Classification of *Parabacteroides distasonis* and other *Bacteroidetes* using O-antigen virulence gene: *RfbA*-Typing and hypothesis for pathogenic vs. probiotic strain differentiation

Nicholas C. Bank, Vaidhvi Singh, and Alex Rodriguez-Palacios

ABSTRACT

*Parabacteroides distasonis* (Pdis) is the type species for the new *Parabacteroides* genus, and a gut commensal of the *Bacteroidetes* phylum. Emerging reports (primarily based on reference strain/ATCC-8503) concerning propose that long-known opportunistic pathogen *Pdis* is a probiotic. We posit there is an urgent need to characterize the pathogenicity of *Pdis* strain variability. Unfortunately, no methods/insights exist to classify *Bacteroidetes for this purpose. Herein, we developed a virulence gene-based classification system for *Pdis* and *Bacteroidetes* to facilitate pathogenic-vs-probiotic characterization. We used DNA in silico methods to develop a system based on the virulence (lipopolysaccharide/bacterial wall) *rfbA* O-antigen-synthesis gene'. We then performed phylogenetic analysis of *rfbA* from fourteen *Pdis* complete genomes (21 genes), other *Parabacteroides*, *Bacteroidetes* and *Enterobacteriaceae*; and proposed a PCR-based Restriction-Fragment Length Polymorphism method. Cluster analysis revealed that *Pdis* can be classified into four lineages (based on gene gaps/insertions) which we designated *rfbA*-Types I, II, III, and IV. In context, we found 14 additional *rfbA*-types (I–XVIII) interspersed with numerous *Bacteroidetes* and pathogenic *Enterobacteriaceae* forming three major *rfbA*-superclusters. For laboratory *rfbA*-Typing implementation, we developed a PCR-primer strategy to amplify *Pdis* *rfbA* genes (100%-specificity) to conduct MboII-RFLP and sub-classify *Pdis*. *In-silico* primers for other *Bacteroidetes* are proposed/discussed. Comparative analysis of lipopolysaccharide/lipid-A gene *lpxK* confirmed *rfbA* as highly discriminant. In conclusion, *rfbA*-Typing classifies *Bacteroidetes/Pdis* into unique clusters/superclusters given *rfbA* copy/sequence variability. Analysis revealed that most pathogenic *Pdis* strains are single-copy *rfbA*-Type I. The relevance of the *rfbA* strain variability in disease might depend on their hypothetical modulatory interactions with other O-antigens/lipopolysaccharides and TLR4 lipopolysaccharide-receptors in human/animal cells.

Introduction

*Parabacteroides distasonis* (Pdis), a gram-negative bacterium of the intestinal tract, is the type strain for the genus *Parabacteroides*, for which there is emerging controversy regarding the role that they play in human and animal health. Although *Pdis* has been recognized as an intestinal commensal since the mid-1930s, there is a recent increase in reports describing contradictory pathogenic (detrimental) and probiotic (beneficial) effects on human and animal health. There is need to identify and characterize factors that could account for reported differences in potential *Pdis* strain pathogenicity, especially because there is an emerging interest in using *Pdis* as a human probiotic, which poses a major risk to public health. To date, such contradictions indicate that the differences in the health effects could be due to strain differences. Unfortunately, there are no classification or cataloging systems to help typify *Pdis* into lineages using a virulence meaningful approach.

Extrapolating from prior research on *E. coli* and *Salmonella*, strain-dependent mechanisms linked to bacterial surface markers, such as the O-antigen, could be used to help guide research and help propose studies to determine the causes...
that lead to the varied effects observed for \textit{Pdis} on human and animal health. The O-antigen is a key virulence molecule of lipopolysaccharides (LPS) constitutively expressed on the cell wall surface of gram-negative bacteria. Lipopolysaccharide is a well understood virulence factor for gram-negative bacteria, consisting of lipid A, an oligosaccharide, and the O-antigen polysaccharide. The O-antigen is the immunogenic component of LPS, and as such can influence the host–bacterium relationship in several ways; potential mechanisms include resisting host complement and phagocytic engulfment, molecular mimicry, and colonization ability.\textsuperscript{21} Additionally, variation in the amount/type of monosaccharides in O-antigens provide major LPS structural diversity\textsuperscript{21} and virulence potential across bacteria (e.g. \textit{E. coli} O157).

Historically, strains of gram-negative bacteria like \textit{E. coli}, \textit{Salmonella}, and \textit{Shigella} have been classified based on O-antigen structure, using antigen and antibody agglutination reactions, in a laboratory method known as O-serotyping.\textsuperscript{22} While O-serotyping is a well-known and standardized practice, this method is still deemed highly variable across laboratories.\textsuperscript{23,24} Established viable alternatives to O-serotyping have been demonstrated for several \textit{Enterobacteriaceae}. In an attempt to streamline the classification schemes to typify \textit{Enterobacteriaceae}, newer techniques have examined the use of DNA sequences of genes involved in the synthesis and processing of the O-antigen.\textsuperscript{25,26}

Genes that encode the enzymes for synthesis of O-antigens are classically clustered in a region known as the \textit{rfb} cluster. Within the cluster, the \textit{rfbA} gene encodes the leading enzyme glucose-1-phosphate thymidyltransferase to produce the O-antigen. This enzyme catalyzes the formation of dTDP-glucose from dTTP and glucose 1-phosphate as the first reaction in the O-antigen synthesis pathway.\textsuperscript{27} Since this is a critical step, we proposed the analysis of this gene to help elucidate strain differences in \textit{Pdis}. This is especially important because the \textit{rfbA} gene was recently identified in two strains of \textit{Pdis}, namely, CavFT-hAR46\textsuperscript{28} and CavFT-hAR56 (this study), which were isolated from gut wall cavitating microlesions of two different patients with severe Crohn’s disease.\textsuperscript{28}

As there are no studies on the actual molecular structures of the O-antigen in \textit{Pdis} to make a proven connection to pathogenesis, here we propose a framework strategy to characterize the variations in length and phylogeny of the \textit{rfbA} gene, since such modifications influence the virulence of the O-antigen products as it is known in \textit{E. coli},\textsuperscript{29,30} and since the presence and length of O-antigen in LPS play an important role in bacterial pathogenesis.\textsuperscript{31,32} Thus, \textit{rfbA} gene variations may help categorize \textit{Pdis} strains for future functional pathogenic vs. probiotic characterization studies. The objective of this study was to develop a classification and cataloging system for \textit{Pdis}, applicable to other \textit{Bacteroidetes}, based on \textit{rfbA} according to variations in gene copy number and polymorphisms. Herein, we use DNA sequence and \textit{in silico} methods to develop a classification system for \textit{Pdis} based on the \textit{rfbA} gene and discovered \textit{Pdis}-specific \textit{rfbA}-types (I–IV), \textit{rfbA}-specific primers for \textit{Pdis} (reverse and forward), three major superclusters when the \textit{Pdis} data was contextu-lized with sequences of numerous \textit{Bacteroidetes} and \textit{Enterobacteriaceae}, and we assessed the discriminatory ability of this system compared to lipid A (of LPS) biosynthesis genes.

\section*{Materials and methods}

\subsection*{Sequence data used}

We performed \textit{in silico} analyses of \textit{rfbA} and \textit{lpxK} gene sequences from bacterial genomes available in NCBI and Pathosystems Resource Integration Center (PATRIC) for \textit{P. distasonis}, other \textit{Parabacteroides} spp.; \textit{Bacteroidetes} (\textit{Bacteroides}, \textit{Alistipes}, \textit{Prevotella}); and \textit{Enterobacteriaceae} (\textit{Escherichia}, \textit{Klebsiella}, \textit{Salmonella}, \textit{Shigella}).

\subsection*{Phylogenetic analyses}

\textit{Parabacteroides distasonis} \textit{rfbA} gene sequences previously collected from NCBI were compiled in CLC Viewer 8.0 (commercially available) and used to construct two alignments and phylogenetic trees: one alignment and tree for the \textit{Pdis} \textit{rfbA} nucleotide sequences and the second for the translated amino acid sequences. Additionally, a third alignment and tree including \textit{rfbA} gene sequences from \textit{Pdis}, other
Parabacteroides spp., Bacteroides spp., Alistipes spp., Prevotella spp., Escherichia spp., Klebsiella spp., Salmonella spp., and Shigella spp., were constructed to provide evolutionary context and for observation of clustering patterns. Results were used to determine phylogenetic relatedness and rfbA gene variance within Pdis strains as well as between Pdis and other bacterial species. A separate phylogenetic analysis of lpxK gene nucleotide and amino acid sequences was performed, the results of which were used to determine the comparative discriminatory ability of the rfbA gene to aid in strain characterization.

**Parabacteroides distasonis sequence cluster analysis**

Results from the Pdis rfbA gene phylogeny were then used to perform sequence cluster analysis. Gene copy analysis was first performed by recording the number of unique rfbA genes present in each represented Pdis strain. Pdis cluster analysis was then performed using the rfbA gene alignment data. Initially, aligned sequences were organized based on gross structure (i.e. matching patterns of nucleotide insertions and/or deletions). A sequence cluster, hereafter referred to as rfbA-Type, we defined as a set of rfbA sequences with an identical pattern of insertions and deletions. For each rfbA-Type, a cluster representative [CR] strain was chosen based on having the fewest rfbA copy number variations (CNVs) of its cluster or unique status (e.g. ATCC 8503 is the reference strain for the entire Pdis species and therefore will be the [CR] strain within its rfbA-Type). Next, within each rfbA-Type, individual rfbA gene sequences were analyzed for nucleotide (Nt) homology in three areas: (i) Nt percent homology with its rfbA-Type [CR] strain, (ii) Nt percent homology with the Pdis reference strain ATCC 8503 rfbA gene sequence, and (iii) rfbA-Type inter-cluster consensus sequence percent homology. Lastly, Pdis strains of each rfbA-Type were assigned subtypes (A-F) based on descending percent homology to their [CR] strain (e.g. highest percent homology was designated subtype A, second highest – subtype B, third highest – subtype C, fourth highest – subtype D, etc.). The collected Pdis rfbA gene sequences were then translated in CLC Viewer 8.0 to amino acid sequences in the +1, +2, and +3 reading frames. For each rfbA-Type, nucleotide and corresponding amino acid sequences were assessed for trends in conservation levels, both within and across reading frames. Additionally, notable amino acid mutations and their corresponding nucleotide polymorphisms within rfbA-Types were recorded.

**Bacteroidetes/Enterobacteriaceae sequence cluster analysis**

Gene copy analysis was also performed on the results from the Pdis, Bacteroidetes, and Enterobacteriaceae rfbA gene phylogeny, followed by observation for clustering patterns through which the previously designed Pdis rfbA-Typing system could be extrapolated to other species and genera. Expanded rfbA-Types were assigned based on bootstrap values and phylogenetic tree morphology (see statistics below), with each apparent cluster receiving a different rfbA-Type designation.

**Primer design for amplification of rfbA gene**

Primer design was conducted by identifying left and right flanking regions of the rfbA gene alignment which were whole (i.e. no gaps or deletions) throughout all rfbA sequences. Then, from the corresponding regions of the rfbA gene alignment consensus sequence, a left flank of 52 base pairs and right flank of 49 base pairs were selected and processed with Primer3Plus to identify optimal primers. Primer sequences were then entered into Basic Local Alignment Search Tool (BLAST) to confirm accuracy in identifying Pdis strains. These methods were then used to design primers for other Bacteroidetes genera (Parabacteroides, Bacteroides, Alistipes, Prevotella).

**In silico rfbA-RFLP analysis**

The collected rfbA genes were uploaded into DNASTAR (commercially available) and processed in two separate agarose gel simulations using the previously validated restriction enzyme MboII. The first gel was simulated with the full-length rfbA gene PCR amplicon, and the following simulation used end-truncated rfbA gene PCR amplicons after...
application of the previously designed primers. The digestion patterns were then used to design the Pdis RFLP typing system, a complementary classification scheme to the rfbA typing system derived from sequence cluster analysis.

Statistics

Sequences were aligned using CLC Viewer 8.0. Then, alignments were used to construct neighbor-joining phylogenetic trees using Jukes-Cantor to account for the nucleotide (Euclidean) distances across sequences. Branch reproducibility was quantified using bootstrapping for 1000 replicates (bootstrap values are shown in trees). Branch morphology and bootstrap values greater than 90 were used as a guide to designate a branch as a distinct rfbA-Type cluster. To assess the reproducibility of cluster assemblage, strain allocation within hierarchical clusters for phylogenetic trees generated with only Pdis sequences (21 rfbA gene sequences, from 14 strains) were compared to the allocation of the same strains within clusters generated in a phylogenetic tree containing a total of 89 rfbA genes from 49 other species, a total of 8 genera, using Fisher’s exact test. Significance was held at $p < .05$. For Pdis sequence cluster analyses, percent homologies for $i$, $ii$, and $iii$ were calculated using the Expasy SIM Alignment Tool (https://web.expasy.org/sim/), and the statistical significance of homologies for $ii$ was assessed using Kruskal–Wallis one-way analysis of variance (ANOVA) in conjunction with Dunn’s test.

Results

Copy number and the phylogeny of rfbA nucleotide sequence in Parabacteroides distasonis

From 15 possible Parabacteroides species (as complete genomes) available to date, rfbA sequences were only available for 13 species (P. distasonis, P. johnsonii, P. merdae, P. golsteinii, P. acidiﬁaciens, P. faecis, P. bouchesdurhonensis, P. chartae, P. massiliensis, P. provencensis, P. timonensis, P. pacaensis, P. gordonii). Sequences for the rfbA gene in P. chongii and P. chinchilla were not available. DNA sequences were analyzed using phylogenetic and hierarchical analysis to illustrate the genetic distances of 21 rfbA gene sequences identified in 14 Pdis genomes. Of interest, illustrating the gene diversity and conservancy within Pdis, we found that the rfbA gene can be present in Pdis genomes as single, double, or triple copies, with distinct or similar gene homologies within each genome, and that different genes have unique reproducible patterns of gaps and insertions which enable the designation of rfbA-Types (Figure 1a). Phylogenetic analysis revealed distinct grouping of Pdis strains into four main clusters. Given the presence of multiple rfbA gene copies, some rfbA sequences from strains FDAARGOS 615, 82G9, NBRC 113806, and CBBP-1 were present in more than one cluster (Figure 1b). Of note, data indicate that when a Pdis genome has >1 rfbA copy, the copies are of different sequence type, except for CBBP-1, which has three rfbA copies only matching two sequence types (Figure 1c). The role of such diversity in health remains unknown.

Classification system based on rfbA-typing for Parabacteroides distasonis

We subsequently examined the structural differences between rfbA gene sequences. Based on rfbA sequence structural variation, our analysis revealed that Pdis could be classified into lineages based on the presence of gaps and insertions in the gene sequence, of which we identified Types I, II, III, IV, and Subtypes A, B, C, D, E, and F (Table 1), among the 21 rfbA genes derived from fourteen complete Pdis genomes. The rfbA-Type I cluster is composed of eleven Pdis strains, including the species reference strain ATCC 8503. rfbA-Types II, III, and IV contain four, four, and two strains, respectively, and have the strains ATCC 82G9, FDAARGOS 1234, and NBRC 113806 as representatives for future analysis. As a measure of inter-cluster distances, Figure 1d illustrates that some clustered sequences have a very low % of DNA homology. Relative to the Pdis reference strain ATCC 8503, individual strand % homologies were highest in rfbA-Type I strains and lowest in rfbA-Types III and IV strains (Figure 1e, K-W, $p < .0001$). Such differences across the rfbA gene could explain differences in O-antigen related virulence across strains, especially if strains have major gene sequence differences as the ones observed...
when comparing rfbA-types I and IV, which can be as low as 43% using the ATCC 8503 rfbA gene sequence as a referent. Inferred amino acid sequence analysis demonstrated, across all Pdis rfbA genes, that the clustering structure observed for the DNA remained unaltered when using the amino acid sequence data (Figure 1f).

**Conservation of amino acid sequences predicted from rfbA genes over decades and geography**

After observing the variance in rfbA DNA sequence homology, analysis of the respective amino acid sequence conservation levels was performed to understand which gene polymorphisms may ultimately affect the final rfbA protein product. The +1-reading frame, relative to the +2 and +3 reading frames, showed the highest conservation levels in an alignment of all twenty-one rfbA amino acid sequences from Pdis genomes. Aside from rfbA-Type IV, which is comprised of only two, 100% homologous rfbA gene sequences, amino acid conservation of the rfbA-Type I cluster (+1-reading frame) was the highest among all rfbA types with only three sites of amino acid mutations across the 292 amino acid-long sequences (with 27 sites of nucleotide polymorphisms present in the rfbA-Type I gene cluster) (Figure 2a-b).
Within the rfbA-Type I amino acid sequence cluster, one of the three mutations, pos. 127 A -> V, is uniquely present in the CavFT-hAR46/46_b/56 strains. Two other mutations, pos. 185 G -> D, and G -> S, are uniquely present in the FDAARGOS 615 and CL06T03C10 strains, respectively. Of interest, the Type I rfbA sequences belonging to the CavFT-hAR46 and CavFT-hAR56 vs. ATCC 8503 strains are 98.6% homologous at the nucleotide level despite these two USA isolates being isolated by different institutions from patients over 85 years apart.\(^\text{28,41}\) The initial presumed designation as ‘pathogenic’ or ‘probiotic’ and the geographical distribution of the reference strains used is listed in Table 2. Taken together, the concurrent presence of high DNA and protein sequence homology may indicate which polymorphic sites are most relevant in altering the pathogenic potential of Pdis.

**Table 1. rfbA-Type and subtype classifications based on sequence homology and *in-silico* RFLP.**

| rfbA-Type | Strains | % Homology to [CR] | % Homology to ATCC 8503 | rfbA-Subtype | RFLP-Type |
|-----------|---------|--------------------|-------------------------|-------------|-----------|
| rfbA-Type I | ATCC 8503 (CR) | 100 | 100 | A | 1 |
| | CL1T00C22 | 99.4 | 99.4 | B | 1 |
| | 82G9* | 99.2 | 99.2 | C | 2 |
| | FDAARGOS_759 | 99.2 | 99.2 | C | 2 |
| | CT06 | 99.2 | 99.2 | C | 1 |
| | NBRC 113806* | 99.1 | 99.1 | D | 1 |
| | FDAARGOS_615* | 99 | 99 | E | 1 |
| | CavFT-hAR46 | 98.6 | 98.6 | F | 1 |
| | CavFT-hAR46_b | 98.6 | 98.6 | F | 1 |
| | CavFT-hAR56 | 98.6 | 98.6 | F | 1 |
| rfbA-Type II | 82G9** (CR) | 100 | 72.1 | A | 3 |
| | FDAARGOS_615** | 100 | 72.1 | A | 3 |
| | CBBP-1* | 91.5 | 71.7 | B | 4 |
| | CBBP-1** | 79.7 | 70.9 | C | 5 |
| rfbA-Type III | FDAARGOS_1234 [CR] | 100 | 42.2 | A | 6 |
| | CBBP-1*** | 99.9 | 43.1 | B | 6 |
| | APCS2_PD | 99.3 | 42.3 | C | 6 |
| | CL03T12C09 | 99.2 | 42.2 | D | 7 |
| rfbA-Type IV | NBRC 113806** [CR] | 100 | 43.5 | A | 8 |
| | FDAARGOS_615*** | 100 | 43.5 | A | 8 |

Additional Parabacteroides spp., Bacteroides spp., Alistipes spp., Prevotella spp., Escherichia spp., Klebsiella spp., Salmonella spp., and Shigella spp. illustrates consistent phylogeny of the four original Pdis rfbA-Types compared to their isolated cluster analysis in Figure 1b, and all four Pdis rfbA-Types are unique to the species (2x4 Fisher’s exact p = 1.0; i.e. the clusters were identical with or without other bacteria). Gene copy analysis of the phylogeny reveals the presence of multiple unique rfbA gene copies in A. shahii, A. onderdonkii, and A. finegoldii; E. coli O1:H42, K-12, O2:H6, and E. fergusonii; S. flexneri; and K. aerogenes. While P. distasonis strains exhibit anywhere from 1 to 3 copies of the rfbA gene in their genome, the twelve other Parabacteroides species have only one rfbA copy per genome (Figure 3, highlighted in green).

Cluster analysis further revealed formation of three ‘rfbA-superclusters’; Supercluster 1 solely contains the Enterobacteriaceae Escherichia, Klebsiella, Salmonella, and Shigella; Supercluster 2 predominantly consists of Pdis (rfbA-Types I and II) as well as a representative from each included genus except for Shigella; Supercluster 3 contains at least one representative from each included genus, the majority of Parabacteroides spp., and the Pdis rfbA-Types III and IV (Figure 3). Within each supercluster, new rfbA-Types, determined by bootstrap values and branch morphology, were assigned to

**Context of rfbA-type classification in P. distasonis vs. Bacteroidetes and Enterobacteriaceae**

To better understand the newly developed P. distasonis rfbA-Typing framework, we examined its context among the larger set of gram-negative bacteria; the closely related Bacteroidetes and more distantly related Enterobacteriaceae (total n = 49 additional species; 8 genera). Genetic mapping of the twenty-one Pdis rfbA genes amongst rfbA genes from 12
extrapolate the rfbA-Typing system beyond P. distasonis. Results from expanded rfbA-Typing show up to eighteen total unique rfbA-Types (Pdis rfbA-Types I–IV plus fourteen newly assigned rfbA-Types) discernable from the eighty-nine-sequence phylogeny, with one rfbA-Type in Supercluster 1, eight rfbA-Types in Supercluster 2, and nine rfbA-Types comprising Supercluster 3. The clustering pattern of new rfbA-Types appears to be predominantly restricted to species of the same genera (rfbA-Types V, VI, IX, XI, XII, XIV, and XVII) with the most notable exception being the rfbA-Types consisting of a mix of Enterobacteriaceae genera (rfbA-Types X, XV, XVI, and XVIII). Species within rfbA-Types VII and VIII appear to be relatively more variable and could indicate the need for additional related species and/or rfbA gene sequences to illustrate more well-defined clusters.

**Design of primers for amplification of ‘end-truncated Pdis rfbA gene’ in laboratory isolates**

We designed Pdis rfbA gene consensus-sequence based primers to facilitate identification of the rfbA gene in future Pdis isolates, regardless of rfbA-Type (Figure 4a-b). The forward primer 5’-CCGCTTGTATCGAT-C-3’ and reverse primer 5’-AAATACTGCGCGTACTGATTCTT-3’ were identified using Primer3Plus and verified with BLAST to identify Pdis strains with 100% specificity (see Supplementary Excel File). Use of
| Strain          | rfbA copy number | rfbA-Type | Presumed pathogenic/probiotic | Year | Isolation country | Isolation source | Remarks | References |
|----------------|------------------|-----------|------------------------------|------|-------------------|------------------|---------|------------|
| ATCC 8503      | 1                | I         | Pathogenic                   | 1933 | USA               | Human feces      | Isolated from distal human gut microbiota and used as reference genome in this study | 44 |
| CavFT-hAR46    | 1                | I         | Pathogenic                   | 2019 | USA               | Human intramural gut wall | Isolated from a gut wall cavitating micro-lesion in a patient with Crohn’s disease | 28 |
| CavFT-hAR46_b  | 1                | I         | Pathogenic                   | 2019 | USA               | Human intramural gut wall | Isolated from a gut wall cavitating micro-lesion in a patient with Crohn’s disease | Same isolate as 28 Re-sequenced, Not Published |
| CavFT-hAR56    | 1                | I         | Pathogenic                   | 2019 | USA               | Human intramural gut wall | Isolated from a gut wall cavitating micro-lesion in a patient with Crohn’s disease | 28 |
| FDAARGOS_759   | 1                | I         | Not specified                | -    | Human feces       | Used as reference genomes in NCBI | - |
| CL11T00C22     | 1                | I         | Not specified                | 2009 | USA               | Human feces       | Isolated from feces of healthy adult | 46 |
| CLO67T03C10    | 1                | I         | Not specified                | 2009 | USA               | Human feces       | Isolated from feces of healthy adult | 46 |
| CLO67T03C09    | 1                | III       | Not specified                | 2009 | USA               | Human feces       | Isolated from feces of healthy adult | 46 |
| FDAARGOS_1234  | 1                | III       | Not specified                | -    | Human feces       | -                  | Laboratory host for propagation of bacteriophage PDS1 | Not Published |
| APCS2/PD       | 1                | III       | Not specified                | 2017 | Ireland           | Human feces       | - |
| 8269           | 2                | I, III    | Pathogenic                   | -    | Japan             | Human feces       | Isolated from human feces | Not Published |
| NBRC 113806    | 2                | I, IV     | Not specified                | -    | Human feces       | Human feces       | Clinical isolate | Not Published |
| FDAARGOS_615   | 3                | I, II     | Pathogenic                   | -    | Human feces       | Human feces       | - |
| CBBP-1         | 3                | II, III   | Probiotic                    | -    | Feces             | Feces             | - |

These primers should generate an expected PCR product of about 850 bp in length (Figure 4c). To examine the potential to identify one general primer that could encompass all species at the genus level, we conducted in-silico primer design and analysis for the major cluster of Bacteroidetes (Parabacteroides, Bacteroides, Alistipes, and Prevotella) shown in Figure 3. A sample of selected primers for the genera and their BLAST performances are listed in the Supplementary Excel File. Of note, the development of primers to cover most species within each genus was more challenging to design due to a wide array of gaps and insertions across multiple species owed to rfbA sequence variability (Supplementary Figure 1).

Enhanced discrimination ability of Parabacteroides distasonis using MboII-RFLP rfbA subtyping

To further aid in rfbA gene-based identification methods of Pdis isolates, we designed a lab-accessible application of the rfbA-typing system based on RFLP. Results from in silico rfbA-RFLP of both full-length genes and end-truncated PCR amplicons demonstrate eight unique patterns by MboII restriction digest. (Figure 4d-e). The rfbA-Type I sequences demonstrated two unique digestion patterns, and rfbA-Type II, III, and IV demonstrated three, two, and one unique digestion patterns, respectively. RFLP in silico analysis revealed that the PCR amplicons produced with our designed primers yielded similar patterns of classification compared to the complete gene sequences derived from the complete Pdis genomes.

Effective discrimination of strains by rfbA compared to the lipid A phosphorylation gene lpxK

Lipopolysaccharides (LPS) are composite molecules consisting of a lipid (Lipid A) and a polysaccharide composed of the O-antigen, and two (outer and inner) oligosaccharide cores linked by covalent bonds. Thus far, we have herein examined the phylogeny of the rfbA gene. To quantify the discriminatory ability of the rfbA gene in differentiating Pdis strains, we compared a potential classification scheme based on the sequence homologies of the same strains using the lpxK gene which encodes a lipid A phosphorylation enzyme. The lpx genes have been implicated in toll-like receptor 4 (TLR4) mediated pathology of other gram-negative...
Analysis shows that i) the \( lpxK \) gene is present as a single copy in all the \( Pdis \) genomes used in this study, and that ii) all strains have high sequence homology and conservation at nucleotide (98.65%) and amino acid level (98.1%), making the \( lpxK \) gene suboptimal as a classification system (Figure 5). This comparative analysis reassures that the classification and cataloging proposed for \( Pdis \) strains based on the O-antigen \( rfbA \) gene is highly discriminant and useful compared to the lipid A \( lpxK \) gene. As an immediate mechanistic application of this \( rfbA \)-Typing system, we examined a potential interaction of O-antigen/LPS molecules (or membrane fractions) of various \( P. distasonis \) strains with that of other O-antigen/LPS molecules of known pathogens, for example \( E. coli \) or other \( Enterobacteriaceae \). To facilitate the understanding of such hypothetical interactions, Figure 6 illustrates how variability in \( rfbA \)/LPS geno/phenotypes could modulate proinflammatory/apoptotic pathways.

**Figure 3.** \( rfbA \)-Superclusters and contextualization of \( rfbA \)-Typing system for the \( P. distasonis \) with respect to other \( Bacteroidetes \) and pathogenic \( Enterobacteriaceae \). Asterisks (*, **, ***)) denote first, second, and third copy of the \( rfbA \) gene in each species and/or strain.
Figure 4. *In silico* MboII restriction digest of full-length and end-truncated *P. distasonis* *rfbA* genes. (A) Alignment of complete *rfbA* genes. (B) Alignment of primer end-truncated *rfbA* genes. (C) RFLP protocol. (D) Gel electrophoresis patterns for full-length *rfbA* genes after MboII digest. (E) Gel electrophoresis patterns for end-truncated *rfbA* genes after MboII digest. Gels shown in panels D and E depict *Pdis* *rfbA*-Types alongside newly designated RFLP Types based on restriction digest patterns. Asterisks (*, **, ****) denote first, second and third copy of the *rfbA* gene in each species and/or strain.
Nomenclature and reporting of \textit{rfbA}-Types

To facilitate the reporting of an \textit{rfbA} profile, we suggest the use of the following designation format: i) \textit{P. distasonis} \textit{rfbA}-Type n1-I for strains having one \textit{rfbA} gene copy (n1) with a nucleotide sequence of the \textit{rfbA}-Type I; ii) \textit{P. distasonis} \textit{rfbA}-Type n3-I, II, II’ for strains with three gene copies, one of each was either types I, II or III; and iii) \textit{P. distasonis} \textit{rfbA}-Type n4-I,IV’ for strains with four copies, with at least one Type I and one Type IV. If more details are desired,
the latter case could be presented as iv) *P. distasonis* ‘rfbA-Type n4-I,IV(2,2)’ or ‘Type n4-I,IV(1,3)’ or ‘Type n4-I,IV(3,1)’ to provide the detailed counts of each unique sequence type in subscript parentheses with numerically ordered digits representing each ordered corresponding *rfbA*-Type.

**Discussion**

*Parabacteroides distasonis* has emerged in recent years for its contradictory dual potential for pathogenicity and probiotic ability, although our current knowledge of the potential for this bacterium to modulate health or cause disease is suboptimal and incomplete. Of the 14 studies cited, only 7 detailed the specific *Pdis* strain examined;5,7,9–12,15,17 the strain being either ATCC 85035,7,9–12,15,17 or a strain not cataloged in NCBI.12,15 Data available in the literature and NCBI on the presently examined 14 strains of *Pdis* indicate that 5 strains are presumed pathogenic, 1 probiotic, whereas 7 were neither presumed to be probiotic or pathogenic. Out of the presumed five pathogenic strains, two were isolated from gut wall cavitating micro-lesions in two patients with severe surgical Crohn’s disease, one was associated with enhancing colitis in mice, and two were human clinical isolates. Of potential relevance to disease, *rfbA*-Type I was a common genotype to all the pathogenic strains of *Pdis*, irrespective of the number of *rfbA* gene copies in the genome (Table 2).

While this bacterium has reported associations with IBD and other diseases, its specific mechanisms are not well understood.2 The fact that *P. distasonis* had been found in extraintestinal lesions (e.g., abscesses) does not necessarily indicate that *Pdis* is a primary pathogen, but rather indicates that the dissemination of this bacterium from the gut lumen may make *Pdis* an opportunistic pro-inflammatory microorganism. To what extent this intestinal commensal promotes inflammation in the gut wall in humans and how this varies with human genetics and predisposition to IBD remains unclear, but strain isolation from pus-containing intramural microscopic lesions (CavFT-hAR46 and CavFT-hAR56) indicates that opportunistic inflammation may depend on the environment where *Pdis* is encountered.28 Furthermore, experiments in animals with genetic deficiencies and induced colitis (peptidoglycan recognition protein *pglyrp* gene, 5% DSS colitis) have shown that *Pdis* (ATCC 8503) is a colitis-promoting species (in BALB/c mice with specific-pathogen free microbiota, with and without antibiotics), compared to mice that did not receive *Pdis*, or to mice that received *Alistipes finegoldii* (another type strain for an emerging *Bacteroidetes* genus56) which protected mice from colitis.5 To contextualize the relevance of the 5% DSS model and the pathogenic effect of live *Pdis* aggravating colitis in this model, it is worth noting that the group of mice treated with *Alistipes finegoldii* (*rfbA*-Type XIII) were protected from colitis, bodyweight loss, and stool (bleeding) scores. Thus, the addition of *Pdis* (ATCC 8503) to mice exposed to DSS-colitis...
exhibited significantly worse effects compared to mice not receiving any bacteria or mice receiving *Alistipes finegoldii*.

Emphasizing the potential dichotomous role of *Pdis* in health, recent studies conducted with the tumor-prone A/J mouse line have shown a beneficial effect using the same *Pdis* strain (ATCC 8503) and freeze-dried *Pdis* membrane fractions (LPS/O-antigen). In the A/J model, *Pdis* beneficially attenuated toll-like receptor 4 signaling (TLR4; present in myeloid cells: monocytes, macrophages, dendritic cells; and nonimmune cells: endothelial cells, adipocytes) and Akt activation, attenuated tumorigenesis, modulated inflammatory markers and promoted intestinal barrier integrity in azoxymethane-treated mice, with and without a high-fat diet.\(^{10,11}\) Additionally, another study utilizing membrane fractions from *Pdis* ATCC 8503 on BALB/c mice attenuated the severity of colitis induced with 3% DSS and prevented increases in proinflammatory cytokines, indicating that membrane components of *Pdis*, though not specifically live *Pdis* cells, could modulate intestinal inflammation.\(^{12}\)

Of interest, studies in cancer cells found that *Pdis* (ATCC 8503) membrane fractions inhibited *E. coli* derived LPS-induced TLR4 activation in a dose-dependent manner.\(^{10,11}\) To explain the anti-TLR4 signaling effects of *Pdis* membrane fractions when added to *E. coli* LPS, our *rfbA/lpxK* (LPS/O-antigen) analysis suggests the novel hypothesis that the LPS from *Pdis*, which may vary with *rfbA*-Type and copy number, may directly compete and/or displace LPS from other pathogens on the surface of LPS receptors (TLR4) on host cells (*in vitro*, *in vivo*). Therein, it follows that potential *Pdis*-LPS/LPS-receptor interactions could reduce and/or modulate the intensity of cell signaling as illustrated in Figure 6. Thus, it is possible that the pathogenic effects induced by live *Pdis* is through mechanisms other than its LPS/O-antigen membrane fractions. *In vivo*, the anticolitic effect of the *Pdis* membrane fractions was not observed in mice with severe combined immunodeficiency (lacking T cells and B cells), which suggests that anti-colitic effects could be due T-regulatory cell modulation.\(^{13}\) If an anti-inflammatory mechanism by membrane fractions depends on this hypothesized competition of LPS, this feature could also occur in other *Pdis* strains. To date, the aforementioned studies have only examined the strain ATCC 8503 which is the reference for the genus. Based on our studies, strain-to-strain variability should be expected within this mechanism because *rfbA* structure and copy numbers vary across the *P. distasonis* species. Future studies could examine the effect of various membrane fractions across the *Bacteroidetes* phylum to determine the extent to which this feature correlates with *rfbA*-Type(s)/copy number(s), and if it is unique to all the *Pdis* strains within the *rfbA*-Type I cluster.

To help determine in the future whether the presumed pathogenicity and probiotic duality of *Pdis* is due to a stable phenotype or a fluctuation of the phenotype (pathogenic or probiotic), we proposed a classification system based on the genetic variability of the O-antigen synthesis *rfbA* gene. Our classification system, accompanied with controlling for variables that include animal genetics, models for disease induction (e.g. DSS concentration/duration (protocol REFs), azoxymethane dose), diet, microbiota, and the use of antibiotics will be needed to determine the mechanisms that may play a dual role in animal and human health. Given the potential pathogenic effects,\(^{4–9}\) it is important to determine disease mechanisms before considering *Pdis* to be a probiotic species for humans.

Phylogenetics reveal that copy number and structure of the *rfbA* gene in *Pdis* can be used as a classification system (*rfbA*-Typing) of bacterial isolates for future studies. The remarkable conservation of the *rfbA*-Type I sequences in isolates that spanned over 85 years (ATCC 8503, 1933 vs. CavFT-hAR46, 2019)\(^{28,44}\) indicate that some *rfbA* genes are highly conserved within *Pdis*. Of interest, the *rfbA*-Type I cluster was composed mostly of strains that contain only one gene copy. Today it remains uncertain to what extent a greater number of *rfbA* gene copies could influence virulence associated with potentially increased O-antigen production. In *E. coli*, the gene deletion has been shown to eliminate O-antigen production,\(^{51}\) and different types of *rfbA* represent different types of antigens; For example, two gene
products, rffH and rmlA, encode glucose-1-phosphate thymidyltransferase, catalyzing the same enzymatic reaction, yet they are part of different operons and function in different pathways. The clinical downstream effects of rfbA gene variance on the Pdis O-antigen structure remains to be elucidated. Additionally, future studies to validate P. distasonis O-antigens are warranted and cannot be conducted at this time since there is currently no available literature on their physical structures.

The relationship between O-antigen structures and subsequent virulence is longstanding and well-characterized in gram negative Enterobacteriaceae, namely, E. coli,51-53 Shigella sonnei52 and Shigella flexneri,31 where the presence and length of the O-antigen of the LPS play a crucial role in pathogenesis. Compared to other Bacteroidetes and Enterobacteriaceae, the recognition of at least three major superclusters, wherein Parabacteroides shares gene homology with that of Enterobacteriaceae highlights the potential virulence contribution of the Bacteroidetes phylum in animal and human health via the rfbA O-antigen synthesis gene. In context with Bacteroidetes and Enterobacteriaceae (the latter in which rfb genes have been well described51,54,55), conserved clustering of the four distinct Pdis rfbA-Types highlights the not only the uniqueness of the rfbA gene in this species, but the specificity of rfbA-Types I–IV to Pdis. Similarly, most rfbA-Types assigned to other Bacteroidetes were unique to their respective genera, but those assigned to Enterobacteriaceae consistently contained at least two genera per rfbA-Type (except rfbA-Type VI which consists only of two Klebsiella spp.). To help with the characterization of Pdis isolates, we propose RFLP analysis using the MboII restriction enzyme which has been validated in E. coli and Shigella,56,57 however the analysis could be expanded in the future with different enzymes.

In conclusion, this is the first study that provides some insight on the relation of O-antigen with the pathogenesis of Parabacteroides distasonis and that of other Bacteroidetes. The novel framework applied here to Pdis could help differentiate strains based on virulence potential linked to LPS production. Sequences and strains comprising the rfbA-Type I cluster are of significant interest for further investigation, and the primers and laboratory RFLP technique we designed should facilitate this and other studies of the rfbA gene in Pdis. Herein, we showed that rfbA gene variability (insertions/deletions) also occurs in other major genera within the Bacteroidetes phylum (Parabacteroides, Bacteroides, Alistipes, and Prevotella), creating unique ‘rfbA-superclusters’ that share homology with known pathogenic Enterobacteriaceae (Escherichia, Klebsiella, Salmonella, and Shigella), indicating the same potential use for ‘rfbA-Typing’ classification of Bacteroidetes in general. As a novel hypothesis, data indicate that for P. distasonis, applicable to other Bacteroidetes, there could be potential interactions between the rfbA-LPS/membrane fractions of Pdis with that of other bacteria to modulate the intensity and direction of cell signaling and inflammatory pathways in immune cells. Therein, it is possible that the pathogenic effects induced by whole Pdis cells in some strains (e.g. rfbA-Type I) could be through mechanisms other than LPS/O-antigen membrane fractions and rfbA-Type variation.

Acknowledgments

We acknowledge the Biorepository Core of the NIH Silvio O. Conte Cleveland Digestive Disease Research Core Center (P30DK097948; PD Fabio Cominelli, M.D. Ph.D.) at Case Western Reserve University for clinical advice and provision of deidentified surgical specimens for the above-mentioned R21 project. V.S. joined A.R.-P. laboratory after completing her Ph.D. program in 2021 at CSIR-Institute of Microbial Technology and Jawaharlal Nehru University, India.

Data availability

The genome sequencing data generated for the strains CavFT-hAR46 (BioSample SAMN11642307, PMID: 31488526), CavFT-hAR46_b (this study, same isolate as CavFT-hAR46, re-sequenced), and CavFT-hAR56 (this study) are available in GenBank within the BioProject number PRJNA542869.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This project was supported by NIH grant R21 DK118373 to A. R.-P., entitled “Identification of pathogenic bacteria in Crohn’s disease.” Partial support earlier originated for A.R.-P. from a career development award [310472] and a student
research fellowship award [644955] from the Crohn’s & Colitis Foundation. Partial support for N.C.B earlier originated from The Digestive Health Research Institute’s T35, “The Case Medical Student Summer Research Program (MSSRPP)” [T35DK111373, Fabio Cominelli, M.D. Ph.D, Theresa Pizarro Ph.D.] at Case Western Reserve University.

References

1. Schmidt TM. Encyclopedia of Microbiology. San Diego 1(USA): Academic Press; 2019.

2. Ezeji JC, Sarikonda DK, Hopperton A, Erkkila HL, Cohen DE, Martinez SP, Cominelli F, Kuwahara T, Dichosa AER, Good CE, et al. *Parabacteroides distasonis* intriguing aero-tolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. Gut Micro. 2021;13:1. DOI:10.1080/19490976.2021.1922241

3. Sakamoto M, Benno Y. Reclassification of bacteroides distasonis, bacteroides goldsteinii and bacteroides merdae as *parabacteroides distasonis* gen. nov., comb. nov. Parabacteroides Goldsteinii Comb Nov and Parabacteroides Merdae Comb Nov. Int J Sys Evolut Microbiol. 2006;56(7):1599-1605.

4. Manukyan GP, Ghazaryan KA, Ktsyan ZA, Khachatryan ZA, Arakelova KA, Kelly D, Grant G, Aminov RI. Elevated systemic antibodies towards commensal gut microbiota in autoimmune condition. PLoS One. 2008;3(9):e3172. doi:10.1371/journal.pone.0003172. Epub 2008/09/10. PubMed PMID: 18779861; PubMed Central PMCID: PMC2525839.

5. Dziarski R, Park SY, Kashyap DR, Dowd SE, Gupta D. Pglyp-regulated gut microflora prevotella falsenii, *parabacteroides distasonis* and bacteroides eggerthii enhance and alstipes finegoldii attenuates colitis in mice. PLoS One. 2016;11(1):e0146162. doi:10.1371/journal.pone.0146162. Epub 2016/01/04. PubMed PMID: 26727498; PubMed Central PMCID: PMC4699708.

6. Weimer SB, Matthews MR, Caruso DM, Foster KN. Retropertitoneal necrotizing fascitis from fourier’s gangrene in an immunocompromised patient. Case Rep Surg. 2017;2017:5290793. doi:10.1155/2017/5290793. Epub 2018/01/16. PubMed PMID: 29333313; PubMed Central PMCID: PMC5733135.

7. Blacher E, Bashirades S, Shapiro H, Rothschild D, Mor U, Dori-Bachash M, Kleimeyer C, Moresi C, Harnik Y, Zur M, Zabari M. Potential roles of gut microbiome and metabolites in modulating ALS in mice. Nature. 2019;572(7770):474–480. doi:10.1038/s41586-019-1443-5.

8. Moreno-Arrones OM, Serrano-Villar S, Perez-Brocal V, Saceda-Corrado D, Morales-Rayos C, Rodrigues-Barata R, Moya A, Jaen-Olasolo P, Vano-Galvan S. Analysis of the gut microbiota in alopecia areata: identification of bacterial biomarkers. J Eur Acad Dermatol Venereol. 2020;34(2):400–405. doi:10.1111/jdv.15885. Epub 2019/08/17. PubMed PMID: 31419351.

9. Zhou C, Zhao H, Xiao XY, Guo RJ, Wang Q, Chen H, Zhao LD, Zhang CC, Jiao YH, Ju YM, YangHX. Metagenomic profiling of the pro-inflammatory gut microbiota in ankylosing spondylitis. J Autoimmun. 2020;107:102360. doi:10.1016/j.jaut.2019.102360.

10. Koh GY, Kane AV, Wu X, Crott JW. *Parabacteroides distasonis* attenuates tumorigenesis, modulates inflammatory markers and promotes intestinal barrier integrity in azoxymethane-treated A/J mice. Carcinogenesis. 2020;41(7):909–917. doi:10.1093/carcin/bgaa018. Epub 2020/03/03. PubMed PMID: 32115637.

11. Koh GY, Kane A, Lee K, Xu Q, Wu X, Roper J, Mason JB, Crott JW. *Parabacteroides distasonis* attenuates toll-like receptor 4 signaling and Akt activation and blocks colon tumor formation in high-fat diet-fed azoxymethane-treated mice. Int J Cancer. 2018;143(7):1797–1805. doi:10.1002/ijc.31559. Epub 2018/04/27. PubMed PMID: 29696632.

12. Cuffaro B, Assouhoun AL, Boutillier D, Säkeniková L, Desramaut J, Boudebouze S, Salmé–Desnoulez S, Hrdý J, Waligora-Dupriet AJ, Maguin E, et al. *In vitro* characterization of gut microbiota-derived commensal strains: selection of *Parabacteroides distasonis* strains alleviating TNBS-induced colitis in mice. Cells. 2020;9(9). doi:10.3390/cells9092104. Epub 2020/09/20. PubMed PMID: 32947881; PubMed Central PMCID: PMC7565435.

13. Kverka M, Zakostelska Z, Klimesova K, Sokol D, Hucdovic T, Hrnícir T, Rossmann P, Mrazek J, Kopecny J, Verdu EF, et al. Oral administration of *parabacteroides distasonis* antigens attenuates experimental murine colitis through modulation of immunity and microbiota composition. Clin Exp Immunol. 2011;163(2):250–259. doi:10.1111/j.1365-2249.2010.04286.x. Epub 2010/11/23. PubMed PMID: 21087444; PubMed Central PMCID: PMC3043316.

14. Gu S, Zaidi S, Hassan MI, Mohammad T, Malta TM, Noushmehr H, Nguyen B, Crandall KA, Srivastav J, Obias V, Lin P. Mutated CEACAMs disrupt transforming growth factor beta signaling and alter the intestinal microbiome to promote colorectal carcinogenesis. Gastroenterology. 2020;158(1):238–252. doi:10.1053/j.gastro.2019.09.023. PubMed PMID: 31585122.

15. Wang K, Liao M, Zhou N, Bao L, Ma K, Zheng Z, Wang Y, Liu C, Wang W, Wang J, Liu SJ. *Parabacteroides distasonis* alleviates obesity and metabolic dysfunctions via production of succinate and secondary Bile Acids. Cell Rep. 2019;26(1):222–35.e5. doi:10.1016/j.celrep.2018.12.028. Epub 2019/01/04. PubMed PMID: 30605678.

16. Hasain Z, Mokhtar NM, Kamaruddin NA, Mohamed Ismail NA, Razalli NH, Gnanaou JV, Raja Ali RA. Gut microbiota and gestational diabetes mellitus: a review of
host-gut microbiota interactions and their therapeutic potential. Front Cell Infect Microbiol. 2020;10:188. doi: 10.3389/fcimb.2020.00188. Epub 2020/06/06. PubMed PMID: 32500037; PubMed Central PMCID: PMC7243459.

17. Cekanaviciute E, Yoo BB, Runia TF, Debelius JW, Singh S, Nelson CA, Kanner R, Bencosme Y, Lee YK, Hauser S & Crabtree-Hartman E. Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. Proceedings of the National Academy of Sciences. 2017 Oct 3;114 (40):10713-8.

18. Nataro JP, Kaper JB. Diarrheagenic Escherichia coli. Clin Microbiol Rev. 1998;11(1):142–201. doi:10.1128/cmr.11.1.142. Epub 1998/02/11. PubMed PMID: 9457432; PubMed Central PMCID: PMC121379.

19. Mueller M, Tainter C. Escherichia Coli. StaphPears [Internet]. Treasure Island (FL): StaphPears Publishing. 2021.

20. Andino A, Hanning I. Salmonella enterica: survival, colonization, and virulence differences among serovars. ScientificWorldJournal. 2015;2015:520179. doi:10.1155/2015/520179. Epub 2015/02/11. PubMed PMID: 25664339; PubMed Central PMCID: PMC4310208.

21. Lerouge I, Vanderleyden J. O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. FEMS Microbiol Rev. 2002;26(1):17–47. doi:10.1111/j.1574-6976.2002.tb00597.x.

22. Kauffmann F. The serology of the coli group. The Journal of Immunology. 1947;57:71–100.

23. Orskov I, Orskov F, Jann B, Jann K. Serology, chemistry, and genetics of O and K antigens of Escherichia coli. Bacteriol Rev. 1977;41:667.

24. Orskov F, Orskov I. Serotyping of Escherichia coli. In: Bergan T, editor. Methods in Microbiology. Vol. 14. London: Academic Press; 1984. p. 43–112.

25. Wu Y, Lau HK, Lee T, Lau DK, Payne J. In Silico serotyping based on whole-genome sequencing improves the accuracy of shigella identification. Appl Environ Microbiol. 2019;85(7). doi: 10.1128/aem.00165-19. Epub 2019/02/03. PubMed PMID: 30709819; PubMed Central PMCID: PMC6585509.

26. Joensen Katrine G, Tetzchner Anna MM, Iiguchi A, Aarestrup Frank M, Scheutz F, Carroll KC. Rapid and easy in Silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J Clin Microbiol. 2015;53(8):2410–2426. doi:10.1128/JCM.00008-15.

27. Zuccotti S, Zanardi D, Rosano C, Sturla L, Tonetti M, Bolognesi M. Kinetic and crystallographic analyses support a sequential-ordered bi bi catalytic mechanism for Escherichia coli glucose-1-phosphate thymidylate synthase. J Mol Biol. 2001;313(4):831–843. doi:10.1006/jmbi.2001.5073. PubMed PMID: 11697907.

28. Yang F, Kumar A, Davenport KW, Kelliher JM, Ezeji JC, Good CE, Jacobs MR, Conger M, West G, Fiocchi C, Cominelli F, Dichosa AEK, Rodriguez-Palacios A. Complete genome sequence of a parabacteroides distasonis strain (CavFT hAR46) isolated from a gut wall-cavitating microlesion in a patient with severe crohn’s disease. Microbiol Resour Announc. 2019;8(36). doi:10.1128/MRA.00585-19. Epub 2019/09/05. PubMed PMID: 31488526; PubMed Central PMCID: PMC6728636.

29. Sivaraman J, Sauvé V, Matte A, Cylger M. Crystal structure of Escherichia coli glucose-1-phosphate thymidylate synthase (RfIh) complexed with dTTP and Mg2+. Journal of Biological Chemistry. 2002 Nov 15;277(46):44214–9.

30. Wang J, Jiao H, Zhang X, Zhang Y, Sun N, Yang Y, Wei Y, Hu B, Guo X. Two Enteropathogenic Escherichia coli Strains Representing Novel Serotypes and Investigation of Their Roles in Adhesion. Journal of Microbiology and Biotechnology. 2021;31(9):1191-9.

31. Morona R, Daniels C & Van Den Bosch L. Genetic modulation of Shigella flexneri 2a lipopolysaccharide O antigen modal chain length reveals that it has been optimized for virulence. Microbiology. 2003 Apr 1;149(4):925–39.

32. Caboni M, Pedron T, Rossi O, Goulding D, Pickard D, Citiulo F, MacLennan CA, Dougan G, Thomson NR, Saul A, Sansonetti PJ. An O antigen capsule modulates bacterial pathogenesis in Shigella sonnei. PLoS pathogens. 2015 Mar 20;11(3):e1004749.

33. Ghodsi M, Liu B, Pop M. DNAclust: accurate and efficient clustering of phylogenetic marker genes. BMC Bioinform. 2011;12(1):271. doi:10.1186/1471-2105-12-271. Epub 2011/07/02. PubMed PMID: 21718538; PubMed Central PMCID: PMC3213679.

34. Unterberger A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3plus, an enhanced web interface to primer3. Nucleic Acids Res. 2007;35(Web Server):W71–4. doi:10.1093/nar/gkm306. Epub 2007/05/09. PubMed PMID: 17485472; PubMed Central PMCID: PMC1933133.

35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215 (3):403–410. doi:10.1016/s0022-2836(05)80360-2. Epub 1990/08/05. PubMed PMID: 2231712.

36. Coimbra RS, Grimont F, Lenormand P, Burguière P, Beutin L, Grimont PAD. Identification of Escherichia coli O-serogroups by restriction of the amplified O-antigen gene cluster (rfb-RFLP). Res Microbiol. 2000;151(8):639–654. doi:10.1016/S0922-2508(00)00134-0.

37. Coimbra RS, Grimont F, Grimont PAD. Identification of shigella serotypes by restriction of amplified O-antigen gene cluster. Res Microbiol. 1999;150(8):543–553. doi:10.1016/S0922-2508(99)00103-5.

38. Munro HN. Mammalian protein metabolism. Elsevier; 2012. Ebook.
