Manuscript number:  DMD-AR-2022-000898R1

Manuscript title: “Gut microbiome-wide search for bacterial azoreductases reveals potentially uncharacterized azoreductases encoded in the human gut microbiome”

Authors: Domenick J. Braccia, Glory Minabou Ndjite, Ashley Weiss, Sophia Levy, Stephenie Abeysinghe, Xiaofang Jiang, Mihai Pop, Brantley Hall

Author affiliations:
DB, MP, BH: Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland, USA
GMN, AW, SL, SA, BH: Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland, USA
XJ: National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA
MP: Department of Computer Science, University of Maryland, College Park, Maryland, USA
RUNNING TITLE PAGE

Running title: Prediction of uncharacterized azo reduction by the gut microbiome

Corresponding Author:
- Name: Dr. Brantley Hall
- Address: 3126 Campus Drive College Park 20742
- Telephone: 301-405-7481
- Email: brantley@umd.edu
- Fax number: N/A

Number of:
- Text pages: 17
- Tables: 2 primary, 5 supplementary
- Figures: 6
- References: 91

Number of words in:
- Abstract: 236
- Introduction: 750 (not including references)
- Discussion: 1423

Nonstandard abbreviations used:
inflammatory bowel disease (IBD)
colorectal cancer (CRC)
Hidden Markov Model (HMM)
Unified Human Gastrointestinal Genomes (UHGG)
5-aminosalicylic acid (5-ASA)
Dihydroquinolone (DHQ)
flavin mononucleotide (FMN)
flavin adenine dinucleotide (FAD)
ulcerative colitis (UC)
Crohn's disease (CD)
Integrative Human Microbiome Project (HMP2)
Prospective Registry of IBD Patients at MGH (PRISM)
Health Professionals Followup Study (HPFS)
Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST)
sulfasalazine reducing (SR)
non-sulfasalazine reducing (nonSR)
Brain Heart Infusion (BHI)
Yeast Casitone Fatty Acids (YCFA)
Analysis of variance (ANOVA)
Genome Taxonomy Database (GTDB)
Genome Taxonomy Database tool kit (GTDB-tk)
counts per million (CPM)
transcripts per million (TPM)
False Discover Rate (FDR)

Recommended section assignments:
- Gastrointestinal, Hepatic, Pulmonary, and Renal
- Metabolism, Transport, and Pharmacogenomics
ABSTRACT

The human gut is home to trillions of microorganisms that are responsible for the modification of many orally administered drugs, leading to a wide range of therapeutic outcomes. Prodrugs bearing an azo bond are designed to treat inflammatory bowel disease (IBD) and colorectal cancer (CRC) via microbial azo reduction, allowing for topical application of therapeutic moieties to the diseased tissue in the intestines. Despite the inextricable link between microbial azo reduction and the efficacy of azo prodrugs, the prevalence, abundance, and distribution of azoreductases have not been systematically examined across the gut microbiome. Here, we curated and clustered amino acid sequences of experimentally confirmed bacterial azoreductases and conducted a Hidden Markov Model (HMM)-driven homolog search for these enzymes across 4,644 genome sequences present in the representative Unified Human Gastrointestinal Genomes (UHGG) collection. We identified 1,958 putative azo-reducing species, corroborating previous findings that azo reduction appears to be a ubiquitous function of the gut microbiome. However, through a systematic comparison of predicted and confirmed azo-reducing strains, we hypothesize the presence of uncharacterized azoreductases in 25 prominent strains of the human gut microbiome. Finally, we confirmed the azo reduction of Acid Orange 7 by multiple strains of *Fusobacterium nucleatum*, *Bacteroides fragilis*, and *Clostridium clostridioforme*. Together, these results suggest the presence and activity of many uncharacterized azoreductases in the human gut microbiome and motivate future studies aimed at characterizing azoreductase genes in prominent members of the human gut microbiome.

Significance Statement: In this work, we systematically examined the prevalence, abundance, and distribution of azoreductases across the healthy and IBD human gut microbiome revealing potentially uncharacterized azoreductase genes. We also confirmed the reduction of Acid Orange 7 by strains of *Fusobacterium nucleatum*, *Bacteroides fragilis*, and *Clostridium clostridioforme*.
INTRODUCTION

Orally administered drugs are an attractive, non-invasive mode of delivery of pharmaceuticals to the intestines. The human gut microbiome plays an important role in drug metabolism (Spanogiannopoulos et al. 2016) and is capable of activating (Peppercorn and Goldman 1972; Morrison, Wright, and John 2012; Sousa et al. 2014), inactivating (Kalman 1978; Dobkin et al. 1983; Haiser et al. 2013), and even toxifying (Wallace et al. 2010) pharmaceutical drugs. Prodrugs containing an azo bond actually require bacterial azoreductase activity to release biologically active compounds (Peppercorn and Goldman 1972). For conditions such as Inflammatory Bowel Disease (IBD) and colorectal cancer (CRC), bacterial azoreductases have been utilized to deliver therapeutics such as 5-aminosalicylic acid (5-ASA), prednisolone (Ruiz et al. 2011), and celecoxib (Marquez Ruiz et al. 2011) topically to diseased intestinal tissues (Figure 1). Following oral administration of sulfasalazine, bacterial azoreductases in the gut reduce azo bonds, liberating 5-ASA and allowing it to confer its anti-inflammatory properties (Mahida et al. 1991; Weber et al. 2000; Rachmilewitz et al. 1992) topically on inflamed intestinal tissue. Direct oral administration of 5-ASA is non-optimal because the majority of the drug is absorbed in the small intestine and is sent through systemic circulation (Peppercorn and Goldman 1973; Tozaki et al. 2002; Friend 2005; Perrotta et al. 2015; Foppoli et al. 2019). Other examples of azo-bonded prodrugs are OPN501 and celecoxib-5-ASA. OPN501 is made of up prednisolone, 5-ASA, and an inert cyclization product, Dihydroquinolone (DHQ). Upon azo reduction, 5-ASA is released and is able to act topically upon the target tissue. Following a spontaneous cyclization reaction, prednisolone and DHQ are released where prednisolone can act upon the target tissue (Ruiz et al. 2011). Celecoxib-5-ASA is made up of celecoxib, 5-ASA, and DHQ, which are all released upon azo reduction and cyclization in a similar mechanistic fashion to OPN501 (Marquez Ruiz et al. 2011).
A recent review article by Suzuki 2019 collected and curated experimentally confirmed azoreductases and described their preferred flavin cofactors and electron donors (H. Suzuki 2019). They found that bacterial azoreductases qualitatively cluster into four main clades, which harbor a preference for either flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) as the flavin cofactor, and either NADH or NADPH as the preferred electron donor. Clades I, II, and III are flavoproteins whereas Clade IV proteins are flavin-free. Clade I azoreductases prefer NADPH as the electron donor, Clade II prefers NADH as its electron donor, and Clades III and IV generally use both. Of the 37 enzymes examined by Suzuki 2019, eight enzymes formed no distinct phylogenetic clade and featured differences in primary sequence length, flavin cofactor, and preferred electron donor. In addition to their relevance in drug delivery and efficacy, azoreductases are involved in nitroreduction (Brown 1981; Rafii, Franklin, and Cerniglia 1990; Rafii and Cerniglia 1995; Liu et al. 2007; Mercier et al. 2013; Chalansonnet et al. 2017), quinone oxidoreduction (Liu et al. 2008; Leelakriangsak et al. 2008; Ryan, Wang, et al. 2010; Ryan et al. 2014), and azo dye reduction. Azo dyes such as Allura Red and Brilliant Black are commonly used in the food and textile industries, and waste from their production and usage pollutes the environment. This has led to a plethora of manuscripts characterizing the activity of azoreductases across the bacterial kingdom (Misal et al. 2014; Cerniglia et al. 1982; Feng, Cerniglia, and Chen 2012; T. Zimmermann, Kulla, and Leisinger 1982; Nakanishi et al. 2001; Y. Suzuki et al. 2001; Silke Blümel, Knackmuss, and Stolz 2002; S. Blümel and Stolz 2003; H. Chen, Wang, and Cerniglia 2004; Nachiyar and Rajakumar 2005; H. Chen, Hopper, and Cerniglia 2005; Ooi et al. 2007; Matsumoto et al. 2010; Misal et al. 2011; Gonçalves et al. 2013; Lang et al. 2013; Zhang, Ng, and Chang 2016; Eslami, Amoozegar, and Asad 2016), many of which exhibit non-negligible sequence similarity to gut microbial azoreductases (H. Suzuki 2019).

There is a growing body of literature suggesting that azo reduction is a ubiquitous function of the human gut microbiome (Javdan et al. 2020) with many prominent bacterial
strains showing significant reduction of sulfasalazine *in vitro* (M. Zimmermann et al. 2019).

Javdan et al. 2020 showed that among 20 different individuals, sulfasalazine reduction was one of the only ubiquitous functions of the gut microbiome. Zimmermann et al. 2019 tested the reduction of sulfasalazine by 76 prominent strains of the gut microbiome and reported a significant (FDR adjusted p-value < 0.05) reduction of sulfasalazine by 62 of these strains. Interestingly, some of these experimentally confirmed azo-reducing strains reported by Zimmermann et al. 2019 have no prior evidence of azo-reduction, and thus, may encode novel or uncharacterized azoreductase genes (Figure 1C). Identification of known azoreductases in newly reported sulfasalazine reducing species can help narrow down strains to target for identification of uncharacterized azoreductases.

The prevalence, abundance, expression, and distribution of azoreductase enzymes in the human gut microbiome have implications for the efficacy of existing prodrugs mentioned above, as well as for the development of future azo prodrugs. While azoreductases have been identified and characterized in many gut bacteria (Supp. Table 1), the distribution of azoreductases has not been systematically explored across currently gut bacterial reference genomes. To address this gap, we conducted a homolog search of known azoreductases across the Unified Human Gastrointestinal Genomes (UHGG) collection (Almeida et al. 2020) to identify putative azoreductases and azo-reducing species in the human gut microbiome. We then assessed the relative abundance and expression of known azoreductases in healthy, ulcerative colitis (UC), and Crohn’s disease (CD) participants of the Integrative Human Microbiome Project (HMP2) (Proctor and Huttenhower 2019), the Prospective Registry of IBD Patients at MGH (PRISM) (Franzosa et al. 2019), and the Health Professionals Followup Study (HPFS) (Abu-Ali et al. 2018). Finally, we tested the *in vitro* azo reduction of Acid Orange 7 by three strains of *Fusobacterium nucleatum* along with two strains of *Bacteroides fragilis* and two strains of *Clostridium clostridioforme*.
MATERIALS AND METHODS

Description of publicly available shotgun metagenomic sequencing data

Shotgun metagenomic sequencing data obtained from the Integrative Human Microbiome Project (HMP2) (N(samples) = 703, N(individuals) = 103) (Proctor and Huttenhower 2019), the Prospective Registry of IBD Patients at MGH (PRISM) (N(samples) = 218, N(individuals) = 218) (Franzosa et al. 2019), and the Health Professionals Follow-up Study (HPFS) (N(samples) = 220, N(individuals) = 220) (Abu-Ali et al. 2018) are used throughout this work. Note that all samples referred to throughout this work are human stool samples.

Curation of HMMs representing azoreductase enzymes

We searched the literature for known and experimentally validated species of bacteria which have azoreductase activity. The list of gene sequences collected, along with the relevant metadata (organism, functional annotation, length, etc.) is available in Supp. Table 1.

Preliminary evidence for azoreductase gene sequence clustering is shown in Suzuki 2019 where sequences were aligned and phylogenetically compared. Next, after collecting 40 sequences of experimentally validated azoreductases, we generated a sequence similarity network (SSN) using EFI-EST (Gerlt et al. 2015) at a 35% amino acid sequence identity threshold for identifying similar clusters of azoreductase genes. This threshold corroborates the preliminary evidence for a diversity of azoreductase sequences put forth by Suzuki 2019. With the exception of Clade IVb sequences, which reached 31% sequence identity, all other clusters of genes had at least 35% sequence identity. The groups shown in Figure 2 were pressed into profile Hidden Markov Models (HMMs) using HMMER version 3.1b2 (Finn et al. 2015). Other genes that did not fall into the Clade I - Clade IV clusters (arsH, yieF, mdaB, azo1, azoR, etc.), were pressed into singular HMMs and included in the homolog search.
**Search for azoreductase genes across human gut microbial genomes**

We searched HMMs of known, experimentally validated azoreductases across 4,644 non-redundant genomes contained in the UHGG collection (Almeida et al. 2020) using HMMER v3.1b2 (Finn et al. 2015). Alignments to queried HMMs with E-value <1x10⁻¹⁰ and 60% coverage of the query sequence were labeled as putative azoreductase gene sequences. Putative azo-reducing bacterial species with experimentally confirmed azoreductase activity (Supp. Table 2) were then labeled as “known” azo-reducing species and classified separately from the putative azo-reducing species. Putative azo-reducing species across the bacterial taxonomy were visualized using the iTOL web interface (Letunic and Bork 2019) and prominent phyla of the gut microbiome were subsetted and presented in Figure 3.

**Relative abundance and azoreductase gene abundance and expression estimation**

Raw sequencing reads for samples from HMP2 (Proctor and Huttenhower 2019), PRISM (Franzosa et al. 2019), and HPFS (Abu-Ali et al. 2018), were downloaded and extracted with NCBI’s SRA toolkit v2.10.9 (“SRA-Tools - NCBI” n.d.). Quality control and adapter trimming of the fastq sequence files were done with the Trim Galore wrapper v0.6.6 (“Babraham Bioinformatics - Trim Galore!” n.d.). To remove potential human contaminants, quality trimmed reads were screened against the human genome (hg19) with Bowtie2 v2.4.2 (Langmead and Salzberg 2012). Putative azoreductase sequences were extracted from UHGG genomes via custom shell and python scripts. Putative azoreductase gene sequences (HMP2, PRISM) and expression levels (HPFS) were quantified using salmon v1.4.0 (Patro et al. 2017) and were normalized and aggregated in R v4.1.1 and were subsequently visualized using the R package ggplot2 (Wickham 2011) (Figure 4). Taxonomy profiling of the cleaned metagenomic reads from HMP2 samples was performed using Kraken 2 v2.0.8-beta (Wood, Lu, and Langmead 2019) to estimate the relative abundance of bacterial species present in each dataset. These relative abundances were then processed in R v4.1.1 and plotted using ggplot2 (Figure 5).
Computational and bioinformatic procedures are open source and are provided at

https://github.com/dombraccia/Azoreductases.

**Statistical analysis of relative metagenomic sequence data from HMP2 and PRISM datasets**

Statistical analyses described in Figure 4 were performed in R 4.1.1 with the wilcoxon.test method using default parameterization. Next, we compared the relative abundances of known and putative azo-reducing species for HMP2 subjects with 20 or more stool samples taken over the course of the study (Figure 5). A linear mixed effects model analysis of variance (ANOVA) was performed on this subset of HMP2 data to determine any statistically significant differences in relative abundance values across nonIBD, UC, and CD subjects. The R package lme4 (Bates et al. 2014) was used to fit the model to the data and the package lmerTest (Kuznetsova, Brockhoff, and Christensen 2017) was used to perform the ANOVA on the model. All statistical analyses were performed in R version 4.1.1 and are provided at

https://github.com/dombraccia/Azoreductases.

**Acid Orange 7 azo reduction assay**

Biological triplicates were grown in a ten ml tube containing 10 ml of Yeast Casitone Fatty Acids (YCFA) broth for *Bacteroides fragilis* and *Clostridium clostridioforme* strains and 10ml of Brain Heart Infusion (BHI) broth for *Fusobacterium nucleatum* strains. Each tube was inoculated with 10 μL of bacteria from glycerol stocks. The final concentrations of each substrate in the bacterial cultures were: 50 μg/ml of FMN, 50 μg/ml of NADH, and 50 μmol/ml of Acid Orange 7. The bacterial cultures were left to grow in an anaerobic chamber for 72 hours, and Acid Orange 7 decolorization was measured once every 24 hours since inoculation. The decolorization of Acid Orange 7 was measured by aliquoting triplicates of 200 μL media aliquants to a 96-well plate for each bacterial culture, and absorbance of 550 nm light was measured using a
spectrophotometer. The raw absorbance values for each biological and technical replicate are reported in Supp. Table 5.
RESULTS

Primary amino acid sequences of bacterial azoreductases group by mechanistic preferences

To begin identifying putative azo-reducing species of the gut microbiome, we searched the literature for experimentally verified azoreductase enzymes, collected amino acid sequences and metadata for these azoreductases (Supp. Table 1), and compared their primary sequences using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (Gerlt et al. 2015). The resulting sequence similarity network captured mechanistic preferences such as flavin dependence and electron donor preference for each azoreductase reported by Suzuki 2019. We saw near complete concordance between the Clade I-IV azoreductases and the sequence similarity clusters at a 35% amino acid identity edge threshold, with the exception of Clade IV, which was split into two separate clusters (Figure 2). The gene families labeled arsH, mdaB, yieF, and azo1 clustered separately from Clade I-IV azoreductases, and we consider each of these clusters as separate sub-families of azoreductases. Multiple sequence alignments were generated for each cluster shown in Figure 2 (with the exception of the “Other Azoreductases” group) using MUSCLE v3.8.425 (Edgar 2004). Hidden Markov Models (HMMs) were then trained on the multiple sequence alignments from the previous step using the HMMER v3.1b2 method hmmpress (Finn et al. 2015) and were queried against the UHGG collection (Almeida et al. 2020).

Homolog search for azoreductases supports evidence for ubiquity of azo reduction by the human gut microbiome

We searched for homologs of azoreductases across 4,644 representative genomes in the UHGG collection (Almeida et al. 2020) using HMMs generated from sequences of experimentally validated azoreductase enzymes (Figure 3). This collection contains 204,938 genome sequences of bacteria known to inhabit the human gut, of which 4,644 are included in
the representative collection (Supp. Table 4). For the remainder of this work, we refer to species receiving statistically significant (E-value < 1x10^{-10}) hits to azoreductase genes as “putative azo-reducing species” or “putative azo-reducing bacteria". Of the 4,644 genomes in the UHGG collection, there are 1,443 (31.1%) with one putative azoreductase gene, 343 (7.4%) with two or more putative azoreductases, and 372 (8.0%) with three or more putative azoreductases, indicating the extensive potential of the gut microbiome to reduce azo bonds.

Most notably, 364 genomes contain hits to the Clade I profile, 452 contain hits to the Clade II profile, 793 genomes contain hits to the Clade III profile, 568 contain hits to the Clade IVa profile, 410 contain hits to the Clade IVb profile, 285 contain hits to the mdaB profile, and 477 contain hits to the yieF profile. Prominent phyla of the gut microbiome such as Proteobacteria and Firmicutes appear particularly rich with Clade I, Clade II, Clade III, Clade IVab, and FAD utilizing azoreductases (purple columns in Figure 3).

**Systematic evaluation of predicted azo-reducing species**

We next sought to evaluate the results of our azoreductase homolog search with recent findings by Zimmermann et al. 2019 regarding sulfasalazine reduction. Zimmermann et al. 2019 tested the degradation of sulfasalazine by 76 prominent gut bacterial strains, 67 of which had corresponding reference genomes present in the UHGG collection. This provided an excellent source of data to compare our bioinformatic predictions against. We determined the sulfasalazine reducing status as either sulfasalazine reducing (SR) or non-sulfasalazine reducing (nonSR) for each of the 67 strains based on the significant (FDR adjusted p-value < 0.05) reduction of sulfasalazine in vitro reported by Zimmermann et al. 2019 (Table 1). We also determined the predicted reducing status as either a predicted reducer (PR) or a non-predicted reducer (nonPR) for each strain based on the presence of a putative azoreductase identified from the homolog search step. For each strain, the sulfasalazine reducing status and predicted reducing status were compared to systematically validate the results of the azoreductase
homolog search (Table 1, columns 7-9). We correctly predicted the sulfasalazine reducing status for 47.8% (32/67) of strains and we incorrectly predicted the sulfasalazine reducing status for 52.2% (35/67) of strains (Table 2). The vast majority (77.1%, 27/35) of incorrectly predicted strains are false negatives, meaning, the strain does reduce sulfasalazine in vitro, but we did not identify an azoreductase in the homolog search step (Table 2). Interestingly, the majority of false positives (75%, 6/8) are members of the Proteobacteria phylum which we previously noted to be particularly rich in azoreductase gene sequences (Figure 3).

Exploratory analysis of azoreductase abundance and expression levels in the human gut microbiome

After identifying putative azo-reducing species of the human gut microbiome, we next sought to examine the abundance and expression of putative azoreductases using publicly available metagenomic and metatranscriptomic datasets. We used shotgun metagenomic sequence data from the Human Microbiome Project 2 (HMP2) (Proctor and Huttenhower 2019) and the Prospective registry of IBD patients at MGH (PRISM) (Franzosa et al. 2019) to quantify azoreductase gene abundance. We also used high throughput metatranscriptomic sequence data from the Health Professionals Follow-up Study (HPFS) (Abu-Ali et al. 2018) to quantify the expression of azoreductases by the human gut microbiota (Figure 4A-C). Briefly, raw genomic and transcriptomic reads were filtered and processed using fastp (S. Chen et al. 2018), and azoreductases were quantified using salmon v1.4.0 (Patro et al. 2017). Please see the Materials and Methods section for more details on the computational and statistical procedures used.

Significant differences in azoreductase gene abundances between disease conditions are displayed in Figure 4A-B with asterisks. We find that Clade I, Clade II, Clade III, Clade IVa, Clade IVb, mdaB, and yieF genes are considerably higher in abundance than ferB, azoR1_cperf, arsH, azoR_ropacus, azo1, and azoR1_llentus within all three disease conditions for both HMP2 (all FDR adjusted p < 2.2x10^{-16}) and PRISM (all FDR adjusted p < 2.6x10^{-7})
Clade IVa is higher in abundance than all other azoreductases across healthy, UC and CD cohorts from HMP2 (Figure 4A), but the same statistically significant difference was not observed in the PRISM study (Figure 4B).

The expression of Clade I, Clade II, Clade III, Clade IVa, mdaB, and yieF azoreductases are significantly higher than those of Clade IVb, ferB, azoR1_cperf, arsH, azoR_ropacus, azo1, and azoR1_llentus azoreductases in healthy individuals (minimum FDR adjusted p-value < 1x10^-7 between mdaB and azoR1_cperf) (Figure 4C). While Clade IVb abundance levels are comparable to those of Clades I, II, III, and IVa, the expression levels of Clade IVb azoreductases are significantly lower in vivo than the expression Clade I, II, III, and IVa azoreductases (all FDR adjusted p < 2.2x10^-16).

The relative abundance of putative azo-reducing species fluctuates over time
We next sought to examine whether relative abundance levels of combined known and putative azo-reducing species are stable or fluctuate over time. The HMP2 dataset provides a unique opportunity to examine the stability of individuals’ gut microbiomes over time, as there are 18 individuals across healthy, UC, and CD cohorts with at least 20 stool samples taken once every two weeks over a six-month period. In order to examine the stability of azo reduction in the human gut, we compared the relative abundance of known and putative azo-reducing species from these participants (Figure 5). The median relative abundance of combined azo-reducing species ranges from 20.3 ± 3.58% to 33.9 ± 19.2% for nonIBD, 21.7 ± 7.6% to 49.0 ± 15.0% for UC, and 34.9 ± 6.39 to 62.3 ± 18.8 for CD subjects. Using linear mixed effects model ANOVA, we found that combined azo-reducing species are significantly more abundant in CD subjects than in nonIBD subjects (p = 0.002) and are not significantly more abundant in UC subjects than in nonIBD subjects (p = 0.064) (Figure 5A). Note that Figure 5B shows the same relative
abundance values displayed in Figure 5A, but over the course of the study, from collection 1 to collection 24.

**Multiple strains of *Fusobacterium nucleatum*, *Bacteroides fragilis*, and *Clostridium clostridioforme* reduce Acid Orange 7 *in vitro***

Finally, we sought to test the azo reduction of Acid Orange 7 by three strains of the health-relevant (Castellarin et al. 2012; Kostic et al. 2012; Bashir et al. 2015; Abed et al. 2016) microbe, *Fusobacterium nucleatum*. As well, we tested the azo reduction of Acid Orange 7 by two positive control species, *Bacteroides fragilis*, and *Clostridium clostridioforme*. Acid Orange 7 is an azo-bonded dye commonly used in the food and textile industries (Bay et al. 2014), and the decolorization of azo-bonded dyes is commonly used to test azo reduction by bacteria *in vitro* (Feng, Cerniglia, and Chen 2012). *F. nucleatum* CTI-06, *F. nucleatum* subsp. animalis D11, and *F. nucleatum* subsp. polymorphum were grown in Brain Heart Infusion (BHI) media, and Acid Orange 7 was added to the culture after four days of growth (Materials and Methods). We also tested the azo reduction of Acid Orange 7 by the known azo-reducing species *Bacteroides fragilis* and *Clostridium clostridioforme*. *B. fragilis* strains 3_1_12 and CL07T00C01 and *C. clostridioforme* strains 2_1_49_FAA and WAL-7855 were grown in Yeast Casitone Fatty Acids (YCFA) broth and served as positive controls. We find that all strains examined in this assay significantly decolorized Acid Orange 7 *in vitro* (Figure 6). To our knowledge, this is the first reporting of azo reduction by *F. nucleatum* CTI-06, *F. nucleatum* subsp. animalis D11, and *F. nucleatum* subsp. polymorphum.
DISCUSSION

The presence of azo-reducing bacteria in the human gut is necessary for the effective delivery and activation of azo-bonded prodrugs. While azoreductase activity has been identified in several prominent phyla of the human gut microbiota (M. Zimmermann et al. 2019) and appears to be ubiquitous across healthy individuals (Javdan et al. 2020), the prevalence, abundance, and distribution of azoreductases have not been systematically examined in the human gut microbiome of healthy individuals nor in individuals living with IBD. In this work, we curated and compiled known azoreductase genes (Figure 2), searched for azoreductase gene families across a non-redundant set of 4,644 human gut bacterial genomes (Almeida et al. 2020), and identified 1,958 putative azo-reducing species (Figure 3). The systematic comparison of our search results to recent experimental evidence of sulfasalazine reduction by prominent gut bacteria (Table 1, Table 2) indicates a disconnect between the current state of azoreductase annotation and experimental evidence of sulfasalazine reduction. Interestingly, the majority (77.1%, 27/35) of incorrectly predicted sulfasalazine reducing strains are false negatives, meaning, these strains did not return a significant hit to an azoreductase gene from the homolog search step but do in fact reduce sulfasalazine in vitro. This inconsistency between annotated azoreductases and experimental evidence of azo reduction suggests that many prominent bacterial strains of the human gut microbiome may encode and express previously uncharacterized azoreductase genes. These genes likely serve other endogenous roles such as nitro reduction (Liu et al. 2007; Chalansonnet et al. 2017) and quinone oxidoreduction (Liu et al. 2008; Leelakriangsak et al. 2008; Ryan, Wang, et al. 2010; Ryan et al. 2014) with the azo reduction being a side mechanism that these enzymes cross-functionally participate in.

We next sought to report the relative abundance and expression of azoreductases in the human gut microbiome for healthy controls and IBD patients. Our analysis of 1,558 metagenomic samples from 326 individuals across healthy, UC, and CD patient cohorts showed that Clade I, II, III, IVa, IVb, mdaB, and yieF azoreductases are significantly more abundant in
the gut microbiome compared to the other azoreductases examined in this study (Figure 4A-B).

We also examined the expression of azoreductases by the human gut microbiota and found that, with the exception of Clade IVb, expression levels of azoreductases roughly match with their corresponding genomic abundance (Figure 4C). The incongruence of Clade IVb abundance and expression levels suggests that, when feasible, shotgun metagenomic sequencing of stool samples should be performed in parallel with metatranscriptomic sequencing to better understand the functional landscape of the gut microbiome and the relative contributions of different azoreductases to overall azo reduction. We also sought to examine the relative abundance of known and putative azo-reducing bacteria in healthy, UC, and CD patients over time. We found that the relative abundance of known and putative azo-reducing bacteria is significantly (p = 0.002) higher in individuals with CD and is modestly (p = 0.06) higher in individuals with UC compared to healthy controls (Figure 5). This bodes well for the future of azo-bonded prodrug development because these therapies are intended to treat individuals afflicted with UC and CD. However, the cumulative relative abundance of known and putative azo-reducing bacteria fluctuates over time (Figure 5B) and future studies should explore whether there exists some minimum necessary abundance of azo-reducing species for adequate prodrug metabolism and activation.

Finally, we tested the reduction of the azo-bonded dye Acid Orange 7 by three strains of *Fusobacterium nucleatum* alongside positive control strains of *Bacteroides fragilis* and *Clostridium clostridioforme* (Figure 6). *Fusobacterium nucleatum* is positively correlated with colorectal cancer (Marchesi et al. 2011; Kostic et al. 2012), is present in and on cancerous tissue (Castellarin et al. 2012), and possibly contributes to the etiology of the disease (McCoy et al. 2013; Rubinstein et al. 2013; Han 2015). We found that *F. nucleatum* CTI-06, *F. nucleatum* subsp. animalis D11, and *F. nucleatum* subsp. polymorphum all significantly reduce Acid Orange 7 in vitro, indicating the encoding and activity of azoreductases in these strains of *F. nucleatum*. The *F. nucleatum* reference strain present in UHGG received significant hits to the...
mdaB (E-value = 4.20x10^{-13}) and yieF (E-value = 1.40x10^{-23}) HMMs, and thus, represents an accurately predicted azo-reducing bacteria (Supp. Table 3). The identification and characterization of these, and possibly other, *F. nucleatum* azoreductases could lead to the eventual development of an azo-bonded colorectal cancer therapeutic designed specifically to activate in the presence of *F. nucleatum* on the surface of colonic tumors.

*Bacteroides fragilis* is a known reducer of azo dyes including Acid Orange 7 (this study), Amaranth, Orange II, Tartrazine (Bragger et al. 1997), as well as of the quinone menadione (Ito et al. 2020). Additionally, many *B. fragilis* strains have been shown to be potent reducers of sulfasalazine *in vitro* (M. Zimmermann et al. 2019). While we did identify a significant (E-value < 1x10^{-45}) hit to the Clade IvH HMM in the two *B. fragilis* reference strains present in UHGG (Supp. Table 3), there may be other *B. fragilis* genes or operons that exhibit azoreductase activity. Ito et al. 2020 described two NADH:quinone oxidoreductase operons, NQR and NUO, and one NADH:quinone oxidoreductase gene, *ndh2* capable of reducing the quinone menadione. Recall that bacterial quinone oxidoreductases are often cross-reactive with azo compounds and have even been proposed to be a part of the same FMN-dependent superfamily of NAD(P)H utilizing oxidoreductase enzymes (Ryan et al. 2014). Future studies are required to confirm or deny that NQR, NUO, and *ndh2* are hitherto uncharacterized azoreductases contributing to the complete azo reduction of sulfasalazine by *B. fragilis* shown in Zimmerman et al. 2019.

Of the seven bacterial strains tested for reduction of Acid Orange 7, the two *Clostridium clostridioforme* strains exhibited by far the most effective reduction of Acid Orange 7 (*Figure 6*). Although *C. clostridioforme* is a known azo dye reducer (Raffi and Cerniglia 1990; Nakamura et al. 2002; Xu et al. 2010), neither of the two reference strains present in UHGG recruited significant alignments to known azoreductase gene families curated in the homolog search step of this work (Supp. Table 3). This could be the result of either (i) strain-level variation between the reference strains and those tested with Acid Orange 7 in this study, or (ii) the presence and...
activity of one or more uncharacterized azoreductases in \textit{C. clostridioforme}. In either case, further research including a comparative genomics analysis and gene knockout experiment on various strains of \textit{C. clostridioforme} could lead to an improved understanding of gut microbial azo reduction.

This study has two primary limitations. 1. The E-value and percent of alignment thresholds for determining a putative azoreductase in the homolog search step are not absolute, but rather, are designed to strike a balance between identifying spurious homologs and missing the identification of true azoreductase homologs. This is an inherent limitation of studies requiring hard cutoffs for homolog classification, and thus, is very difficult to avoid. 2. Bacterial azoreductases exhibit different substrate specificities (Ryan, Wang, et al. 2010; Lang et al. 2013; Ryan, Laurieri, et al. 2010; Bin et al. 2004; Deller et al. 2006; SUGIURA et al. 2006; Joshi et al. 2008; Mendes et al. 2011), and thus, have varying affinities for different azo prodrugs as well as azo dyes. Though we show a significant reduction of Acid Orange 7 by three strains of \textit{Fusobacterium nucleatum} in this work, future experiments showing the reduction of azo drugs such as sulfasalazine would further bolster the hypothesis that \textit{F. nucleatum} encodes and expresses one or more uncharacterized azoreductases.

In conclusion, we show that known azoreductases are widely distributed in the human gut microbiome and that there are likely many more uncharacterized azoreductases encoded and expressed in the human gut microbiome. These results both (i) bolster previous findings suggesting the ubiquity of azo-reduction in the gut microbiome (Javdan et al. 2020) and (ii) suggest the presence and activity of many hitherto uncharacterized azoreductases in the human gut microbiome. The list of false negative strains identified in our systematic comparison analysis can serve as a resource for future studies focused on identifying azoreductases encoded by the human gut microbiome (Table 1). Overall, this work describes the abundance and distribution of known azoreductases in the human gut microbiome and motivates the need for future studies focused on annotating hitherto uncharacterized azoreductases encoded in the
human gut microbiome. Further validation and annotation of putative azoreductases encoded by prominent members of the gut flora such as *Bacteroides fragilis*, *Fusobacterium nucleatum*, and *Clostridium clostridioforme* are important for functional characterization of azo reduction by the human gut microbiome and for the future of azo prodrug development.
ACKNOWLEDGEMENTS
None.

AUTHOR CONTRIBUTIONS

*Participated in research design:* Braccia, Minabou Ndjite, Jiang, Pop, Hall
*Conducted experiments:* Minabou Ndjite, Weiss, Levy, Abeysinghe
*Performed data analysis:* Braccia, Minabou Ndjite
*Wrote or contributed to the writing of the manuscript:* Braccia, Minabou Ndjite, Weiss, Levy, Abeysinghe, Jiang, Pop, Hall
FOOTNOTES

Funding

DB was supported in part by the National Science Foundation [DGE-1632976] and in part by BH startup funding from the University of Maryland. XJ was supported in part by the Intramural Research Program of the National Library of Medicine (NLM), National Institutes of Health. MP was supported by the National Institutes of Health [R01-AI-100947]. BH, GMN, AW, SL, and SA were supported by startup funding from the University of Maryland.

Financial Disclosure Statement

No author has an actual or perceived conflict of interest with the contents of this article.
REFERENCES


Cerniglia, Carl E., J.P. Freeman, Wirt Franklin, and L.David Pack. 1982. “Metabolism of Azo Dyes Derived from Benzidine, 3,3’-Dimethylbenzidine and 3,3’-Dimethoxybenzidine to


LEGENDS FOR FIGURES

Figure 1. Azo reduction by gut microbiota.
A) Pathways of azo-bonded drug activation by bacterial azoreductase activity. Dihydroquinolone (DHQ) is produced via cyclization of an intermediate of OPN501 and celecoxib-5-ASA prodrug following intramolecular lactamization (Ruiz et al. 2011; Marquez Ruiz et al. 2011). B) Description of downstream metabolites of bacterial azo reduction and the mechanisms of action in inflammatory bowel disease (IBD) and colorectal cancer (CRC). References for each molecular function described in this subfigure: 5-ASA (Mahida et al. 1991; Cominelli et al. 1992; Rachmilewitz et al. 1992), prednisolone (Cohen et al. 2000), celecoxib (A.-W. Wu et al. 2003, 200; W. K. K. Wu et al. 2010; Gustafsson et al. 2010), sulfapyridine (Nielsen 1982), DHQ (Ruiz et al. 2011; Marquez Ruiz et al. 2011). (C) Presence of azoreductase-containing bacteria is required for prodrug activation in the IBD gut and CRC gut (left). Many azo-reducing bacteria have been characterized, however, some species have shown experimental evidence of azo-reduction without the full characterization of the genes responsible for azo reduction (right).
Figure 2. Bacterial azoreductases cluster by cofactor and electron donor preferences.
Following an extensive literature search for experimentally confirmed bacterial azoreductases, amino acid sequences of 40 azoreductase enzymes were collected and clustered with EFI-EST (Gerlt et al. 2015) at 35% sequence identity. Each node in the figure above is a single azoreductase gene and the edges between nodes indicate at least 35% sequence identity between the two amino acid sequences. The colored clusters, Clade I through Clade IVa and IVb, are groups of azoreductases previously described by Suzuki 2019 as mechanistically similar groups based on cofactor and electron donor preferences. Clusters labeled with gene names (mdaB, yieF, etc.) represent homologous gene sequences found in two or more organisms. Each mechanistically characterized group of azoreductases were subsequently pressed into profile Hidden Markov Models (HMMs) using HMMER v3.1b2 (Finn et al. 2015) which formed the basis of the homologue search. The group labeled “other azoreductases” contains sequences that did not fall into any cluster at the 35% identity threshold, and were pressed into singular HMMs prior to the homologue search.
Figure 3. Azoreductases are widely distributed across gut bacterial taxonomy.

Presence/absence of azoreductases across prominent phyla of the human gut microbiome. The taxonomic tree is obtained from the Unified Human Gastrointestinal Genome collection (UHGG) (Almeida et al. 2020) which is built on the Genome Taxonomy Database (GTDB) (Chaumeil et al. 2020). Phyla names are annotated on the left side. Phyla names followed by a capital letter, e.g., Firmicutes (A), indicate a novel phylum classified by the GTDB-tk. The bar chart in the center indicates the number of species contained in each genus shown in the tree. The size of the circles indicates the number of species which contain hits to the azoreductase genes specified. The color of the circles indicates the cofactor and preferred electron (e⁻) donor of the enzyme.
Figure 4. Abundance and expression of azoreductase genes by human gut microbiota.

(A,B) Visualization of shotgun metagenomic sequencing data from the Human Microbiome Project 2 also known as the Integrative Human Microbiome Project (HMP2) and the Prospective Registry in IBD Study at MGH (PRISM). We used salmon 1.4.0 (Patro et al. 2017) to quantify the abundance of azoreductase genes from hundreds of stool samples across healthy controls (nonIBD) ulcerative colitis (UC) and Crohn’s Disease (CD) participant cohorts. Raw DNA read alignment counts were normalized to counts per million (CPM), analogous to transcripts per million (TPM) normalization. Asterisks above each boxplot indicate statistical significance (* = p < 0.05, ** = p < 0.01, *** = p < 0.001, Wilcoxon rank sum test, all FDR adjusted using Benjamini-Hochberg method)

(C) Visualization of high throughput metatranscriptomic data obtained from the Health Professionals Follow-up Study (HPFS). We quantified the expression of bacterial azoreductases using salmon v1.4.1 (Patro et al. 2017) and normalized the raw read alignment statistics to TPM. Please see the materials and methods section for a more detailed description of the computational and statistical methods employed.
Figure 5. Known and putative azo-reducing species are more abundant in the IBD gut.
Relative abundance of known and putative azo-reducing species in HMP2 subjects with more than 20 total stool collections. (A) Relative abundance of known and putative azo-reducing species for qualifying participants across nonIBD, UC, and CD populations. Subjects with Crohn’s Disease (CD) have significantly higher relative abundances of known+putative azo-reducing species than healthy subjects (nonIBD) per linear mixed effects model ANOVA, F(1,10) = 17.09, p < 0.003. (B) Relative abundance of putative azo-reducing species over time for healthy (nonIBD), Ulcerative Colitis (UC), and Crohn’s Disease (CD) participants. Each line represents a single participant and each point is the summed relative abundance of known+putative azo-reducing species at that collection point. The key on the right links relative abundance distributions for each subject with the same data point shown over collection numbers. Collections were taken approximately every 14 days.
Figure 6. Three putative azo-reducing strains of *Fusobacterium nucleatum* degrade Acid Orange 7 in vitro. The absorbance of light at 550nm (corresponding to the absorbance spectra of Acid Orange 7) was measured in cultures of *Fusobacterium nucleatum*, *Bacteroides fragilis*, and *Clostridium clostridiforme* isolate cultures. *Fusobacterium nucleatum* strains were grown in BHI media and were compared to BHI-blank control mixture whereas *Bacteroides fragilis* and *Clostridium clostridiforme* strains were grown in YCFA media and were thus compared to a YCFA-blank. Each strain was grown and tested in biological and technical triplicates. Each data point on the plot above is the average of three technical replicates from a single biological replicate per strain. Please see the materials and methods section for more details regarding our experimental methodology. Asterisks indicate statistical significance calculated via two-sided t-tests (* = p < 0.05, ** = p < 0.01, *** = p < 0.001).
# TABLES

Table 1. Systematic comparison with Zimmermann et al. 2019 sulfasalazine consumption results. Complete results from the systematic comparison of predicted sulfasalazine reducing bacteria to actual sulfasalazine reducing bacteria. The columns labeled FC (fold change), FC_STD (standard deviation in fold change), p_FDR (FDR adjusted p-value), pct_consumed (percent consumed), and pct_consumed_STD (standard deviation of the percent consumed) were all obtained directly from Zimmermann et al. 2019 Supp. Table 3. The SR_status column contains values SR (sulfasalazine reducer) and nonSR (non sulfasalazine reducer) which were determined based on significant (p_FDR < 0.05) or non-significant (p_FDR >= 0.05) sulfasalazine reduction. The PR_status column contains the values PR (predicted sulfasalazine reducer) and nonPR (predicted sulfasalazine non-reducer) which were determined based on the presence or absence of one or more azoreductase homologs determined from the homolog search step. The final column, result, contains the values TP (true positive), TN (true negative), FP (false positive), and FN (false negative). Correctly predicted SR strains have a result of TP and correctly predicted nonSR strains have a result of TN whereas incorrectly predicted SR strains have a result of FP and incorrectly predicted nonSR strains have a result of FN.

<table>
<thead>
<tr>
<th>strain_name</th>
<th>FC</th>
<th>FC_STD</th>
<th>p_FDR</th>
<th>pct_consumed</th>
<th>pct_consumed_STD</th>
<th>SR_status</th>
<th>PR_status</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akkermansia muciniphila ATCCBAA-835</td>
<td>-0.419</td>
<td>0.361</td>
<td>0.19</td>
<td>25.222</td>
<td>18.707</td>
<td>nonSR</td>
<td>nonPR</td>
<td>TN</td>
</tr>
<tr>
<td>Alistipes indistinctus DSM 22520</td>
<td>-9.01</td>
<td>0.11</td>
<td>0.003</td>
<td>99.806</td>
<td>0.015</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Anaerococcus hydrogenalis DSM7454</td>
<td>-8.826</td>
<td>0.198</td>
<td>0.008</td>
<td>99.78</td>
<td>0.03</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Anaerotruncus cohominis DSM17241</td>
<td>-8.088</td>
<td>1.575</td>
<td>0.016</td>
<td>99.633</td>
<td>0.401</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides caccae ATCC43185</td>
<td>-9.247</td>
<td>0.262</td>
<td>0.008</td>
<td>99.835</td>
<td>0.03</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides cellulosolyticus DSM14838</td>
<td>-1.502</td>
<td>0.308</td>
<td>0.002</td>
<td>64.691</td>
<td>7.532</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides coprophilus DSM18228</td>
<td>-0.006</td>
<td>0.216</td>
<td>0.01</td>
<td>3.366</td>
<td>14.918</td>
<td>nonSR</td>
<td>nonPR</td>
<td>TN</td>
</tr>
<tr>
<td>Bacteroides dorei DSM17555</td>
<td>-4.631</td>
<td>0.748</td>
<td>0.013</td>
<td>95.965</td>
<td>2.091</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides eggerthii DSM20869</td>
<td>-9.59</td>
<td>0.104</td>
<td>0.002</td>
<td>99.87</td>
<td>0.059</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides finegoldii DSM17565</td>
<td>-0.79</td>
<td>0.296</td>
<td>0.042</td>
<td>42.176</td>
<td>11.877</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides fragilis 539T T10</td>
<td>-0.974</td>
<td>0.348</td>
<td>0.009</td>
<td>49.084</td>
<td>12.282</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides fragilis ATCC43859</td>
<td>-10.576</td>
<td>0.088</td>
<td>0.003</td>
<td>99.934</td>
<td>0.004</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides fragilis DS-208</td>
<td>-9.327</td>
<td>0.57</td>
<td>0.006</td>
<td>99.844</td>
<td>0.062</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides fragilis HMW610</td>
<td>-10.522</td>
<td>0.159</td>
<td>0.004</td>
<td>99.932</td>
<td>0.067</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides fragilis HMW615</td>
<td>-10.398</td>
<td>0.256</td>
<td>0.013</td>
<td>99.926</td>
<td>0.013</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides fragilis NCTC9343</td>
<td>-6.184</td>
<td>0.835</td>
<td>0.006</td>
<td>98.625</td>
<td>0.796</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides fragilis TBp9</td>
<td>-9.252</td>
<td>0.144</td>
<td>0.005</td>
<td>99.836</td>
<td>0.016</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides intestinalis DSM17393</td>
<td>-1.206</td>
<td>0.398</td>
<td>0.005</td>
<td>99.267</td>
<td>11.246</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides ovatus ATCC8483</td>
<td>-9.285</td>
<td>0.473</td>
<td>0.546</td>
<td>17.951</td>
<td>26.915</td>
<td>nonSR</td>
<td>PR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides pectinophilus ATCC43243</td>
<td>-0.249</td>
<td>0.241</td>
<td>0.208</td>
<td>15.832</td>
<td>14.042</td>
<td>nonSR</td>
<td>nonPR</td>
<td>TN</td>
</tr>
<tr>
<td>Bacteroides stercolis ATCC43183</td>
<td>-0.285</td>
<td>0.326</td>
<td>0.053</td>
<td>33.46</td>
<td>15.033</td>
<td>nonSR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides thetaiotaomicron 3731</td>
<td>-5.828</td>
<td>0.248</td>
<td>0.055</td>
<td>96.461</td>
<td>6.961</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides thetaiotaomicron 7330</td>
<td>-0.132</td>
<td>0.237</td>
<td>0.003</td>
<td>51.082</td>
<td>8.044</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides thalaeacameroni VPI-5452</td>
<td>-1.252</td>
<td>0.155</td>
<td>0.006</td>
<td>58.006</td>
<td>4.511</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Bacteroides uniformis ATCC25986</td>
<td>-5.605</td>
<td>0.389</td>
<td>0.001</td>
<td>83.558</td>
<td>4.436</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bacteroides xylanisolvens DSM18836</td>
<td>-9.663</td>
<td>0.115</td>
<td>0.002</td>
<td>99.977</td>
<td>0.017</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bifidobacterium adolescentis ATCC15703</td>
<td>-0.7</td>
<td>0.271</td>
<td>0.014</td>
<td>38.442</td>
<td>11.584</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bifidobacterium breve DSM20213</td>
<td>-9.241</td>
<td>0.182</td>
<td>0.008</td>
<td>99.835</td>
<td>0.021</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Blautia hansenii DSM20583</td>
<td>-9.234</td>
<td>0.155</td>
<td>0.005</td>
<td>99.834</td>
<td>0.018</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Bryntgis formalisoxgens DSM14469</td>
<td>-6.722</td>
<td>0.432</td>
<td>0.113</td>
<td>39.359</td>
<td>18.171</td>
<td>nonSR</td>
<td>PR</td>
<td>FN</td>
</tr>
<tr>
<td>Clostridium asparagiforme DSM15991</td>
<td>-9.184</td>
<td>0.203</td>
<td>0.01</td>
<td>99.828</td>
<td>0.024</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Clostridium bollea ATCCBaA-613</td>
<td>-7.113</td>
<td>0.611</td>
<td>0.005</td>
<td>99.277</td>
<td>0.306</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Clostridium difficile 120</td>
<td>-6.539</td>
<td>0.703</td>
<td>0.013</td>
<td>98.479</td>
<td>0.741</td>
<td>SR</td>
<td>nonPR</td>
<td>TN</td>
</tr>
<tr>
<td>Clostridium ramosum DSM1402</td>
<td>-9.046</td>
<td>0.189</td>
<td>0.006</td>
<td>99.811</td>
<td>0.022</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Clostridium scindens ATCC35704</td>
<td>-9.084</td>
<td>0.324</td>
<td>0.004</td>
<td>99.813</td>
<td>0.042</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Clostridium spiroforme DSM15632</td>
<td>-2.478</td>
<td>0.572</td>
<td>0.001</td>
<td>82.056</td>
<td>7.119</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Clostridium sporogenes ATCC15579</td>
<td>-9.072</td>
<td>0.158</td>
<td>0.006</td>
<td>99.814</td>
<td>0.02</td>
<td>SR</td>
<td>nonPR</td>
<td>TN</td>
</tr>
<tr>
<td>Clostridium symbiosum ATCC14040</td>
<td>-9.256</td>
<td>0.209</td>
<td>0.009</td>
<td>99.836</td>
<td>0.024</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Collinsella aerofaciens ATCC25986</td>
<td>-5.229</td>
<td>0.828</td>
<td>0.046</td>
<td>98.999</td>
<td>0.44</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Collinsella intestinalis DSM13280</td>
<td>-8.916</td>
<td>0.118</td>
<td>0.003</td>
<td>99.793</td>
<td>0.017</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Coprococcus comes ATCC27758</td>
<td>-6.182</td>
<td>0.199</td>
<td>0.009</td>
<td>99.828</td>
<td>0.024</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Clostridium pasteuri DSM20583</td>
<td>-4.071</td>
<td>0.57</td>
<td>0.015</td>
<td>94.051</td>
<td>2.351</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Edwardsiella tarda ATCC23685</td>
<td>-0.107</td>
<td>0.287</td>
<td>0.722</td>
<td>7.125</td>
<td>18.483</td>
<td>nonSR</td>
<td>PR</td>
<td>FN</td>
</tr>
<tr>
<td>Enterobacter aerogenes ATCC35316</td>
<td>-0.853</td>
<td>0.579</td>
<td>0.076</td>
<td>44.621</td>
<td>22.214</td>
<td>nonSR</td>
<td>PR</td>
<td>FN</td>
</tr>
<tr>
<td>Enterobacter cloacae VSE3</td>
<td>-8.61</td>
<td>0.228</td>
<td>0.01</td>
<td>99.744</td>
<td>0.04</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
</tr>
<tr>
<td>Escherichia coli K-12</td>
<td>-1.71</td>
<td>0.311</td>
<td>0.038</td>
<td>39.097</td>
<td>13.995</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Eubacterium biforme DSM13899</td>
<td>-8.927</td>
<td>0.098</td>
<td>0.008</td>
<td>99.795</td>
<td>0.014</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Eubacterium hallii DSM11534</td>
<td>-9.031</td>
<td>0.381</td>
<td>0.028</td>
<td>99.809</td>
<td>0.05</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Eubacterium rectale ATCC33658</td>
<td>-9.002</td>
<td>0.322</td>
<td>0.022</td>
<td>99.805</td>
<td>0.044</td>
<td>SR</td>
<td>nonPR</td>
<td>TN</td>
</tr>
<tr>
<td>Eubacterium ventriosum ATCC97560</td>
<td>-4.653</td>
<td>0.827</td>
<td>0.061</td>
<td>96.024</td>
<td>2.278</td>
<td>nonSR</td>
<td>nonPR</td>
<td>FN</td>
</tr>
<tr>
<td>Species/Strain</td>
<td>Log2FC</td>
<td>p-value</td>
<td>FDR</td>
<td>Raw16S</td>
<td>RawQI90</td>
<td>Strain Type</td>
<td>Classification</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>--------</td>
<td>---------</td>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>-------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>Odontobacter splanchnius ATCC8503</td>
<td>-7.892</td>
<td>0.004</td>
<td>99.579</td>
<td>0.217</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
<td></td>
</tr>
<tr>
<td>Parabacteroides distasonis DSM18315</td>
<td>-1.007</td>
<td>0.22</td>
<td>50.253</td>
<td>10.272</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Parabacteroides johnsonii DSM18315</td>
<td>-0.529</td>
<td>0.123</td>
<td>30.702</td>
<td>13.51</td>
<td>nonSR</td>
<td>nonPR</td>
<td>TN</td>
<td></td>
</tr>
<tr>
<td>Parabacteroides merdae ATCC43194</td>
<td>-0.749</td>
<td>0.006</td>
<td>40.508</td>
<td>8.593</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Photobacterium sp. DSM8205</td>
<td>-8.693</td>
<td>0.015</td>
<td>99.758</td>
<td>0.069</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Photobacterium pelvenri ATCC35198</td>
<td>-1.657</td>
<td>0.099</td>
<td>68.289</td>
<td>8.884</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
<td></td>
</tr>
<tr>
<td>Providencia alcalifaciens DSM30120</td>
<td>-0.379</td>
<td>0.158</td>
<td>23.094</td>
<td>13.183</td>
<td>nonSR</td>
<td>PR</td>
<td>FP</td>
<td></td>
</tr>
<tr>
<td>Providencia rettgeri DSM1131</td>
<td>-0.205</td>
<td>0.25</td>
<td>13.245</td>
<td>12.273</td>
<td>nonSR</td>
<td>PR</td>
<td>FP</td>
<td></td>
</tr>
<tr>
<td>Providencia stuartii ATCC25627</td>
<td>-0.072</td>
<td>0.807</td>
<td>4.854</td>
<td>16.212</td>
<td>nonSR</td>
<td>PR</td>
<td>FP</td>
<td></td>
</tr>
<tr>
<td>Roseburia intestinalis L1-82</td>
<td>-8.876</td>
<td>0.006</td>
<td>99.787</td>
<td>0.623</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Ruminococcus gnavus ATCC29149</td>
<td>-8.639</td>
<td>0.008</td>
<td>99.749</td>
<td>0.008</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Ruminococcus lactaris ATCC29176</td>
<td>-9.522</td>
<td>0.014</td>
<td>99.864</td>
<td>0.814</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Ruminococcus torques ATCC27756</td>
<td>-2.884</td>
<td>0.016</td>
<td>80.841</td>
<td>3.561</td>
<td>SR</td>
<td>PR</td>
<td>TP</td>
<td></td>
</tr>
<tr>
<td>Salmonella Typhimurium LT2</td>
<td>-0.673</td>
<td>0.252</td>
<td>37.287</td>
<td>30.936</td>
<td>nonSR</td>
<td>PR</td>
<td>FP</td>
<td></td>
</tr>
<tr>
<td>Subdoligranulum variabile DSM18176</td>
<td>-8.854</td>
<td>0.008</td>
<td>99.784</td>
<td>0.008</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Victivallis vandenis ATCC BAA-548</td>
<td>-1.844</td>
<td>0.004</td>
<td>72.146</td>
<td>7.574</td>
<td>SR</td>
<td>nonPR</td>
<td>FN</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Summarized results of systematic Zimmermann 2019 comparison. This table displays the summarized results of the systematic comparison of predicted sulfasalazine reducers to experimentally confirmed sulfasalazine reducers reported by Zimmermann et al. 2019. \(^a\) The number of true positives, \(^b\) The number of true false positives, \(^c\) The number of false negatives, \(^d\) The number of true negatives.

<table>
<thead>
<tr>
<th></th>
<th>PR</th>
<th>nonPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR</td>
<td>38.8% (26/67) (^a)</td>
<td>40.3% (27/67) (^c)</td>
</tr>
<tr>
<td>nonSR</td>
<td>11.9% (8/67) (^b)</td>
<td>9.0% (6/67) (^d)</td>
</tr>
</tbody>
</table>
**Figure 1**

**A**

- Sulfasalazine → AZR → sulfapyridine + 5-ASA
- OPN501 → AZR → prednisolone + 5-ASA
- Celecoxib-5-ASA prodrug → AZR → celecoxib + 5-ASA

**B**

- **5-ASA**: Blocks binding of interleukin-1 (IL-1) and tumor necrosis factor-alpha (TNF-alpha) to relative receptors, thus inhibiting NF-kb. Also known as mesalamine.
- **Prednisolone**: An anti-inflammatory corticosteroid commonly prescribed to those with inflammatory bowel disease.
- **Celecoxib**: Inhibitor of cyclooxygenase 2 (COX-2) expression by cancer cells, thus limiting eicosanoid production and colorectal cancer progression.
- **Sulfapyridine**: Carrier molecule for delivery of 5-ASA to distal colon. Has been linked with negative side effects such as exanthem, itchiness and fever.
- **DHQ**: Product of cyclization reaction after azo-reduction of OPN501 and celecoxib-5-ASA prodrug.
- **Azo-bonded prodrugs**: Not readily absorbed in the lumen of the small intestine.
- **AZR**: Azoreductase activity.
- **N=N**: Azo-bond.

**C**

- **Azoreductase activity in the human gut microbiome**
  - Sulfasalazine → 5-ASA
  - OPN501 → prednisolone
  - Celecoxib-5-ASA prodrug → celecoxib

- **Uncharacterized azoreductases in the gut microbiome**
  - Known azoreducers:
    - *Pseudomonas aeruginosa*
    - *Bacillus subtilis*
    - *Bacteroides fragilis*
    - *Enterococcus faecium*
  - Uncharacterized azoreducers:
    - *Fusobacterium nucleatum*
    - *Ruminococcus gravis*
    - *Anaerococcus hydrogenalis*
    - *Bacteroides dorei*
Figure 3

This article has not been copyedited and formatted. The final version may differ from this version.
Figure 4

A. HMP2, N(samples) = 703, N(individuals) = 104

B. PRISM, N(samples) = 218, N(individuals) = 218

C. HPFS, N(samples) = 220, N(individuals) = 220
Figure 5

(A) Linear mixed effect model ANOVA, $F(1,10) = 17.09, p = 0.002$

(B) Relative abundance %

Diagnosis:
- nonIBD Subjects
  - M2072
  - H4023
  - M2042
  - H4019
  - M2084
  - C3022
  - H4035
  - M2064
  - P6018
- UC Subjects
  - H4019
  - H4035
  - M2064
  - C3022
  - H4024
  - M2084
  - H4020
  - P6013
- CD Subjects
  - M2068
  - P6010
  - C3027
  - H4020
  - H4015

Collection number:
- 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Downloaded from ASPET Journals on November 26, 2022.
Figure 6

The graph shows absorbance (505 nm) measurements for different bacterial strains. The absorbance values are compared to a blank (BHI) condition. The bars represent the absorbance values for:

- F. nucleatum subsp. animalis D11
- F. nucleatum subsp. polymorphum
- Blank (YCFA)
- B. fragilis CL07T00C01
- B. fragilis 3_1_12
- C. clostridiiforme 2_1_49FAA
- C. clostridiiforme WAL-7855

Significance levels are indicated as *** (p < 0.001) and ** (p < 0.01).