DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

TITLE PAGE

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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*.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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).

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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*.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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**). All

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

spectrophotometer. The raw absorbance values for each biological and technical replicate are reported in Supp. Table 5.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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 $p < 2.2 \times 10^{-16}$) and PRISM (all FDR adjusted $p < 2.6 \times 10^{-7}$)

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

(**Figure 4A-B**). 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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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 clostridiforme*. *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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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 IVa 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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

activity of one or more uncharacterized azoreductases in *C. clostridioforme*. In either case, further research including a comparative genomics analysis and gene knockout experiment on various strains of *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 *Fusobacterium nucleatum* in this work, future experiments showing the reduction of azo drugs such as sulfasalazine would further bolster the hypothesis that *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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

REFERENCES

Abed, Jawad, Johanna E. M. Emgård, Gideon Zamir, Mouhammad Faroja, Gideon Almogy, Amalie Grenov, Asaf Sol, et al. 2016. “Fap2 Mediates Fusobacterium Nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc.” *Cell Host & Microbe* 20 (2): 215–25. https://doi.org/10.1016/j.chom.2016.07.006.

Abu-Ali, Galeb S., Raaj S. Mehta, Jason Lloyd-Price, Himel Mallick, Tobyn Branck, Kerry L. Ivey, David A. Drew, et al. 2018. “Metatranscriptome of Human Faecal Microbial Communities in a Cohort of Adult Men.” *Nature Microbiology* 3 (3): 356–66. https://doi.org/10.1038/s41564-017-0084-4.

Almeida, Alexandre, Stephen Nayfach, Miguel Boland, Francesco Strozzi, Martin Beracochea, Zhou Jason Shi, Katherine S. Pollard, et al. 2020. “A Unified Catalog of 204,938 Reference Genomes from the Human Gut Microbiome.” *Nature Biotechnology*, July, 1–10. https://doi.org/10.1038/s41587-020-0603-3.

“Babraham Bioinformatics - Trim Galore!” n.d. Accessed August 18, 2021. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Bashir, Arif, Abid Y. Miskeen, Ashaqullah Bhat, Khalid M. Fazili, and Bashir A. Ganai. 2015. “Fusobacterium Nucleatum: An Emerging Bug in Colorectal Tumorigenesis.” *European Journal of Cancer Prevention* 24 (5): 373–85.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. “Fitting Linear Mixed-Effects Models Using Lme4.” *ArXiv:1406.5823 [Stat]*, June. http://arxiv.org/abs/1406.5823.

Bay, Hui Han, Chi Kim Lim, Thuan Chien Kee, Ismail Ware, Giek Far Chan, Shafinaz Shahir, and Zaharah Ibrahim. 2014. “Decolourisation of Acid Orange 7 Recalcitrant Auto-Oxidation Coloured by-Products Using an Acclimatised Mixed Bacterial Culture.” *Environmental Science and Pollution Research* 21 (5): 3891–3906. https://doi.org/10.1007/s11356-013-2331-4.

Bin, Yan, Zhou Jiti, Wang Jing, Du Cuihong, Hou Hongman, Song Zhiyong, and Bao Yongming. 2004. “Expression and Characteristics of the Gene Encoding Azoreductase from Rhodobacter Sphaeroides AS1.1737.” *FEMS Microbiology Letters* 236 (1): 129–36. https://doi.org/10.1111/j.1574-6968.2004.tb09638.x.

Blümel, S., and A. Stolz. 2003. “Cloning and Characterization of the Gene Coding for the Aerobic Azoreductase from Pigmentiphaga Kullae K24.” *Applied Microbiology and Biotechnology* 62 (2): 186–90. https://doi.org/10.1007/s00253-003-1316-5.

Blümel, Silke, Hans-Joachim Knackmuss, and Andreas Stolz. 2002. “Molecular Cloning and Characterization of the Gene Coding for the Aerobic Azoreductase from Xenophilus Azovorans KF46F.” *Applied and Environmental Microbiology* 68 (8): 3948–55. https://doi.org/10.1128/AEM.68.8.3948-3955.2002.

Bragger, J. L, A. W Lloyd, S. H Soozandehfar, S. F Bloomfield, C Marriott, and G. P Martin. 1997. “Investigations into the Azo Reducing Activity of a Common Colonic Microorganism.” *International Journal of Pharmaceutics* 157 (1): 61–71. https://doi.org/10.1016/S0378-5173(97)00214-7.

Brown, Joseph P. 1981. “Reduction of Polymeric Azo and Nitro Dyes by Intestinal Bacteria.” *Applied and Environmental Microbiology* 41 (5): 1283–86. https://doi.org/10.1128/aem.41.5.1283-1286.1981.

Castellarin, Mauro, René L. Warren, J. Douglas Freeman, Lisa Dreolini, Martin Krzywinski, Jaclyn Strauss, Rebecca Barnes, et al. 2012. “Fusobacterium Nucleatum Infection Is Prevalent in Human Colorectal Carcinoma.” *Genome Research* 22 (2): 299–306. https://doi.org/10.1101/gr.126516.111.

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

Potentially Carcinogenic Aromatic Amines by Intestinal Bacteria." *Carcinogenesis* 3 (11): 1255–60. https://doi.org/10.1093/carcin/3.11.1255.

Chalansonnet, Valérie, Claire Mercier, Sylvain Orenga, and Christophe Gilbert. 2017. "Identification of Enterococcus Faecalis Enzymes with Azoreductases and/or Nitroreductase Activity." *BMC Microbiology* 17 (1): 1–10. https://doi.org/10.1186/s12866-017-1033-3.

Chaumeil, Pierre-Alain, Aaron J Mussig, Philip Hugenholtz, and Donovan H Parks. 2020. "GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database." *Bioinformatics* 36 (6): 1925–27. https://doi.org/10.1093/bioinformatics/btz848.

Chen, Huizhong, Sherryll L. Hopper, and Carl E. Cerniglia. 2005. "Biochemical and Molecular Characterization of an Azoreductase from Staphylococcus Aureus, a Tetrameric NADPH-Dependent Flavoprotein." *Microbiology (Reading, England)* 151 (Pt 5): 1433–41. https://doi.org/10.1099/mic.0.27805-0.

Chen, Huizhong, Rong-Fu Wang, and Carl E Cerniglia. 2004. "Molecular Cloning, Overexpression, Purification, and Characterization of an Aerobic FMN-Dependent Azoreductase from Enterococcus Faecalis." *Protein Expression and Purification* 34 (2): 302–10. https://doi.org/10.1016/j.pep.2003.12.016.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. "Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor." *Bioinformatics* 34 (17): i884–90. https://doi.org/10.1093/bioinformatics/bty560.

Cohen, Russell D, Douglas M Woseth, Ronald A Thisted, and Stephen B Hanauer. 2000. "A Meta-Analysis and Overview of the Literature on Treatment Options for Left-Sided Ulcerative Colitis and Ulcerative Proctitis." *The American Journal of Gastroenterology* 95 (5): 1263–76. https://doi.org/10.1016/S0002-9270(00)00732-2.

Cominelli, Fabio, Cynthia C. Nast, Andrea Duchini, and Marjorie Lee. 1992. "Recombinant Interleukin-1 Receptor Antagonist Blocks the Proinflammatory Activity of Endogenous Interleukin-1 in Rabbit Immune Colitis." *Gastroenterology* 103 (1): 65–71. https://doi.org/10.1016/0016-5085(92)91096-M.

Deller, Sigrid, Sonja Sollner, Rosemarie Trenker-El-Toukhy, Ilian Jelesarov, Georg M. Gübitz, and Peter Macheroux. 2006. "Characterization of a Thermostable NADPH:FMN Oxidoreductase from the Mesophilic Bacterium Bacillus Subtilis." *Biochemistry* 45 (23): 7083–91. https://doi.org/10.1021/bi052478r.

Dobkin, Jay F., Jnan R. Saha, Vincent P. Butler, Harold C. Neu, and John Lindenbaum. 1983. "Digoxin-Inactivating Bacteria: Identification in Human Gut Flora." *Science* 220 (4594): 325–27. https://doi.org/10.1126/science.6836275.

Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Eslami, Maryam, Mohammad Ali Amoozegar, and Sedigheh Asad. 2016. "Isolation, Cloning and Characterization of an Azoreductase from the Halophilic Bacterium Halomonas Elongata." *International Journal of Biological Macromolecules* 85 (April): 111–16. https://doi.org/10.1016/j.ijbiomac.2015.12.065.

Feng, Jinhui, Carl E. Cerniglia, and Huizhong Chen. 2012. "Toxicological Significance of Azo Dye Metabolism by Human Intestinal Microbiota." *Frontiers in Bioscience (Elite Edition)* 4 (January): 568–86.

Finn, Robert D., Jody Clements, William Arndt, Benjamin L. Miller, Travis J. Wheeler, Fabian Schreiber, Alex Bateman, and Sean R. Eddy. 2015. "HMMER Web Server: 2015 Update." *Nucleic Acids Research* 43 (W1): W30–38. https://doi.org/10.1093/nar/gkv397.

Foppoli, Anastasia, Alessandra Maroni, Saliha Moutaharrik, Alice Melocchi, Lucia Zema, Luca Palugan, Matteo Cerea, and Andrea Gazzaniga. 2019. "In Vitro and Human Pharmacoscintigraphic Evaluation of an Oral 5-ASA Delivery System for Colonic

Release." *International Journal of Pharmaceutics* 572 (December): 118723. https://doi.org/10.1016/j.ijpharm.2019.118723.

Franzosa, Eric A., Alexandra Sirota-Madi, Julian Avila-Pacheco, Nadine Fornelos, Henry J. Haiser, Stefan Reinker, Tommi Vatanen, et al. 2019. "Gut Microbiome Structure and Metabolic Activity in Inflammatory Bowel Disease." *Nature Microbiology* 4 (2): 293–305. https://doi.org/10.1038/s41564-018-0306-4.

Friend, David R. 2005. "New Oral Delivery Systems for Treatment of Inflammatory Bowel Disease." *Advanced Drug Delivery Reviews*, Drug Delivery in Advancing the Treatment of Inflammatory Bowel Disease, 57 (2): 247–65. https://doi.org/10.1016/j.addr.2004.08.011.

Gerlt, John A., Jason T. Bouvier, Daniel B. Davidson, Heidi J. Imker, Boris Sadkhin, David R. Slater, and Katie L. Whalen. 2015. "Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1854 (8): 1019–37. https://doi.org/10.1016/j.bbapap.2015.04.015.

Gonçalves, Ana Maria D., Sónia Mendes, Daniele de Sanctis, Lígia O. Martins, and Isabel Bento. 2013. "The Crystal Structure of Pseudomonas Putida Azoreductase – the Active Site Revisited." *The FEBS Journal* 280 (24): 6643–57. https://doi.org/10.1111/febs.12568.

Gustafsson, Annika, Marianne Andersson, Kristina Lagerstedt, Christina Lönnroth, Svante Nordgren, and Kent Lundholm. 2010. "Receptor and Enzyme Expression for Prostanoid Metabolism in Colorectal Cancer Related to Tumor Tissue PGE2." *International Journal of Oncology* 36 (2): 469–78. https://doi.org/10.3892/ijo_00000521.

Haiser, Henry J., David B. Gootenberg, Kelly Chatman, Gopal Sirasani, Emily P. Balskus, and Peter J. Turnbaugh. 2013. "Predicting and Manipulating Cardiac Drug Inactivation by the Human Gut Bacterium Eggerthella Lenta." *Science* 341 (6143): 295–98.

Han, Yiping W. 2015. "Fusobacterium Nucleatum: A Commensal-Turned Pathogen." *Current Opinion in Microbiology*, Host–microbe interactions: bacteria • Genomics, 23 (February): 141–47. https://doi.org/10.1016/j.mib.2014.11.013.

Ito, Takeshi, Rene Gallegos, Leigh M. Matano, Nicole L. Butler, Noam Hantman, Matthew Kaili, Michael J. Coyne, Laurie E. Comstock, Michael H. Malamy, and Blanca Barquera. 2020. "Genetic and Biochemical Analysis of Anaerobic Respiration in Bacteroides Fragilis and Its Importance *In Vivo*." Edited by Derek R. Lovley. *MBio* 11 (1): e03238-19. https://doi.org/10.1128/mBio.03238-19.

Javdan, Bahar, Jaime G. Lopez, Pranatchareeya Chankhamjon, Ying-Chiang J. Lee, Raphaella Hull, Qihao Wu, Xiaojuan Wang, Seema Chatterjee, and Mohamed S. Donia. 2020. "Personalized Mapping of Drug Metabolism by the Human Gut Microbiome." *Cell* 181 (7): 1661-1679.e22. https://doi.org/10.1016/j.cell.2020.05.001.

Joshi, Taruna, Leela Iyengar, Karunakar Singh, and Sanjeev Garg. 2008. "Isolation, Identification and Application of Novel Bacterial Consortium TJ-1 for the Decolourization of Structurally Different Azo Dyes." *Bioresource Technology* 99 (15): 7115–21. https://doi.org/10.1016/j.biortech.2007.12.074.

Kalman, Sumner M. 1978. "Digoxin Metabolism in Patients," 3.

Kostic, Aleksandar D., Dirk Gevers, Chandra Sekhar Pedamallu, Monia Michaud, Fujiko Duke, Ashlee M. Earl, Akinyemi I. Ojesina, et al. 2012. "Genomic Analysis Identifies Association of Fusobacterium with Colorectal Carcinoma." *Genome Research* 22 (2): 292–98. https://doi.org/10.1101/gr.126573.111.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. "LmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82 (December): 1–26. https://doi.org/10.18637/jss.v082.i13.

Lang, Weeranuch, Sarote Sirisansaneeyakul, Lukana Ngiwsara, Sónia Mendes, Lígia O. Martins, Masayuki Okuyama, and Atsuo Kimura. 2013. “Characterization of a New Oxygen-Insensitive Azoreductase from Brevibacillus Laterosporus TISTR1911: Toward Dye Decolorization Using a Packed-Bed Metal Affinity Reactor.” *Bioresource Technology* 150 (December): 298–306. https://doi.org/10.1016/j.biortech.2013.09.124.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Leelakriangsak, Montira, Nguyen Thi Thu Huyen, Stefanie Töwe, Nguyen Van Duy, Dörte Becher, Michael Hecker, Haike Antelmann, and Peter Zuber. 2008. “Regulation of Quinone Detoxification by the Thiol Stress Sensing DUF24/MarR-like Repressor, YodB in Bacillus Subtilis.” *Molecular Microbiology* 67 (5): 1108–24. https://doi.org/10.1111/j.1365-2958.2008.06110.x.

Letunic, Ivica, and Peer Bork. 2019. “Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments.” *Nucleic Acids Research* 47 (W1): W256–59. https://doi.org/10.1093/nar/gkz239.

Liu, Guangfei, Jiti Zhou, Ruofei Jin, Mi Zhou, Jing Wang, Hong Lu, and Yuanyuan Qu. 2008. “Enhancing Survival of Escherichia Coli by Expression of Azoreductase AZR Possessing Quinone Reductase Activity.” *Applied Microbiology and Biotechnology* 80 (3): 409. https://doi.org/10.1007/s00253-008-1555-6.

Liu, Guangfei, Jiti Zhou, Hong Lv, Xuemin Xiang, Jing Wang, Mi Zhou, and Yuanyuan Qv. 2007. “Azoreductase from Rhodobacter Sphaeroides AS1.1737 Is a Flavodoxin That Also Functions as Nitroreductase and Flavin Mononucleotide Reductase.” *Applied Microbiology and Biotechnology* 76 (6): 1271–79. https://doi.org/10.1007/s00253-007-1087-5.

Mahida, Y. R., C. E. Lamming, A. Gallagher, A. B. Hawthorne, and C. J. Hawkey. 1991. “5-Aminosalicylic Acid Is a Potent Inhibitor of Interleukin 1 Beta Production in Organ Culture of Colonic Biopsy Specimens from Patients with Inflammatory Bowel Disease.” *Gut* 32 (1): 50–54. https://doi.org/10.1136/gut.32.1.50.

Marchesi, Julian R., Bas E. Dutilh, Neil Hall, Wilbert H. M. Peters, Rian Roelofs, Annemarie Boleij, and Harold Tjalsma. 2011. “Towards the Human Colorectal Cancer Microbiome.” *PLOS ONE* 6 (5): e20447. https://doi.org/10.1371/journal.pone.0020447.

Marquez Ruiz, Juan F., Kinga Kedziora, Brian Keogh, Jacqueline Maguire, Mary Reilly, Henry Windle, Dermot P. Kelleher, and John F. Gilmer. 2011. “A Double Prodrug System for Colon Targeting of Benzenesulfonamide COX-2 Inhibitors.” *Bioorganic & Medicinal Chemistry Letters* 21 (22): 6636–40. https://doi.org/10.1016/j.bmcl.2011.09.071.

Matsumoto, Ken’ichiro, Yuichi Mukai, Daiki Ogata, Fumi Shozui, John Masani Nduko, Seiichi Taguchi, and Toshihiko Ooi. 2010. “Characterization of Thermostable FMN-Dependent NADH Azoreductase from the Moderate Thermophile Geobacillus Stearothermophilus.” *Applied Microbiology and Biotechnology* 86 (5): 1431–38. https://doi.org/10.1007/s00253-009-2351-7.

McCoy, Amber N., Félix Araújo-Pérez, Andrea Azcárate-Peril, Jen Jen Yeh, Robert S. Sandler, and Temitope O. Keku. 2013. “Fusobacterium Is Associated with Colorectal Adenomas.” *PLOS ONE* 8 (1): e53653. https://doi.org/10.1371/journal.pone.0053653.

Mendes, Sónia, Luciana Pereira, Carlos Batista, and Lígia O. Martins. 2011. “Molecular Determinants of Azo Reduction Activity in the Strain Pseudomonas Putida MET94.” *Applied Microbiology and Biotechnology* 92 (2): 393–405. https://doi.org/10.1007/s00253-011-3366-4.

Mercier, C., V. Chalansonnet, S. Orenga, and C. Gilbert. 2013. “Characteristics of Major Escherichia Coli Reductases Involved in Aerobic Nitro and Azo Reduction.” *Journal of Applied Microbiology* 115 (4): 1012–22. https://doi.org/10.1111/jam.12294.

Misal, Santosh A., Devendra P. Lingojwar, Mahendra N. Lokhande, Pradeep D. Lokhande, and Kachru R. Gawai. 2014. “Enzymatic Transformation of Nitro-Aromatic Compounds by a Flavin-Free NADH Azoreductase from Lysinibacillus Sphaericus.” *Biotechnology Letters* 36 (1): 127–31. https://doi.org/10.1007/s10529-013-1338-8.

Misal, Santosh A., Devendra P. Lingojwar, Ravindra M. Shinde, and Kachru R. Gawai. 2011. “Purification and Characterization of Azoreductase from Alkaliphilic Strain Bacillus Badius.” *Process Biochemistry* 46 (6): 1264–69. https://doi.org/10.1016/j.procbio.2011.02.013.

Morrison, Jessica M., Cristee M. Wright, and Gilbert H. John. 2012. “Identification, Isolation and Characterization of a Novel Azoreductase from Clostridium Perfringens.” *Anaerobe*, Sydney M Finegold-90th birthday symposium, 18 (2): 229–34. https://doi.org/10.1016/j.anaerobe.2011.12.006.

Nachiyar, C. Valli, and G. Suseela Rajakumar. 2005. “Purification and Characterization of an Oxygen Insensitive Azoreductase from Pseudomonas Aeruginosa.” *Enzyme and Microbial Technology* 36 (4): 503–9. https://doi.org/10.1016/j.enzmictec.2004.11.015.

Nakamura, Joe, Yoshihiko Kubota, Masaaki Miyaoka, Toshihiko Saitoh, Fumio Mizuno, and Yoshimi Benno. 2002. “Comparison of Four Microbial Enzymes in Clostridia and Bacteroides Isolated from Human Feces.” *Microbiology and Immunology* 46 (7): 487–90. https://doi.org/10.1111/j.1348-0421.2002.tb02723.x.

Nakanishi, Masayuki, Chizuko Yatome, Norihisa Ishida, and Yukio Kitade. 2001. “Putative ACP Phosphodiesterase Gene (AcpD) Encodes an Azoreductase *.” *Journal of Biological Chemistry* 276 (49): 46394–99. https://doi.org/10.1074/jbc.M104483200.

Nielsen, O. H. 1982. “Sulfasalazine Intolerance.” *Scandinavian Journal of Gastroenterology* 17 (3): 389–93. https://doi.org/10.3109/00365528209182073.

Ooi, Toshihiko, Takeshi Shibata, Reiko Sato, Hiroaki Ohno, Shinichi Kinoshita, Tran Linh Thuoc, and Seiichi Taguchi. 2007. “An Azoreductase, Aerobic NADH-Dependent Flavoprotein Discovered from Bacillus Sp.: Functional Expression and Enzymatic Characterization.” *Applied Microbiology and Biotechnology* 75 (2): 377–86. https://doi.org/10.1007/s00253-006-0836-1.

Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression.” *Nature Methods* 14 (4): 417–19. https://doi.org/10.1038/nmeth.4197.

Peppercorn, Mark A., and Peter Goldman. 1972. “The Role of Intestinal Bacteria in the Metabolism of Salicylazosulfapyridine.” *Journal of Pharmacology and Experimental Therapeutics* 181 (3): 555–62.

———. 1973. “Distribution Studies of Salicylazosulfapyridine and Its Metabolites.” *Gastroenterology* 64 (2): 240–45. https://doi.org/10.1016/S0016-5085(73)80035-6.

Perrotta, Cristiana, Paolo Pellegrino, Eliana Moroni, Clara De Palma, Davide Cervia, Piergiorgio Danelli, and Emilio Clementi. 2015. “Five-Aminosalicylic Acid: An Update for the Reappraisal of an Old Drug.” *Gastroenterology Research and Practice* 2015 (January): e456895. https://doi.org/10.1155/2015/456895.

Proctor, Lita, and Curtis Huttenhower. 2019. “The Integrative Human Microbiome Project.” *Nature* 569 (7758): 641–48. https://doi.org/10.1038/s41586-019-1238-8.

Rachmilewitz, D., F. Karmeli, L. W. Schwartz, and P. L. Simon. 1992. “Effect of Aminophenols (5-ASA and 4-ASA) on Colonic Interleukin-1 Generation.” *Gut* 33 (7): 929–32. https://doi.org/10.1136/gut.33.7.929.

Raffi, Fatemeh, and Carl E. Cerniglia. 1990. “An Anaerobic Nondenaturing Gel Assay for the Detection of Azoreductase from Anaerobic Bacteria.” *Journal of Microbiological Methods* 12 (2): 139–48. https://doi.org/10.1016/0167-7012(90)90024-Z.

Rafii, F., and C. E. Cerniglia. 1995. “Reduction of Azo Dyes and Nitroaromatic Compounds by Bacterial Enzymes from the Human Intestinal Tract.” *Environmental Health Perspectives* 103 (suppl 5): 17–19. https://doi.org/10.1289/ehp.95103s417.

Rafii, F, W Franklin, and C E Cerniglia. 1990. “Azoreductase Activity of Anaerobic Bacteria Isolated from Human Intestinal Microflora.” *Applied and Environmental Microbiology* 56 (7): 2146–51. https://doi.org/10.1128/aem.56.7.2146-2151.1990.

Rubinstein, Mara Roxana, Xiaowei Wang, Wendy Liu, Yujun Hao, Guifang Cai, and Yiping W. Han. 2013. “Fusobacterium Nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via Its FadA Adhesin.” *Cell Host & Microbe* 14 (2): 195–206. https://doi.org/10.1016/j.chom.2013.07.012.

Ruiz, Juan F Marquez, Kinga Kedziora, Henry Windle, Dermot P Kelleher, and John F Gilmer. 2011. “Investigation into Drug Release from Colon-Specific Azoreductase-Activated Steroid Prodrugs Using in-Vitro Models.” *Journal of Pharmacy and Pharmacology* 63 (6): 806–16. https://doi.org/10.1111/j.2042-7158.2011.01289.x.

Ryan, Ali, Elise Kaplan, Jean-Christophe Nebel, Elena Polycarpou, Vincenzo Crescente, Edward Lowe, Gail M. Preston, and Edith Sim. 2014. “Identification of NAD(P)H Quinone Oxidoreductase Activity in Azoreductases from P. Aeruginosa: Azoreductases and NAD(P)H Quinone Oxidoreductases Belong to the Same FMN-Dependent Superfamily of Enzymes.” *PLOS ONE* 9 (6): e98551. https://doi.org/10.1371/journal.pone.0098551.

Ryan, Ali, Nicola Laurieri, Isaac Westwood, Chan-Ju Wang, Edward Lowe, and Edith Sim. 2010. “A Novel Mechanism for Azoreduction.” *Journal of Molecular Biology* 400 (1): 24–37. https://doi.org/10.1016/j.jmb.2010.04.023.

Ryan, Ali, Chan-Ju Wang, Nicola Laurieri, Isaac Westwood, and Edith Sim. 2010. “Reaction Mechanism of Azoreductases Suggests Convergent Evolution with Quinone Oxidoreductases.” *Protein & Cell* 1 (8): 780–90. https://doi.org/10.1007/s13238-010-0090-2.

Sousa, Tiago, Vipul Yadav, Vanessa Zann, Anders Borde, Bertil Abrahamsson, and Abdul W. Basit. 2014. “On the Colonic Bacterial Metabolism of Azo-Bonded Prodrugsof 5-Aminosalicylic Acid.” *Journal of Pharmaceutical Sciences* 103 (10): 3171–75. https://doi.org/10.1002/jps.24103.

Spanogiannopoulos, Peter, Elizabeth N. Bess, Rachel N. Carmody, and Peter J. Turnbaugh. 2016. “The Microbial Pharmacists within Us: A Metagenomic View of Xenobiotic Metabolism.” *Nature Reviews. Microbiology* 14 (5): 273–87. https://doi.org/10.1038/nrmicro.2016.17.

“SRA-Tools - NCBI.” n.d. Accessed August 18, 2021. http://ncbi.github.io/sra-tools/.

SUGIURA, Wataru, Tomoko YODA, Takashi MATSUBA, Yoshinori TANAKA, and Yasuhiko SUZUKI. 2006. “Expression and Characterization of the Genes Encoding Azoreductases from Bacillus Subtilis and Geobacillus Stearothermophilus.” *Bioscience, Biotechnology, and Biochemistry* 70 (7): 1655–65. https://doi.org/10.1271/bbb.60014.

Suzuki, Hirokazu. 2019. “Remarkable Diversification of Bacterial Azoreductases: Primary Sequences, Structures, Substrates, Physiological Roles, and Biotechnological Applications.” *Applied Microbiology and Biotechnology* 103 (10): 3965–78. https://doi.org/10.1007/s00253-019-09775-2.

Suzuki, Yasuhiko, Tomoko Yoda, Amin Ruhul, and Wataru Sugiura. 2001. “Molecular Cloning and Characterization of the Gene Coding for Azoreductase from Bacillus Sp. OY1-2 Isolated from Soil *.” *Journal of Biological Chemistry* 276 (12): 9059–65. https://doi.org/10.1074/jbc.M008083200.

Tozaki, Hideyuki, Tomokazu Odoriba, Naoki Okada, Takuya Fujita, Akira Terabe, Tsutomu Suzuki, Susumu Okabe, Shozo Muranishi, and Akira Yamamoto. 2002. “Chitosan Capsules for Colon-Specific Drug Delivery: Enhanced Localization of 5-Aminosalicylic

Acid in the Large Intestine Accelerates Healing of TNBS-Induced Colitis in Rats." *Journal of Controlled Release* 82 (1): 51–61. https://doi.org/10.1016/S0168-3659(02)00084-6.

Wallace, Bret D., Hongwei Wang, Kimberly T. Lane, John E. Scott, Jillian Orans, Ja Seol Koo, Madhukumar Venkatesh, et al. 2010. "Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme." *Science*, November. https://www.science.org/doi/abs/10.1126/science.1191175.

Weber, Christoph K., Susanne Liptay, Thomas Wirth, Guido Adler, and Roland M. Schmid. 2000. "Suppression of NF-KB Activity by Sulfasalazine Is Mediated by Direct Inhibition of IκB Kinases α and β." *Gastroenterology* 119 (5): 1209–18. https://doi.org/10.1053/gast.2000.19458.

Wickham, Hadley. 2011. "Ggplot2." *WIREs Computational Statistics* 3 (2): 180–85. https://doi.org/10.1002/wics.147.

Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. "Improved Metagenomic Analysis with Kraken 2." *Genome Biology* 20 (1): 257. https://doi.org/10.1186/s13059-019-1891-0.

Wu, Ai-Wen, Jin Gu, Jia-Fu Ji, Zhen-Fu Li, and Guang-Wei Xu. 2003. "Role of COX-2 in Carcinogenesis of Colorectal Cancer and Its Relationship with Tumor Biological Characteristics and Patients' Prognosis." *World Journal of Gastroenterology* 9 (9): 1990–94. https://doi.org/10.3748/wjg.v9.i9.1990.

Wu, William Ka Kei, Joseph Jao Yiu Sung, Chung Wa Lee, Jun Yu, and Chi Hin Cho. 2010. "Cyclooxygenase-2 in Tumorigenesis of Gastrointestinal Cancers: An Update on the Molecular Mechanisms." *Cancer Letters* 295 (1): 7–16. https://doi.org/10.1016/j.canlet.2010.03.015.

Xu, Haiyan, Thomas M. Heinze, Donald D. Paine, Carl E. Cerniglia, and Huizhong Chen. 2010. "Sudan Azo Dyes and Para Red Degradation by Prevalent Bacteria of the Human Gastrointestinal Tract." *Anaerobe* 16 (2): 114–19. https://doi.org/10.1016/j.anaerobe.2009.06.007.

Zhang, Xia, I-Son Ng, and Jo-Shu Chang. 2016. "Cloning and Characterization of a Robust Recombinant Azoreductase from Shewanella Xiamenensis BC01." *Journal of the Taiwan Institute of Chemical Engineers* 61 (April): 97–105. https://doi.org/10.1016/j.jtice.2016.01.002.

Zimmermann, Michael, Maria Zimmermann-Kogadeeva, Rebekka Wegmann, and Andrew L. Goodman. 2019. "Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes." *Nature* 570 (7762): 462–67. https://doi.org/10.1038/s41586-019-1291-3.

Zimmermann, Thomas, Hans G. Kulla, and Thomas Leisinger. 1982. "Properties of Purified Orange II Azoreductase, the Enzyme Initiating Azo Dye Degradation by Pseudomonas KF46." *European Journal of Biochemistry* 129 (1): 197–203. https://doi.org/10.1111/j.1432-1033.1982.tb07040.x.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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 inflammato1ry 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).

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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 (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

strain_name	FC	FC_STD	p_FDR	pct_consumed	pct_consumed_STD	SR_status	PR_status	result
Akkermansia muciniphila ATCCBAA-835	-0.419	0.361	0.19	25.222	18.707	nonSR	nonPR	TN
Alistipes indistinctus DSM 22520	-9.01	0.11	0.003	99.806	0.015	SR	PR	TP
Anaerococcus hydrogenalis DSM7454	-8.826	0.198	0.008	99.78	0.03	SR	nonPR	FN
Anaerotruncus colihominis DSM17241	-8.088	1.575	0.016	99.633	0.401	SR	nonPR	FN
Bacteroides caccae ATCC43185	-9.247	0.262	0.008	99.835	0.03	SR	PR	TP
Bacteroides cellulosilyticus DSM14838	-1.502	0.308	0.002	64.691	7.532	SR	PR	TP
Bacteroides coprophilus DSM18228	-0.006	0.216	0.991	0.386	14.918	nonSR	nonPR	TN
Bacteroides dorei DSM17855	-4.631	0.748	0.013	95.965	2.091	SR	nonPR	FN
Bacteroides eggerthii DSM20697	-9.59	0.104	0.002	99.87	0.009	SR	nonPR	FN
Bacteroides finegoldii DSM17565	-0.79	0.296	0.042	42.176	11.877	SR	nonPR	FN
Bacteroides fragilis 3397 T10	-0.974	0.348	0.009	49.084	12.282	SR	PR	TP
Bacteroides fragilis ATCC43859	-10.576	0.088	0.003	99.934	0.004	SR	PR	TP

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

Bacteroides fragilis DS-208	-9.327	0.57	0.006	99.844	0.062	SR	PR	TP
Bacteroides fragilis HMW610	-10.522	0.159	0.004	99.932	0.007	SR	PR	TP
Bacteroides fragilis HMW615	-10.398	0.256	0.013	99.926	0.013	SR	PR	TP
Bacteroides fragilis NCTC9343	-6.184	0.835	0.006	98.625	0.796	SR	PR	TP
Bacteroides fragilis T(B)9	-9.252	0.144	0.005	99.836	0.016	SR	PR	TP
Bacteroides intestinalis DSM17393	-1.296	0.398	0.005	59.267	11.246	SR	PR	TP
Bacteroides ovatus ATCC8483	-0.285	0.473	0.546	17.951	26.915	nonSR	PR	FP
Bacteroides pectinophilus ATCC43243	-0.249	0.241	0.268	15.832	14.042	nonSR	nonPR	TN
Bacteroides stercoris ATCC43183	-0.588	0.326	0.053	33.46	15.033	nonSR	nonPR	TN
Bacteroides thetaiotaomicron 3731	-1.303	0.248	0.005	59.461	6.961	SR	PR	TP
Bacteroides thetaiotaomicron 7330	-1.032	0.237	0.003	51.082	8.044	SR	PR	TP
Bacteroides thetaiotaomicron VPI-5482	-1.252	0.155	0.006	58.006	4.511	SR	PR	TP
Bacteroides uniformis ATCC8492	-2.605	0.389	0.001	83.558	4.436	SR	nonPR	FN
Bacteroides xylanisolvens DSM18836	-9.663	0.115	0.002	99.877	0.01	SR	PR	TP
Bifidobacterium adolescentis ATCC15703	-0.7	0.271	0.014	38.442	11.584	SR	nonPR	FN
Bifidobacterium breve DSM20213	-9.241	0.182	0.008	99.835	0.021	SR	nonPR	FN
Blautia hansenii DSM20583	-9.234	0.155	0.005	99.834	0.018	SR	nonPR	FN
Bryantia formatexigens DSM14469	-0.722	0.432	0.113	39.359	18.171	nonSR	PR	FP
Clostridium asparagiforme DSM15981	-9.184	0.203	0.01	99.828	0.024	SR	nonPR	FN
Clostridium bolteae ATCCBAA-613	-7.113	0.611	0.005	99.277	0.306	SR	nonPR	FN
Clostridium difficile 120	-6.039	0.703	0.013	98.479	0.741	SR	PR	TP
Clostridium ramosum DSM1402	-9.046	0.169	0.006	99.811	0.022	SR	PR	TP
Clostridium scindens ATCC35704	-9.064	0.324	0.004	99.813	0.042	SR	nonPR	FN
Clostridium spiroforme DSM1552	-2.478	0.572	0.001	82.055	7.119	SR	nonPR	FN
Clostridium sporogenes ATCC15579	-9.072	0.158	0.006	99.814	0.02	SR	PR	TP
Clostridium symbiosum ATCC14940	-9.256	0.209	0.009	99.836	0.024	SR	nonPR	FN
Collinsella aerofaciens ATCC25986	-6.629	0.628	0.046	98.99	0.44	SR	PR	TP
Collinsella intestinalis DSM13280	-8.916	0.118	0.003	99.793	0.017	SR	nonPR	FN
Coprococcus comes ATCC27758	-9.182	0.199	0.009	99.828	0.024	SR	nonPR	FN
Dorea formicigenerans ATCC27755	-4.071	0.57	0.015	94.051	2.351	SR	nonPR	FN
Edwardsiella tarda ATCC23685	-0.107	0.287	0.722	7.125	18.483	nonSR	PR	FP
Eggerthella lenta ATCC25559	-0.435	0.217	0.038	26.042	11.107	SR	PR	TP
Enterobacter cancerogenus ATCC35316	-0.853	0.579	0.076	44.621	22.214	nonSR	PR	FP
Enterococcus faecalis V583	-8.61	0.228	0.01	99.744	0.04	SR	PR	TP
Escherichia coli K-12	-0.714	0.331	0.038	39.057	13.995	SR	PR	TP
Eubacterium biforme DSM3989	-8.927	0.098	0.008	99.795	0.014	SR	nonPR	FN
Eubacterium hallii DSM3353	-9.031	0.381	0.028	99.809	0.05	SR	nonPR	FN
Eubacterium rectale ATCC33656	-9.002	0.322	0.022	99.805	0.044	SR	PR	TP
Eubacterium ventriosum ATCC27560	-4.653	0.827	0.051	96.024	2.278	nonSR	nonPR	TN

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

Odoribacter splanchnius	-7.892	0.744	0.004	99.579	0.217	SR	PR	TP
Parabacteroides distasonis ATCC8503	-1.007	0.298	0.022	50.253	10.272	SR	nonPR	FN
Parabacteroides johnsonii DSM18315	-0.529	0.281	0.123	30.702	13.51	nonSR	nonPR	TN
Parabacteroides merdae ATCC43184	-0.749	0.208	0.006	40.508	8.593	SR	nonPR	FN
Pretovella copri DSM18205	-8.693	0.41	0.015	99.758	0.069	SR	nonPR	FN
Proteus penneri ATCC35198	-1.657	0.313	0.009	68.289	6.884	SR	PR	TP
Providencia alcalifaciens DSM30120	-0.379	0.247	0.108	23.094	13.183	nonSR	PR	FP
Providencia rettgeri DSM1131	-0.205	0.204	0.25	13.245	12.273	nonSR	PR	FP
Providencia stuartii ATCC25827	-0.072	0.246	0.807	4.854	16.212	nonSR	PR	FP
Roseburia intestinalis L1-82	-8.876	0.156	0.006	99.787	0.023	SR	nonPR	FN
Ruminococcus gnavus ATCC29149	-8.639	0.045	0	99.749	0.008	SR	nonPR	FN
Ruminococcus lactaris ATCC29176	-9.522	0.145	0.005	99.864	0.014	SR	nonPR	FN
Ruminococcus torques ATCC27756	-2.384	0.298	0.016	80.841	3.961	SR	PR	TP
Salmonella Typhimurium LT2	-0.673	0.712	0.252	37.287	30.936	nonSR	PR	FP
Subdoligranulum variabile DSM15176	-8.854	0.189	0.008	99.784	0.028	SR	nonPR	FN
Victivallis vadensis ATCC BAA-548	-1.844	0.392	0.004	72.146	7.574	SR	nonPR	FN

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.

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.

	PR	nonPR
SR	38.8% (26/67) [a]	40.3% (27/67) [c]
nonSR	11.9% (8/67) [b]	9.0% (6/67) [d]

Figure 1

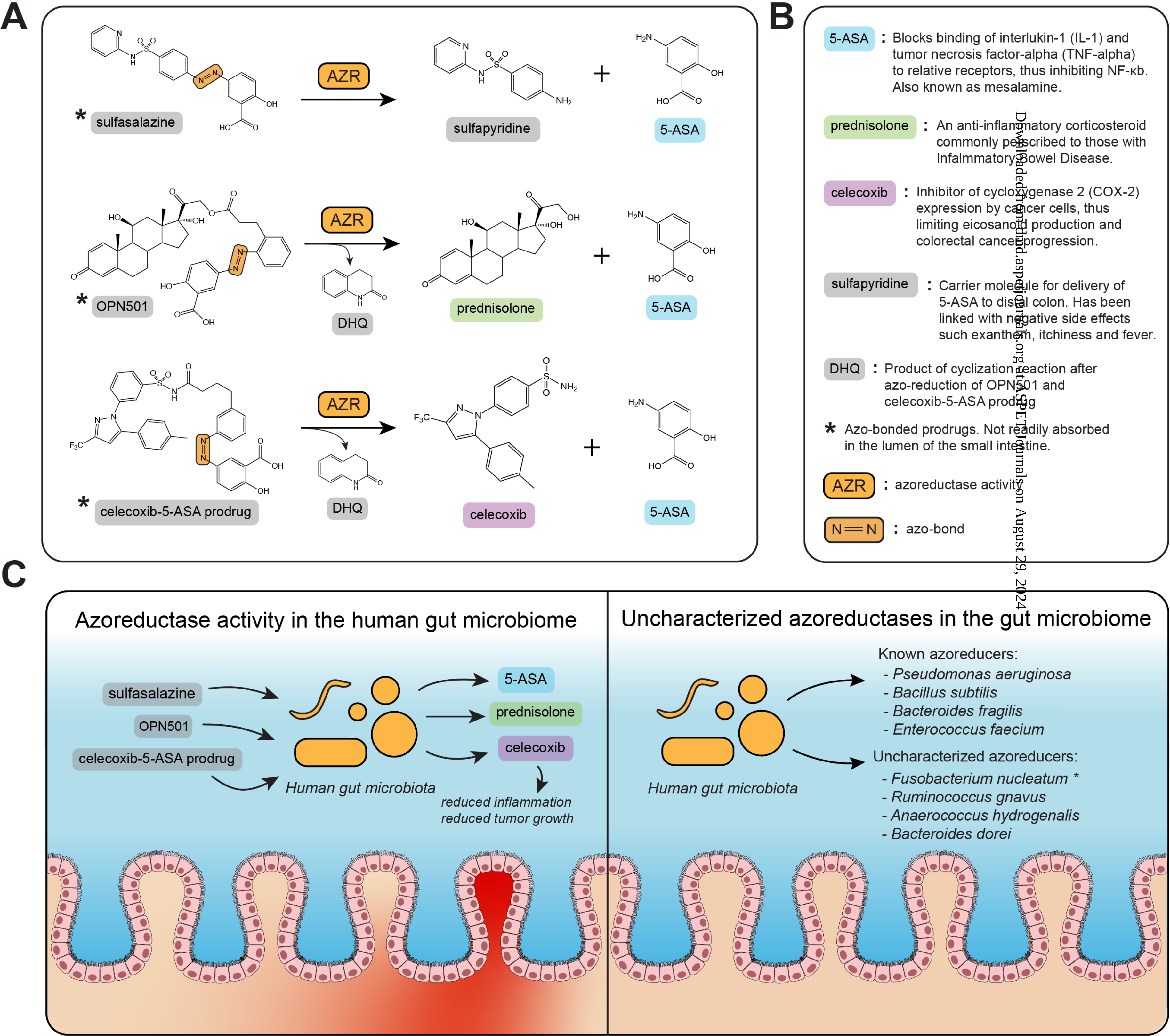
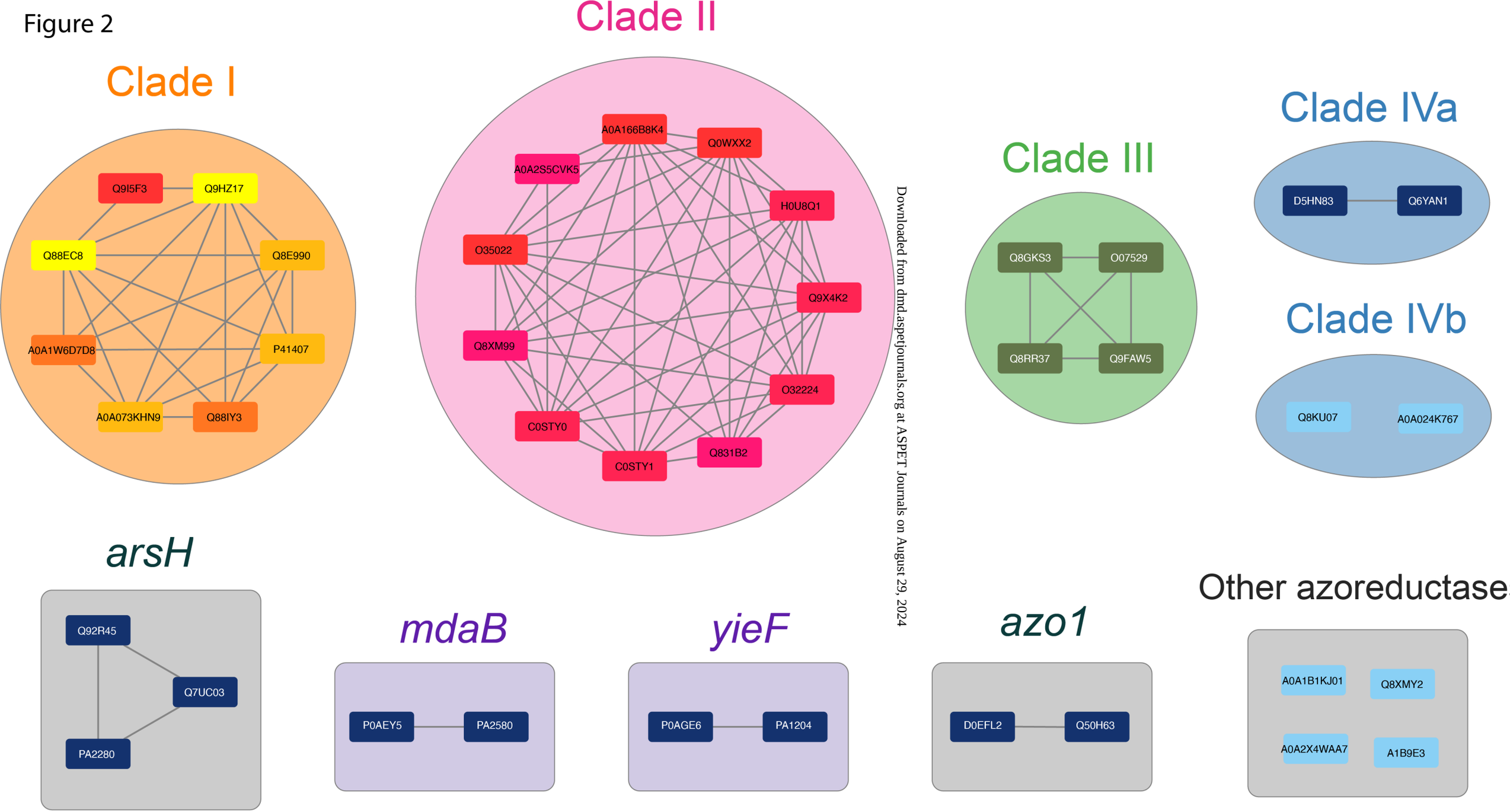


Figure 2



- Cofactor: FMN; Preferred e^- donor: NADPH
- Cofactor: FMN; Preferred e^- donor: NADH
- Cofactor: FMN; Preferred e^- donor: NAD(P)H
- Cofactor: Flavin-free; Preferred e^- donor: NADH
- Cofactor: FAD; Preferred e^- donor: NADPH
- Uncharacterized

Figure 3

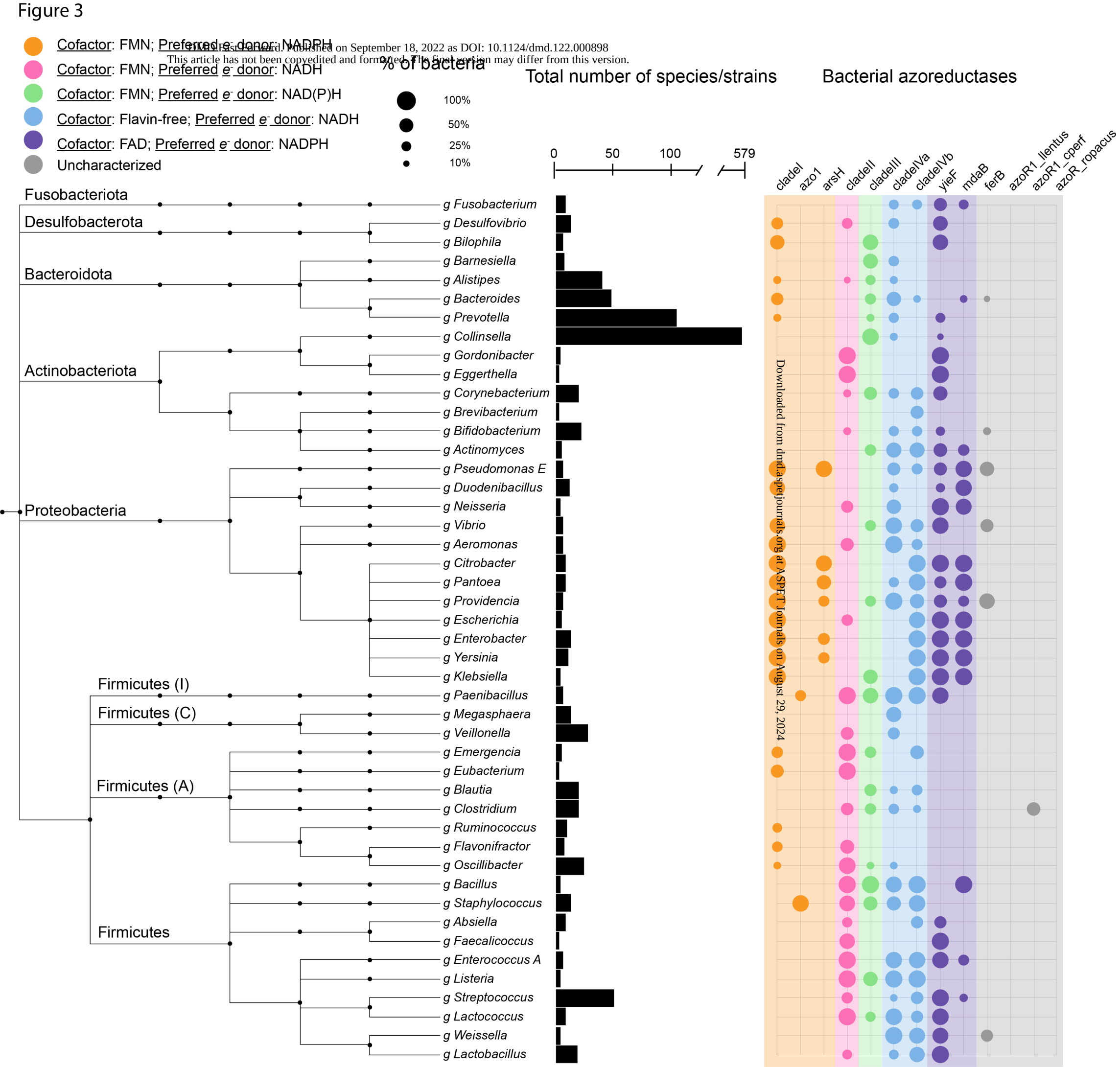


Figure 4

Figure 5

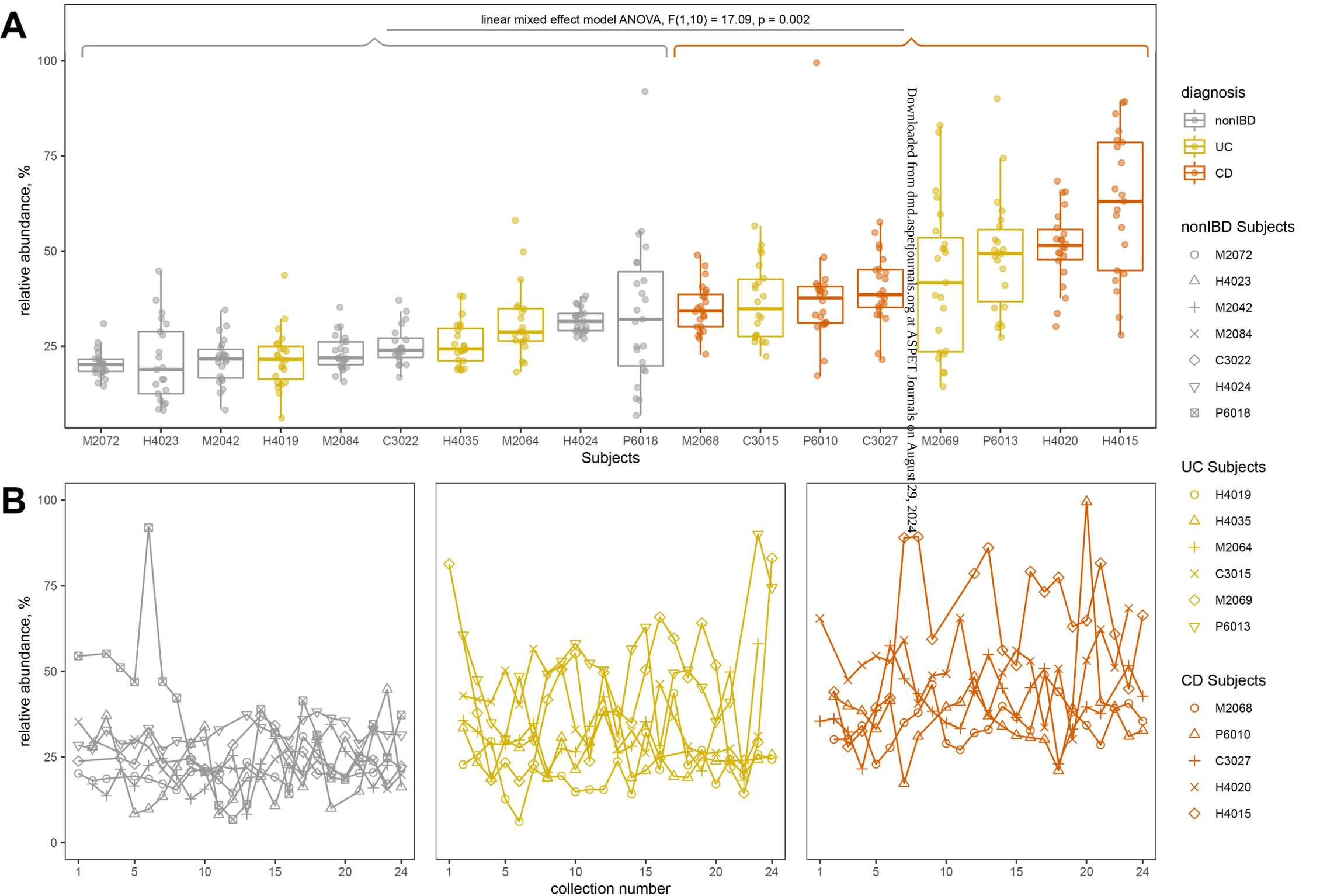


Figure 6

DMD Fast Forward. Published on September 18, 2022 as DOI: 10.1124/dmd.122.000898
This article has not been copyedited and formatted. The final version may differ from this version.