges and an overall dominance in the strains belonging to other phylotypes. This result is consistent with the fact that B2 strains showed an enhanced ability to persist in the intestinal microbiota, as demonstrated in infants [22,23]. B1 strains, on the contrary, showed lower performance in colonization, generally representing minor *E. coli* subpopulations within each sample.

Phylogroup classification and PFGE pulsotypes were consistent, but the strains belonging to the same phylogroups were spread over different clades of the PFGE dendrogram, with only D isolates clustered together. PFGE analysis targeted the whole genome, most of which is composed of accessory genes that can be acquired by horizontal gene transfer. The PFGE fingerprint of *E. coli* results from the digestion of a genome, the size of which, including plasmids and prophages, ranges approximately from 4.6 to 5.9 Mbp, encoding from 4200 to 5500 genes [3]. This huge difference is mainly due to the genes associated with bacteriophage elements and involved in virulence or resistance to antimicrobials. The phylotypes recognized in this study were shuffled in PFGE, likely because of the genome evolution, which was mainly due to the recombination and horizontal transfer that broke the old phylotype relationships, thereby emphasizing more recent changes. On the other hand, MLST typing, the gold-standard approach for *E. coli* classification, allows a phylotyping-consistent analysis, because MLST is based on the conserved nature of the housekeeping genes of the core genome [24]. However, when both the genetic and functional features of each strain were elaborated together in this study (Figure 7), the different phylogroups clustered separately in a PCoA space.

The majority of the strains were sensitive to all the tested antibiotics, including amoxicillin plus clavulanic acid and cephalosporins, the most diffused antibiotics for the first management of infections. Two sole exceptions, resistant to gentamycin and ciprofloxacin, were isolated from the same subject. The absence of the AR phenotype does not exclude the presence of AR genes, which may be expressed

in vivo or can be involved in the diffusion and spread of AR genes [25]. However, the rapid emergence of resistant bacteria occurring worldwide attributed to the overuse and misuse of antibiotics, healthy subjects harbor an endogenous *E. coli* population still sensitive to a wide range of antibiotics, contrary to the pathogenic strains isolated from different clinical specimens [26–28]. The scarce relevance of commensal *E. coli* as an AR carrier disagrees with other studies [29–32] but is consistent with a recent analysis of metagenomes that investigated AR genes in the gut microbiota of healthy people. As a whole, the results show that commensal *E. coli* may not pose a threat by itself in terms of AR, though major differences are registered among countries [33,34].

The great variability of pathogenicity-associated features suggests that the genetic determinants of virulence had a role in shaping the genome of the isolates. A significant co-occurrence was observed in a set of B2 and F strains for *pap* genes (*papAH*, *papC*, *papEF*), encoding proteins of the fimbria favoring the ascension of the urinary tract and promoting colonization and infection, *sat*, which is associated with a cytopathic secreted autotransported toxin exerting effect, *kpsMTII*, the marker of K2 capsular polysaccharides, which plays a main role in pathogenesis, *chuA* (heme receptor), *yfcV* (Yfc fimbria), along with *fyuA*, *malX*, *usp*, *ompT* (Figure 3; Figure 6). Among them, twelve strains, belonging to the phylogroups B2 (03.25, 04.06, 06.16, 10.17, 12.04, 18.09, and 19.02) and F (08.57, 08.109, 13.20, 16.04, and 16.27), presented a pattern of virulence genes similar to that of the strains responsible for urinary tract infections, on the basis of the presence of the genes *chuA* (heme receptor), *yfcV* (Yfc fimbria), along with *fyuA* (*Yersinia* siderophore receptor) and *ompT* (aspartyl protease) [35]. This reinforces the idea that it is possible that pathogenic *E. coli* potentially inhabit the gut of healthy subjects without instigating infections, although they can act as etiologic agents of extra-intestinal infections, without a clear distinction between commensal *E. coli* and pathogens.

Capsular polysaccharides represent a class of macromolecules contributing to the surface properties. They are involved in important biological processes including adhesion and resistance to the host's immune responses, such as complement-mediated killing and phagocytosis [5]. Many strains (27/51) carried *kpsMTII*, among which six strains also carried *kpsMTIII*, encoding another capsular antigen. Macromolecules on the surface of bacteria confer ultrastructural stability are important for recognition by, and interaction with, the environment and in pathogenic bacteria form a defensive barrier against the host's immune system. All but one strain was negative to the *afa/draBC* gene, excluding the presence of afimbrial adhesins encoded by the *afa* operons [36]. Likewise, for capsular polysaccharides, curli are involved in adhesion and strengthen *E. coli* defenses, particularly against complement-mediated killing by the host, thereby contributing to the pathogenic potential of strains [37]. The positive relationship between curli and biofilm formation was confirmed in the present study, although the latter feature was not frequent among the observed strains. Interestingly, the phylogroups B2 and F, which include the strains with the largest set of potential pathogenic determinants, are less prone to produce curli, a condition that may increase their susceptibility to host defenses.

E. coli pathogenicity islands serve as integration sites for genetic elements. Thus, virulence genes and AR determinants can be rapidly and simultaneously acquired, changing commensal strains into major risk agents [38]. Nineteen stains, representing most of the B2 or F isolates, harbored a PAI island, which produced a positive result for the maltose- and glucose-specific component IIa of the phosphoenolpyruvate dependent phosphotransferase system (*malX*). Furthermore, all the isolates also carried the gene *usp*, encoding a putative bacteriocin generally located on the PAI [31]. Eleven out of 51 strains, belonging to the phylogroups A and B2, harbored the 54-kb polyketide synthases *pks* pathogenicity island, encoding the multi-enzymatic machinery for the synthesis of a peptidepolyketide hybrid genotoxin called colibactin [39]. *hlyF*, encoding for the toxin α-emolysin [40], occurred in other strains, which were spread over most of the phylotypes.

The potentiality of the isolates to exchange genetic material was assessed by searching the gene *traD* and challenging the β-glucuronidase negative *E. coli* strains as recipients in conjugation experiments. *traD* is a key gene in the conjugation process, encoding a main component in the transferosome of type IV secretion systems [19]. The presence of *traD* has been chosen to screen the presence of the

conjugation apparatus and the potentiality of the strains to exchange DNA. The majority of the strains (32/51) carried *traD*, regardless of the phylotype group. When more than a single *E. coli* strain was identified in the same subject (Table 1), some isolates encoded *traD* and others did not. The 9 strains harboring the *iss* gene (responsible for the increased survival of bacteria in the serum and generally carried by the big virulence plasmid, ColV) were all positive to *traD*, according to the fact that ColV plasmids are usually conjugative [41]. The majority of β-glucuronidase-negative strains, assayed as recipients in the conjugation experiment, received the plasmid (8/10), confirming their receptive aptitude for the genetic exchange and DNA shuffling of commensal *E. coli*. Interestingly, 2 of the 8 transconjugant strains harbored *traD*, a putative marker of a conjugative plasmid that is expected to exclude the acquirement of another plasmid through conjugation.

The fitness and competitiveness of the bacteria can be improved by bacteriocin production, which is generally associated with the counterpart immunity protein, thereby protecting the producing host cell from the lethal action of the bacteriocin. Colicins are plasmid encoded toxins produced by *E. coli* under conditions of stress and able to kill related bacteria competing for niches and nutrients. Colicin E7, and its immunity counterpart, ImmE7, have been used to develop bacterial conjugation-based antimicrobial agents [42,43]. The antibacterial activity of the colicin ColE7, bacterial "kill"–"anti-kill" antimicrobial system has been determined, thereby offering new perspectives in the development of *E. coli* targeted antimicrobials. The strains isolated in this study were screened for the presence of the *colE7* and *immE7* genes, in order to investigate the susceptibility of commensal strains to the action of recombinant antimicrobials. Most of the strains were negative for both the genes, suggesting that, if necessary, they could be potentially vulnerable to this new approach.

4. Materials and Methods

4.1. Isolation and Enumeration and of E. coli

Fresh fecal samples were collected from 20 healthy adult subjects, who were all caucasians: 10 males and 10 females aged 35 to 45, following a western omnivore diet, who had not been treated with prebiotics and/or probiotics for 1 month and antibiotics for 3 months. They did not have any relevant diseases or drug treatments in their medical histories. These subjects were enrolled as volunteers among the employees of the University of Modena and Reggio Emilia and their relatives and were not in a relationship with the researchers. All human clinical samples were collected after the subjects gave written informed consent for their participation in the study. All personal data used in the study were anonymized.

Fresh fecal samples were homogenized (10% *w/v*) and serially diluted in isotonic Buffered Peptone Water (Sigma, Steinheim, Germany). Then, they were spread onto HiCrome Coliform Agar (HCCA, Sigma) and incubated. After 24 h at 37 °C, the putative *E. coli* colonies were checked for indole production with Kovac's reagent (Sigma). Colonies that exhibited consistent reactions were ascribed to the species *E. coli*. Isolates were maintained and propagated in a Luria Bertani (LB) medium aerobically at 37 °C.

A total of 96 *E. coli* clones per sample were subjected to ERIC-PCR [44] and RAPD-PCR [45] and clustered into biotypes with a similarity level of 75% using the Pearson correlation coefficient. For each strain, the concentration was determined by multiplying the relative abundance and total plate count of *E. coli*.

For qPCR quantification, the feces were suspended in PBS pH 7.8 (10% *w/w*). Bacterial gDNA was extracted with the QIAmp DNA Stool Mini Kit (Qiagen, Hilden, Germany), quantified with a NanoPhotometer *p*-Class (Implen GmbH, Munchen, Germany), diluted to 2.5 ng/μL in a TE buffer (pH 8), and subjected to qPCR analysis, with primers targeting eubacteria and *E. coli*. [46]. The mixture contained 10 μL of SsoFast EvaGreen Supermix, 4 μL of each 2 μM primer, and 2 μL of the template. qPCR reactions were carried out with the CFX96 Real-Time System (Bio-Rad Laboratories, Redmond,

WA, USA) according to the following program: 98 °C for 2 min; 45 cycles at 98 °C for 0.05 min, 60 °C for 0.05 min, and 95 °C for 1 min; 65 °C for 1 min.

4.2. Phyloptyping and PFGE Genotyping

Standard multiplex PCR was used for all the isolates, according to the revisited phylotyping method published by Clermont et al. [17]. PFGE was performed according to the PulseNet protocol [47]. The genomic DNA was digested with 50 U of *Xba*I at 37 °C for 3 h. Macrorestriction fragments were resolved by counter-clamped homogeneous electric field electrophoresis in a CHEF-DRIII apparatus (Bio-Rad, USA). The run time was 24 h at 6.0 V/cm, with initial and final switch times of 2.2 s and 54.2 s, respectively. PFGE images were digitally captured and analyzed with the GelCompare II 6.0 software (Applied Maths NV, Belgium). Dice's coefficients were calculate based on the band profiles, setting the position tolerance at 1% and the optimization at 1%. A similarity dendrogram was derived using the unweighted pair group method with arithmetic means (UPGMA). Strains were ascribed to the same pulsotype if the PFGE profile possessed >85% similarity.

4.3. Virulence Genotyping

All the isolates were screened by multiplex-PCR for a carriage of 34 genes associated to virulence factors: *afa/draBC*, *cdtB*, *chuA*, *cnf1*, *cvaC*, *fimH*, *focG*, *fyuA*, *gafD*, *hlyD*, *hlyF*, *hra*, *ibeA*, *iha*, *ire*, *iroN*, *iss*, *iutA*, *KpsMTII*, *kpsMTIII*, *malX* (*PAI*), *ompT*, *papAH*, *papC*, *papEF*, *pic*, *rfc*, *sat*, *sfa*, *sfaS*, *traT*, *usp*, *vat*, and *yfcV*. Primer sequences, the pool of primers, and amplification reactions were set up according to Johnson et al. [48].

4.4. Detection of the Genes traD, colE7, immE7, and of pks Island

A triplex-PCR was utilized to screen the isolates for the presence of the genes *colE7*, *immE7*, and t*raD*, utilizing the primers, colE7-F/colE7-R2 (5′-AAGTCAGATGCTGATGTTGC-3′/ 5′-ATAGACACCACCATTTTGGC-3′), immE7-F/immE7-R2 (5′-GCAACTGATGATGTGTTAGATG-3′/ 5′-TGTTTAAATCCTGGCTTACCG-3′), and traD-F/traD-R2 (5′-GATAAATACATGATCTGGTGCGG-3′/ 5′-TTGATGGAAAGCAGTGTGTC-3′), respectively, specifically developed in this study. The amplification reaction was carried out in a final volume of 50 µL containing 5 µL of 10X Dream Taq Green Buffer, 5 µL dNTP Mix (2 mM each), 0.25 µL Dream Taq, 9 µL primer mix (2 µM each), and 50 ng of template DNA. The thermocycle was: 95 °C for 5 min, 30 cycles at 95 °C for 30 sec, 58 °C for 30 sec, and 72 °C for 1 min; and 72 °C for 7 min.

The polyketide synthases *pks* pathogenicity island was searched by duplex-PCR, amplifying the *clbB* and *clbN* genes located at the 5′ and 3′ regions of the island, according to Sarshar et al. [49].

4.5. Antibiotic Susceptibility

Antimicrobial susceptibility was tested with a Vitek2 semi-automated system (bioMerieux, France). Minimum inhibitory concentrations (MICs) were interpreted according to EUCAST (European Committee on Antimicrobial Susceptibility Testing—www.eucast.org). Clinical breakpoints for the tested antibiotics were expressed as mg/L (S, susceptible; R, resistant): for amikacin, S ≤ 8 and R > 16; for amoxicillin/clavulanic acid, S ≤ 8 and R > 8; for cefotaxime, 1 ≤ 8 and R > 2; for ceftazidime, S ≤ 8 and R > 8; for ciprofloxacin, S ≤ 0.5 and R > 1; for gentamicin, S ≤ 2 and R > 4; for piperacillin + tazobactam, S ≤ 8 and R > 16; and for trimethoprim/sulfamethoxazole, S ≤ 40 and R > 80.

4.6. Biofilm and Phenotype Assays

Biofilm formation was assayed by crystal violet (CV) staining, as described in [50]. *E. coli* strains were cultured in LB without NaCl and in an M9 medium (BD Difco, Sparks, Maryland, USA) containing 4 g/L glucose and 0.25 g/L yeast extract (hereinafter referred to as LBWS and M9glu, respectively). The specific biofilm formation (SBF) index was calculated as the ratio between CV absorbance at 570 nm

and the culture's turbidity at 600 nm, setting a threshold of 1. The data reported here are the means of 3 independent experiments, each carried out in triplicate.

Curli and cellulose formation were assayed in LBWS agar plates supplemented with congo red (CR) or calcofluor white (CF), respectively, according to [51]. Red colonies were considered positive to CR. Emission of fluorescence as result of UV exposure (315–400 nm) was considered positive to CF.

4.7. Solid Mating Conjugation Experiments

The donor strain *E. coli* N4i pOX38: Cm (N4i: EcN immE7 Gmr; pOX38: Cm: Tra+ RepFIA+ Cmr) was derived from *E. coli* Nissle 1917 [42,43]. The donor was cultured aerobically at 37 °C in LB supplemented with 15 μg/mL gentamycin and 20 μg/mL chloramphenicol, while the recipients were cultured in LB. Overnight donor and recipient cultures were seeded in their respective media (1% v/v) and aerobically incubated for 2 h. One mL of the recipient exponential culture was centrifuged (10,000× *g* for 5 min at 4 °C) and resuspended in 400 μL of the donor exponential culture. The whole volume was spread onto an LB–agar plate and incubated at 37 °C for 4 h, and then the cells were collected from the surface with 1 mL of PBS, serially diluted, and spread onto HCCA to differentiate the recipient and transconjugant from the donor colonies and HCCA supplemented with 20 μg/mL of chloramphenicol to select and differentiate the transconjugant from the donor colonies.

4.8. Statistical Analysis

Dice's coefficient was utilized to gauge the distance between strains, based on the sets of binary data, i.e., the presence/absence of genes or phenotypic properties. Distance matrices between strains were utilized to compute and display UPGMA dendrograms and Principal Coordinates Analysis plots. The co-occurrence of genetic determinants and phenotypic properties were evaluated by means of contingency tables and Cramér's V metrics, considering the limit of significance as $p < 0.01$ and $p < 0.05$. Statistical analyses were performed with the SPSS Statistics 21 software (Armonk, New York, USA).

5. Conclusions

A characterization of *E. coli* isolated from fecal samples of healthy subjects was performed. Several strains had the potential to cause extraintestinal infections because of the presence of genes associated with adhesins, siderophores, and toxins. The recurrence of strains sharing a pattern of virulence genes similar to that of potentially pathogenic strains may pose a health threat. However, a main outcome of this study shows the great sensitivity of the isolates to antibiotics, which are susceptible to the common antimicrobial therapy used for Gram-negative bacteria, such as amoxicillin/clavulanate. This suggests that the *E. coli* populations inhabiting the gut of healthy subjects and of patients are more likely differentiated on the basis of AR than virulence factors, emphasizing the importance of the prudent use of antibiotics in the general population.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2607/7/8/251/s1. Table S1: Dataset of all the genetic determinants and the phenotypic features of *E. coli* isolates.

Author Contributions: Conceptualization, M.R; data curation, S.R., F.C. and A.A.; investigation, L.R., E.M. and F.B.; methodology, S.R., L.R. and G.G.; supervision, S.R.; writing—original draft, S.R., A.A. and M.R.; writing—review & editing, G.G. and M.S.E.

Funding: No dedicated funds.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Russo, T.A.; Johnson, J.R. A proposal for an inclusive designation for extraintestinal pathogenic *Escherichia coli*: ExPEC. *J. Infect. Dis.* **2000**, *181*, 1753–1754. [CrossRef] [PubMed]
2. Manges, A.R.; Geum, H.M.; Guo, A.; Edens, T.J.; Fibke, C.D.; Pitout, J.D.D. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin. Microbiol. Rev.* **2019**, *32*. [CrossRef] [PubMed]

3. Leimbach, A.; Hacker, J.; Dobrindt, U. *E. coli* as an all-rounder: The thin line between commensalism and pathogenicity. *Curr. Top Microbiol. Immunol.* **2013**, *358*, 3–32. [CrossRef] [PubMed]
4. Croxen, M.A.; Law, R.J.; Scholz, R.; Keeney, K.M.; Wlodarska, M.; Finlay, B.B. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* **2013**, *26*, 822–880. [CrossRef] [PubMed]
5. Sarowska, J.; Futoma-Koloch, B.; Jama-Kmiecik, A.; Frej-Madrzak, M.; Ksiazczyk, M.; Bugla-Ploskonska, G.; Choroszy-Krol, I. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: Recent reports. *Gut Pathog.* **2019**, *11*, 10. [CrossRef]
6. Searle, L.J.; Méric, G.; Porcelli, I.; Sheppard, S.K.; Lucchini, S. Variation in siderophore biosynthetic gene distribution and production across environmental and faecal populations of *Escherichia coli*. *PLoS ONE* **2015**, *10*. [CrossRef]
7. Welch, R.A. Uropathogenic *Escherichia coli*-Associated Exotoxins. *Microbiol. Spectr.* **2016**, *4*. [CrossRef]
8. Kaas, R.S.; Friis, C.; Ussery, D.W.; Aarestrup, F.M. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genom.* **2012**, *13*, 577. [CrossRef]
9. Koraimann, G. Spread and persistence of virulence and antibiotic resistance genes: A ride on the F plasmid conjugation module. *EcoSal Plus* **2018**, *8*, 1. [CrossRef]
10. Johnson, J.R.; Stell, A.L. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* **2000**, *181*, 261–272. [CrossRef]
11. Kudinha, T.; Johnson, J.R.; Andrew, S.D.; Kong, F.; Anderson, P.; Gilbert, G.L. Distribution of phylogenetic groups, sequence type ST131, and virulence-associated traits among *Escherichia coli* isolates from men with pyelonephritis or cystitis and healthy controls. *Clin. Microbiol. Infect.* **2013**, *19*. [CrossRef] [PubMed]
12. Lee, J.G.; Han, D.S.; Jo, S.V.; Lee, A.R.; Park, C.H.; Eun, C.S.; Lee, Y. Characteristics and pathogenic role of adherent-invasive *Escherichia coli* in inflammatory bowel disease: Potential impact on clinical outcomes. *PLoS ONE* **2019**, *14*. [CrossRef]
13. Bailey, J.K.; Pinyon, J.L.; Anantham, S.; Hall, R.M. Commensal Escherichia coli of healthy humans: A reservoir for antibiotic-resistance determinants. *J. Med. Microbiol.* **2010**, *59*, 1331–1339. [CrossRef] [PubMed]
14. Mathai, D.; Kumar, V.A.; Paul, B.; Sugumar, M.; John, K.R.; Manoharan, A.; Kesavan, L.M. Fecal carriage rates of extended-spectrum β-lactamase-producing *Escherichia coli* among antibiotic naive healthy human volunteers. *Microb. Drug Resist.* **2015**, *21*, 59–64. [CrossRef] [PubMed]
15. González, D.; Gallagher, E.; Zúñiga, T.; Leiva, J.; Vitas, A.I. Prevalence and characterization of β-lactamase-producing Enterobacteriaceae in healthy human carriers. *Int. Microbio.* **2019**, in press. [CrossRef]
16. Clermont, O.; Bonacorsi, S.; Bingen, E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **2000**, *66*, 4555–4558. [CrossRef]
17. Clermont, O.; Christenson, J.K.; Denamur, E.; Gordon, D.M. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **2013**, *5*, 58–65. [CrossRef]
18. Wassenaar, T.M. *E. coli* and colorectal cancer: A complex relationship that deserves a critical mindset. *Crit. Rev. Microbiol.* **2018**, *44*, 619–632. [CrossRef]
19. Frost, L.S.; Ippen-Ihler, K.; Skurray, R.A. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **1994**, *58*, 162–210.
20. Chak, K.F.; Kuo, W.S.; Lu, F.M.; James, R. Cloning and characterization of the ColE7 plasmid. *J. Gen. Microbiol.* **1991**, *137*, 91–100. [CrossRef]
21. Raimondi, S.; Amaretti, A.; Gentilomi, G.; Bonvicini, F.; Rossi, M. Antibiotic resistance, virulence factors, phenotyping, and genotyping of Enterobacteriaceae (other than *E. coli*) isolated from the feces of healthy subjects. Unpublished; manuscript in preparation.
22. Nowrouzian, F.L.; Wold, A.E.; Adlerberth, I. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J. Infect. Dis.* **2005**, *191*, 1078–1083. [CrossRef]
23. Ostblom, A.; Adlerberth, I.; Wold, A.E.; Nowrouzian, F.L. Pathogenicity island markers, virulence determinants *malX* and *usp*, and the capacity of *Escherichia coli* to persist in infants' commensal microbiotas. *Appl. Environ. Microbiol.* **2011**, *77*, 2303–2308. [CrossRef]

24. Robins-Browne, R.M.; Holt, K.E.; Ingle, D.J.; Hocking, D.M.; Yang, J.; Tauschek, M. Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Front. Cell. Infect. Microbiol.* **2016**, *6*, 141. [CrossRef]
25. Suzuki, S.; Horinouchi, T.; Furusawa, C. Prediction of antibiotic resistance by gene expression profiles. *Nat. Commun.* **2014**, *5*, 5792. [CrossRef]
26. Sahuquillo-Arce, J.M.; Selva, M.; Perpiñán, H.; Gobernado, M.; Armero, C.; López-Quílez, A.; González, F.; Vanaclocha, H. Antimicrobial resistance in more than 100,000 *Escherichia coli* isolates according to culture site and patient age, gender, and location. *Antimicrob. Agents Chemother.* **2011**, *55*, 1222–1228. [CrossRef]
27. Cole, B.K.; Ilikj, M.; McCloskey, C.B.; Chavez-Bueno, S. Antibiotic resistance and molecular characterization of bacteremia *Escherichia coli* isolates from newborns in the United States. *PLoS ONE* **2019**, *14*. [CrossRef]
28. Morley, V.J.; Woods, R.J.; Read, A.F. Bystander selection for antimicrobial resistance: Implications for patient health. *Trends in Microbiol.* **2019**, in press. [CrossRef]
29. Dyar, O.J.; Hoa, N.Q.; Trung, N.V.; Phuc, H.D.; Larsson, M.; Chuc, N.T.; Lundborg, C.S. High prevalence of antibiotic resistance in commensal *Escherichia coli* among children in rural Vietnam. *BMC Infect. Dis.* **2012**, *12*, 92. [CrossRef]
30. Gurnee, E.A.; Ndao, I.M.; Johnson, J.R.; Johnston, B.D.; Gonzalez, M.D.; Burnham, C.A.; Hall-Moore, C.M.; McGhee, J.E.; Mellmann, A.; Warner, B.B.; et al. Gut colonization of healthy children and their mothers with pathogenic ciprofloxacin-resistant *Escherichia coli*. *J. Infect. Dis.* **2015**, *212*, 1862–1868. [CrossRef]
31. Bag, S.; Ghosh, T.S.; Banerjee, S.; Mehta, O.; Verma, J.; Dayal, M.; Desigamani, A.; Kumar, P.; Saha, B.; Kedia, S.; et al. insights into antimicrobial resistance traits of commensal human gut microbiota. *Microb. Ecol.* **2019**, *77*, 546–557. [CrossRef]
32. Purohit, M.R.; Lindahl, L.F.; Diwan, V.; Marrone, G.; Lundborg, C.S. High levels of drug resistance in commensal *E. coli* in a cohort of children from rural central India. *Sci. Rep.* **2019**, *9*, 6682. [CrossRef]
33. Feng, J.; Li, B.; Jiang, X.; Yang, Y.; Wells, G.F.; Zhang, T.; Li, X. Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. *Environ. Microbiol.* **2018**, *20*, 355–368. [CrossRef]
34. Pormohammad, A.; Nasiri, M.J.; Azimi, T. Prevalence of antibiotic resistance in *Escherichia coli* strains simultaneously isolated from humans, animals, food, and the environment: A systematic review and meta-analysis. *Infect. Drug Resist.* **2019**, *12*, 1181–1197. [CrossRef]
35. Spurbeck, R.R.; Dinh, P.C.; Walk, S.T.; Stapleton, A.E.; Hooton, T.M.; Nolan, L.K. *Escherichia coli* isolates that carry *vat*, *fyuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. *Infect. Immun.* **2012**, *80*, 4115–4122. [CrossRef]
36. Servin, A.L. Pathogenesis of *Afa/Dr.* diffusely adhering *Escherichia coli*. *Clin. Microbiol. Rev.* **2005**, *18*, 264–292. [CrossRef]
37. Biesecker, S.G.; Nicastro, L.K.; Wilson, R.P.; Tükel, Ç. The functional amyloid curli protects *Escherichia coli* against complement-mediated bactericidal activity. *Biomolecules* **2018**, *8*, 5. [CrossRef]
38. Kaushik, M.; Kumar, S.; Kapoor, R.K.; Virdi, J.S.; Gulati, P. Integrons in Enterobacteriaceae: Diversity, distribution and epidemiology. *Int. J. Antimicrob. Agents* **2018**, *51*, 167–176. [CrossRef]
39. Johnson, J.R.; Johnston, B.; Kuskowski, M.A.; Nougayrede, J.P.; Oswald, E. Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island. *J. Clin. Microbiol.* **2008**, *46*, 3906–3911. [CrossRef]
40. Murase, K.; Martin, P.; Porcheron, G.; Houle, S.; Helloin, E.; Pénary, M.; Nougayrède, J.P.; Dozois, C.M.; Hayashi, T.; Oswald, E. *HlyF* produced by extraintestinal pathogenic *Escherichia coli* is a virulence factor that regulates outer membrane vesicle biogenesis. *J. Infect. Dis.* **2016**, *213*, 856–865. [CrossRef]
41. Gophna, U.; Parket, A.; Hacker, J.; Ron, E.Z. A novel *ColV* plasmid encoding type IV pili. *Microbiol.* **2003**, *149*, 177–184. [CrossRef]
42. Starčič Erjavec, M.; Petkovšek, Ž.; Kuznetsova, M.V.; Maslennikova, I.L.; Žgur-Bertok, D. Strain ŽP—The first bacterial conjugation-based "kill"-" anti-kill" antimicrobial system. *Plasmid* **2015**, *82*, 28–34. [CrossRef]
43. Maslennikova, I.L.; Kuznetsova, M.V.; Toplak, N.; Nekrasova, I.V.; Žgur Bertok, D.; Starčič Erjavec, M. Estimation of the bacteriocin ColE7 conjugation-based "kill" - "anti-kill" antimicrobial system by real-time PCR, fluorescence staining and bioluminescence assays. *Lett. Appl. Microbiol.* **2018**, *67*, 47–53. [CrossRef]
44. Wilson, L.A.; Sharp, P.M. Enterobacterial repetitive intergenic consensus (ERIC) sequences in *Escherichia coli*: Evolution and implications for ERIC-PCR. *Mol. Biol. Evol.* **2006**, *23*, 1156–1168. [CrossRef]

45. Quartieri, A.; Simone, M.; Gozzoli, C.; Popovic, M.; D'Auria, G.; Amaretti, A.; Raimondi, S.; Rossi, M. Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe* **2016**, *38*, 130–137. [CrossRef]
46. Simone, M.; Gozzoli, C.; Quartieri, A.; Mazzola, G.; Di Gioia, D.; Amaretti, A.; Raimondi, S.; Rossi, M. The probiotic *Bifidobacterium breve* B632 inhibited the growth of Enterobacteriaceae within colicky infant microbiota cultures. *Biomed. Res. Int.* **2014**. [CrossRef]
47. Standard Operating Procedure for PulseNet PFGE of Escherichia coli O157:H7, Escherichia coli non-O157 (STEC), Salmonella serotypes, Shigella sonnei and Shigella flexneri. Available online: http://www.cdc.gov/pulsenet/PDF/ecoli-shigella-salmonella-pfge-protocol-508c.pdf (accessed on 19 July 2019).
48. Johnson, J.R.; Porter, S.; Johnston, B.; Kuskowski, M.A.; Spurbeck, R.R.; Mobley, H.L.; Williamson, D.A. Host characteristics and bacterial traits predict experimental virulence for *Escherichia coli* bloodstream isolates from patients with urosepsis. *Open Forum Infect. Dis.* **2015**, 2. [CrossRef]
49. Sarshar, M.; Scribano, D.; Marazzato, M.; Ambrosi, C.; Aprea, M.R.; Aleandri, M.; Pronio, A.; Longhi, C.; Nicoletti, M.; Zagaglia, C.; et al. Genetic diversity, phylogroup distribution and virulence gene profile of *pks* positive *Escherichia coli* colonizing human intestinal polyps. *Microb. Pathog.* **2017**, *112*, 274–278. [CrossRef]
50. Sicard, J.F.; Vogeleer, P.; Le Bihan, G.; Rodriguez Olivera, Y.; Beaudry, F.; Jacques, M.; Harel, J. N-Acetyl-glucosamine influences the biofilm formation of *Escherichia coli*. *Gut Pathog.* **2018**, *22*, 10–26. [CrossRef]
51. Vogeleer, P.; Tremblay, Y.D.N.; Jubelin, G.; Jacques, M.; Harel, J. Biofilm-forming abilities of shiga toxin-producing *Escherichia coli* isolates associated with human infections. *Appl. Environ. Microbiol.* **2015**, *28*, 1448–1458. [CrossRef]