Dysbiosis in Metabolic Genes of the Gut Microbiomes of Patients with an Ileo-anal Pouch Resembles That Observed in Crohn's Disease

Vadim Dubinsky,[a] Leah Reshef,[a] Keren Rabinowitz,[b,c] Karin Yadgar,[b,c] Lihi Godny,[b,c] Keren Zonensain,[b,c] Nir Wasserberg,[d,e] Iris Dotan,[b,d] (iD) Uri Gophna[a]

[a]Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel Aviv, Israel
[b]Division of Gastroenterology, Rabin Medical Center, Petah-Tikva, Israel
[c]Felsenstein Medical Research Center, Rabin Medical Center, Petah Tikva, Israel
[d]Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel
[e]Colorectal Unit, Division of Surgery, Rabin Medical Center, Petah-Tikva, Israel

ABSTRACT Crohn's disease (CD), ulcerative colitis (UC), and pouchitis are multifactorial and chronic inflammatory bowel diseases (IBD). Pouchitis develops in former UC patients after proctocolectomy and ileal-pouch-anal anastomosis and is characterized by inflammation of the previously normal small intestine comprising the pouch. The extent to which microbial functional alteration (dysbiosis) in pouchitis resembles that of CD or UC has not been investigated, and the pathogenesis of pouchitis remains unknown. We collected 208 fecal metagenomes from 69 patients with a pouch (normal pouch and pouchitis) and compared them to publicly available metagenomes of patients with CD ($n = 88$), patients with UC ($n = 76$), and healthy controls ($n = 56$). Patients with pouchitis presented the highest alterations in species, metabolic pathways, and enzymes, which was correlated with intestinal inflammation. *Ruminococcus gnavus* strains encoding mucin-degrading glycoside hydrolases were highly enriched in pouchitis. Butyrate and secondary bile acid biosynthesis pathways were decreased in IBD phenotypes and were especially low in pouchitis. Pathways such as amino acid biosynthesis and degradation of aromatic compounds and sugars, encoded by members of the *Enterobacteriaceae*, were enriched in pouch and CD but not in UC. We developed microbial feature-based classifiers that can distinguish between patients with a normal pouch and pouchitis and identified species and genes that were predictive of pouchitis. We propose that the noninflamed pouch is already dysbiotic and microbially is similar to CD. Our study reveals microbial functions that outline the pathogenesis of pouchitis and suggests bacterial groups and functions that could be targeted for intervention to attenuate small intestinal inflammation present in pouchitis and CD.

IMPORTANCE Crohn's disease (CD), ulcerative colitis (UC), and pouchitis are chronic inflammatory conditions of the bowel. Pouchitis develops in former UC patients after proctocolectomy and ileal-pouch-anal anastomosis and is characterized by inflammation of the previously normal small intestine comprising the pouch. The extent to which microbial dysbiosis in patients with pouchitis resembles that of CD or UC and the pathogenesis of pouchitis remains unclear. We investigated the functions in the gut microbiomes of these patients using metagenomics. We found that the noninflamed pouch is already dysbiotic and microbially is similar to CD. Our study reveals microbial functions with a potential role in pouchitis pathogenesis such as depletion in butyrate and secondary bile acid synthesis and enrichment of amino acid synthesis and degradation of aromatic compounds, encoded by members of the *Enterobacteriaceae*. We developed microbial feature-based classifiers that can distinguish between patients

with a normal pouch and pouchitis and identified species and genes that were predictive of pouchitis. We suggest species and functions that could be targeted for intervention to attenuate small intestinal inflammation present in pouchitis and CD.

KEYWORDS pouchitis, UC, CD, mucin, butyrate, bile acids, oxidative stress, classifier, Crohn's disease, microbiome, ulcerative colitis

Inflammatory bowel diseases (IBD), including Crohn's disease (CD), ulcerative colitis (UC), and pouchitis, are chronic, relapsing, and remitting inflammatory conditions of the intestine. CD mainly affects the small intestine and colon, while UC is localized to the colon. The etiology of IBD is multifactorial, involving genetic predisposition (1), environmental triggers (2), and an aberrant immune-mediated response to specific antigens of the intestinal microbiota (3).

Approximately 25% of patients with complicated UC may undergo total large bowel resection. Reconstruction of intestinal continuity is achieved by the creation of a reservoir ("pouch") from the healthy small intestine which is connected to the anus (ileal pouch-anal anastomosis, also called pouch surgery) (4). Up to 70% of these former UC patients may develop inflammation of the previously normal small intestine comprising the pouch (pouchitis) (5). Pouchitis is the most common complication after pouch surgery, but its pathogenesis is still largely unknown (6).

It was shown previously that the gut microbiome in patients with IBD is characterized by lower diversity, lower stability, and compositional changes in specific taxa compared to that in healthy individuals (7). Levels of facultative anaerobes from the *Enterobacteriaceae* are increased, while beneficial obligate anaerobes, such as *Faecalibacterium prausnitzii* and *Eubacterium rectale*, are decreased (8–12). Shifts in microbial functions in the metagenomes of patients with IBD have also been demonstrated, e.g., lower presence of short-chain fatty acid (SCFA) synthesis pathways, increase in oxidative stress response genes, and alteration in amino acid metabolism genes (9, 13). These taxonomic and functional changes are broadly referred to as dysbiosis.

Since most patients with a pouch will develop inflammation in what has been normal uninflamed small bowel tissue (5), these patients represent a unique condition modeling the development of *de novo* small intestinal inflammation, such as that characterizing ileal CD (14). We demonstrated previously that patients with a pouch and CD might have several mechanisms of pathogenesis in common, including mucosal gene and microRNA expression (14, 15), antiglycan serologic responses (16), and CD-like bacterial dysbiosis (11). However, the extent to which microbial dysbiosis in pouchitis resembles that of CD has not been investigated in depth. Here, we analyzed the shotgun metagenomes of patients with a pouch, patients with IBD, and healthy subjects with the following aims. (i) We wanted to identify microbial signatures (species and functional genes) that might be common to disease phenotypes and thus to place patients with a pouch in the context of "primary" IBD. More specifically, we wanted to test whether pouchitis has microbial features that are most similar to those of CD. (ii) We wanted to discover microbial features which might be predictive of pouchitis (untargeted approach). (iii) We wanted to examine key bacterial functions that may underlie the pathogenesis of pouchitis: excessive mucin degradation, reduced butyrate and secondary bile acids production, and increased bacterial resistance to oxidative stress.

RESULTS AND DISCUSSION

We enrolled 69 patients with a pouch in this study; the median age was 43.8 years (interquartile range, 33.8 to 62.2 years), and 60% were males (Data Set S1, sheet 1). Two hundred eight fecal samples were collected longitudinally (median of 3 samples per patient) over a period of up to 5.8 years. Based on our classification system, these patients were divided into two groups (discovery cohort): patients with a normal

pouch (noninflamed; $n = 35$ patients; 103 samples) and those with pouchitis (inflamed; $n = 34$; 105 samples). We also used 20 additional patients with a pouch from our center (Rabin Medical Center [RMC]) and 15 pouch patients from an American cohort (17) as a validation cohort (Data Set S1, sheet 2).

To explore the microbiomes of patients with a pouch in the context of primary IBD, we analyzed shotgun metagenomes derived from fecal samples of patients with a pouch, UC, and CD and healthy subjects in terms of species composition and functional gene repertoires. Metagenomes of patients with CD ($n = 88$), those with UC ($n = 76$), and healthy subjects ($n = 56$) were obtained from previously published data (18), which are part of the PRISM, LifeLines DEEP, and NLIBD cohorts (cross-sectional) (Data Set S1, sheet 3). To minimize potential biases, raw metagenomic reads from all cohorts were processed in parallel through the same bioinformatic pipeline, in line with recent recommendations for multicohort studies (19).

Microbial dysbiosis across disease phenotypes and its association with inflammation. The microbial community composition of disease phenotypes and that of healthy subjects (beta-diversity, analyzed by principal-coordinate analysis [PCoA]) were well separated according to profiles of species, genes, and pathways (Fig. 1A to C; Fig. S1A to C), with some degree of overlap, in particular between samples from UC and CD patients, as was observed previously (18). Strikingly, we observed a significant stratification (Kruskal-Wallis, $P < 0.01$) of different disease phenotypes based on the first axis of microbiome variation, shifting away from the healthy subjects, in the following order: UC, CD, normal pouch, and pouchitis (Fig. 1A to C). The latter was the most distant from the microbiomes of healthy controls; i.e., it had the highest level of dysbiosis. Interestingly, according to Shannon diversity (Fig. 1D), no significant difference was observed between patients with CD (median Shannon diversity index, 2.6) and UC (median Shannon diversity index, 2.87) to patients with a normal pouch (median Shannon diversity index, 2.73). In contrast, patients with pouchitis had significantly lower diversity (median Shannon diversity index, 2.31; Kruskal-Wallis, $P < 0.001$) than all other phenotypes.

The subject's clinical phenotype explained a substantial part of the observed variation in microbial composition (beta-diversity based on Bray-Curtis distance) in species, metabolic pathway, and enzymes (Fig. S1D) (permutational multivariate analysis of variance [PERMANOVA]; $R^2 = 12.1\%$, 13.2%, and 10.4%, respectively; $P < 0.001$), while patient age and gender (the latter was available only for the pouch cohort) explained only 0.39% to 0.71% and 0.42% to 0.52%, respectively. Antibiotic use (1 month or 6 months off antibiotics) accounted for only 0.45% to 0.62% of the variation. Fecal calprotectin, a biomarker for intestinal inflammation, explained a low yet significant degree of microbial variation ($R^2 = 2.5\%$ to 4.1%, $P < 0.001$). Consistent with previous studies of patients with UC and CD (13, 20), intersubject variation explained the highest degree of microbiome variation (Fig. S1D) (PERMANOVA; $R^2 = 46.6\%$ to 48.7%; $P < 0.001$) among the longitudinal samples of patients with a pouch cohort. The cohort from which the samples were derived had a considerable effect as well, explaining 8.7% to 11.2% of the variation. Finally, in patients with a pouch, pouch phenotype and disease activity contributed 4.2% to 4.8% and 1.4% to 1.9%, respectively, to the microbiome variation (Fig. S1D). These observations indicate that besides subject and cohort differences, disease type and inflammation severity are the strongest contributors to microbial variability and the resulting dysbiosis.

If dysbiosis is associated with intestinal inflammation, one would expect a correlation between the microbiome variation and inflammation biomarkers. Indeed, we observed a moderate and significant correlation between the fecal calprotectin level and the first axis of microbiome variation for species, metabolic pathways, and enzymes (Fig. 1E to G) (Spearman $r = 0.44$, 0.39, and 0.375, respectively; $P < 0.001$). A weaker negative correlation was detected between calprotectin and Shannon diversity (Fig. 1H) (Spearman $r = -0.186$; $P = 4.01 \times 10^{-4}$). These data suggest that patients with a normal pouch already show signs of inflammation and dysbiosis that may parallel those observed in CD and that these are intensified in patients with pouchitis. As
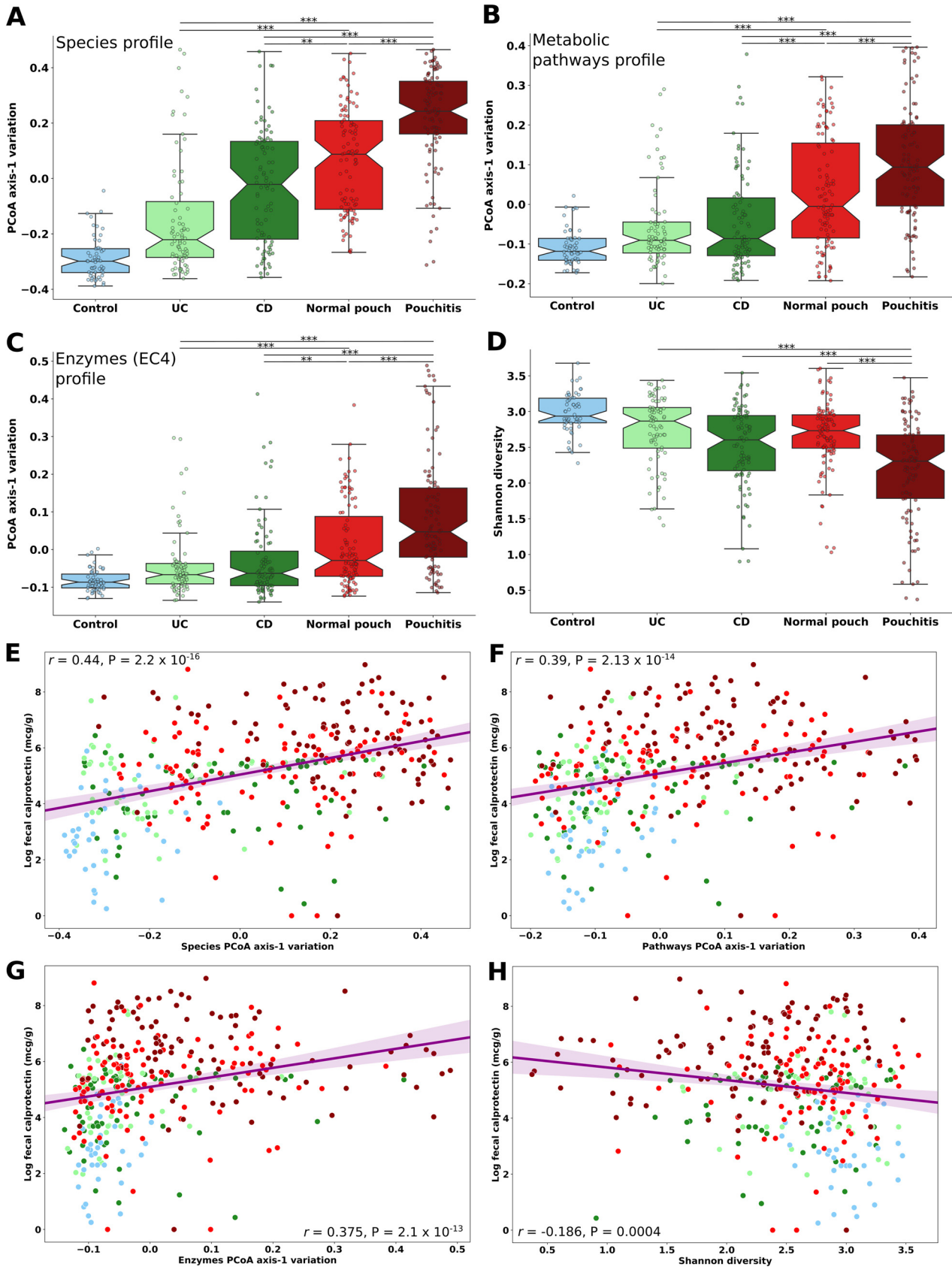
mSystems®

FIG 1 Microbial community stratification among IBD disease phenotypes and healthy subjects. (A to C) Microbiome variation according to the first axis of principal-coordinate analysis (PCoA) of (A) species, (B) metabolic pathways, and (C) enzyme profiles, based on Bray-Curtis dissimilarity

(Continued on next page)

patients with a pouch had the highest level of calprotectin (Fig. S2A), we wanted to assess how much of the observed dysbiosis in those patients was due to intestinal inflammation level or pouchitis. We divided these patients based on similar levels of calprotectin and analyzed the variation in microbial composition and diversity. Pouch phenotype was the only factor having a statistically significant effect on the microbiome regardless of the level of inflammation (Fig. S2B to E).

To rule out the possibility that the observed differences between the pouch patients' and nonpouch patients' microbiomes are inflated due to center or country effects, we obtained and analyzed 15 publicly available pouch metagenomes (17) (from an American cohort). Reassuringly, based on microbiome species, pathway, and enzyme variation (Bray-Curtis) and Shannon diversity (Fig. S3A to D), the pouch samples from the American cohort clustered within the range of the pouches from our RMC pouch cohort.

Bacterial species associated with IBD. We aimed to characterize the altered species composition in the four IBD groups compared to the composition in healthy subjects, and also to compare patients with a pouch to those with primary IBD, using linear modeling of shotgun metagenomic data. As expected, bacterial species composition showed extensive and significant differences involving many taxa between healthy subjects and patients with UC, with CD, and with a pouch (Fig. 2; Data Set S1, sheet 4). We confirmed past observations of decreased abundance in beneficial taxa such as *Faecalibacterium prausnitzii* and increased abundance of potential pathobionts such as *Escherichia coli* in CD and in patients with a pouch (8, 10–12). Patients with a pouch (in particular, those with pouchitis) showed the strongest extremes in terms of the levels of the above-mentioned species, further supporting the existence of a strong dysbiosis in pouchitis.

In addition, we identified significant depletion of an unclassified species of *Subdoligranulum*, a known butyrate-producing genus related to *Faecalibacterium* (21), in all disease phenotypes (Fig. 2), with the lowest mean relative abundance in patients with a normal pouch (0.64%) and those with pouchitis (0.22%). Additionally, *Ruminococcus bromii*, a keystone species that is highly efficient in the degradation of resistant starch, making its degradation products available for other beneficial bacteria via cross-feeding (22), was highly depleted in all disease phenotypes, especially in patients with a pouch (relative abundances of 0.75% and 0.3% in normal pouch and pouchitis, respectively). Interestingly, we found a trend for the enrichment of *Ruminococcus gnavus* and *Ruminococcus torques* in patients with a pouch and for the depletion of *Akkermansia muciniphila* (Fig. 2). These three species have the potential to degrade human secretory mucin, and an observation of an increase in *R. gnavus* and *R. torques* coupled with a decrease in *A. muciniphila* in biopsy specimens of UC and CD patients was previously reported (23). A higher proportion of mucolytic species in patients with IBD may contribute to increased intestinal inflammation (24). Notably, a distinct clade of *R. gnavus* strains was previously found to be enriched in CD and UC patients and harbored potential virulence-associated genes (25). To critically examine whether the mucin degradation potential of the microbiome is higher in IBD, we moved on to quantify the total mucin-degrading bacterial enzymes in the gut microbiomes of these patients.

Mucin degradation and *R. gnavus* glycoside hydrolases. We specifically analyzed glycoside hydrolase (GH) genes that are involved in the breakdown and utilization of mucin-derived glycans and are encoded in the genomes of mucolytic bacteria (Fig. S4). According to these criteria, bacterial community mucin degradation potential in patients with a normal pouch was similar to those of patients with CD and of healthy

FIG 1 Legend (Continued)
of the fecal metagenomes. (D) Shannon diversity in the gut microbiome based on species composition. (E to H) Spearman correlation calculated between fecal calprotectin level (log) and the first axis of microbiome variation for species (E), metabolic pathways (F), enzymes (G), and Shannon diversity (H). Lines are linear regression fitted with a 95% confidence interval. Colors represent the different IBD phenotypes (UC, CD, normal pouch, and pouchitis), as well as healthy subjects (control). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Kruskal-Wallis, applying Dunn's multiple correction test. For a list of all pairwise P values, see Data Set S1, sheet 12. Box plot whiskers mark observations within the 1.5 interquartile range of the upper and lower quartiles.
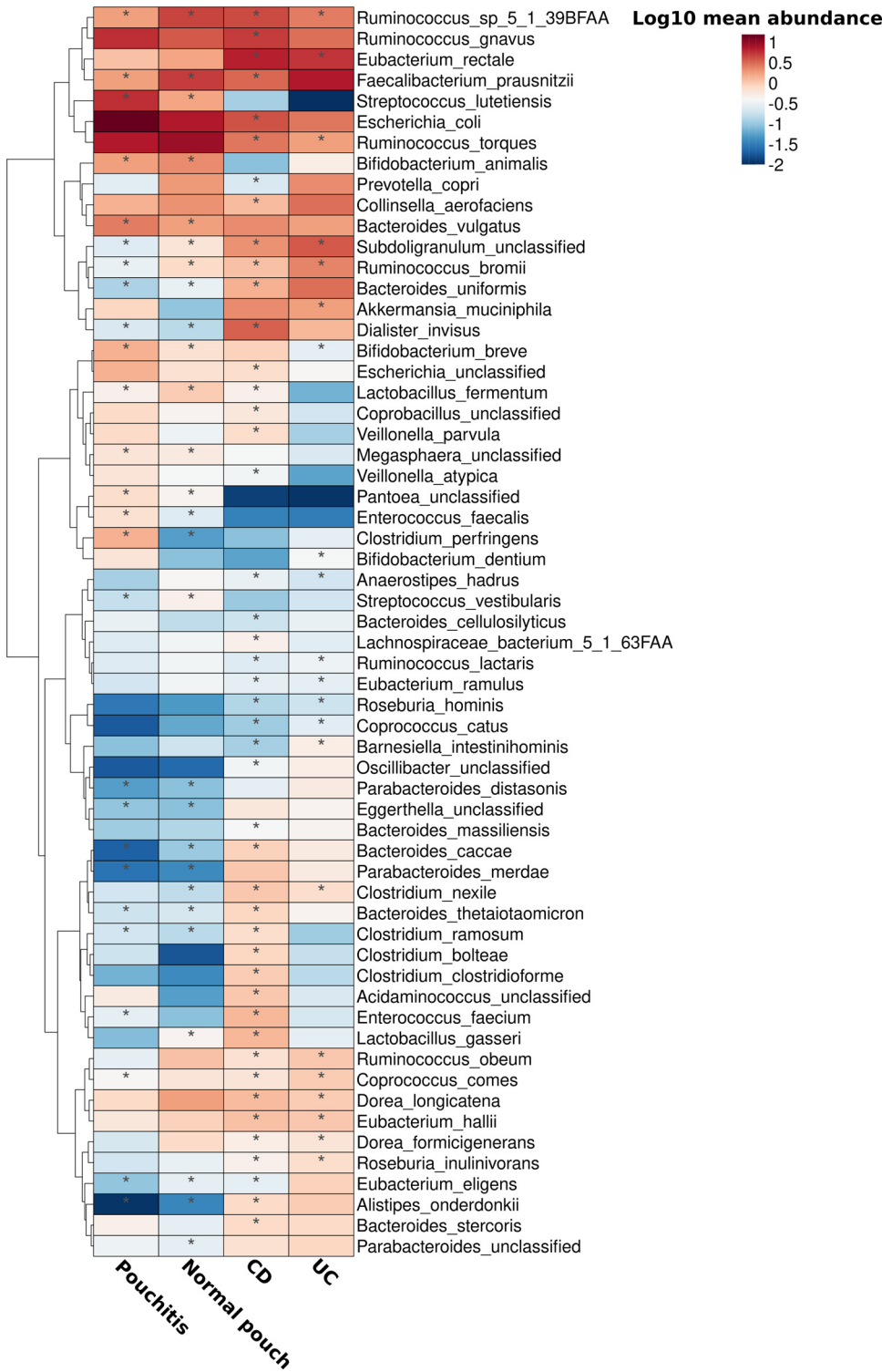
FIG 2 Bacterial species associated with IBD phenotypes. A generalized linear mixed-effects model was built for each bacterial species, and IBD phenotypes were set as a predictor (including age and antibiotic use as fixed effects). Healthy subjects were used as a reference point, and individual subjects and cohort were set as random effects. Only species with a mean relative abundance across all groups of ≥0.15% with at least one significant association with an IBD phenotype are presented. For a full list of species that were significantly associated with IBD phenotypes see Data Set S1, sheet 4. Asterisks indicate patient groups for which the association was significant (FDR < 0.05).
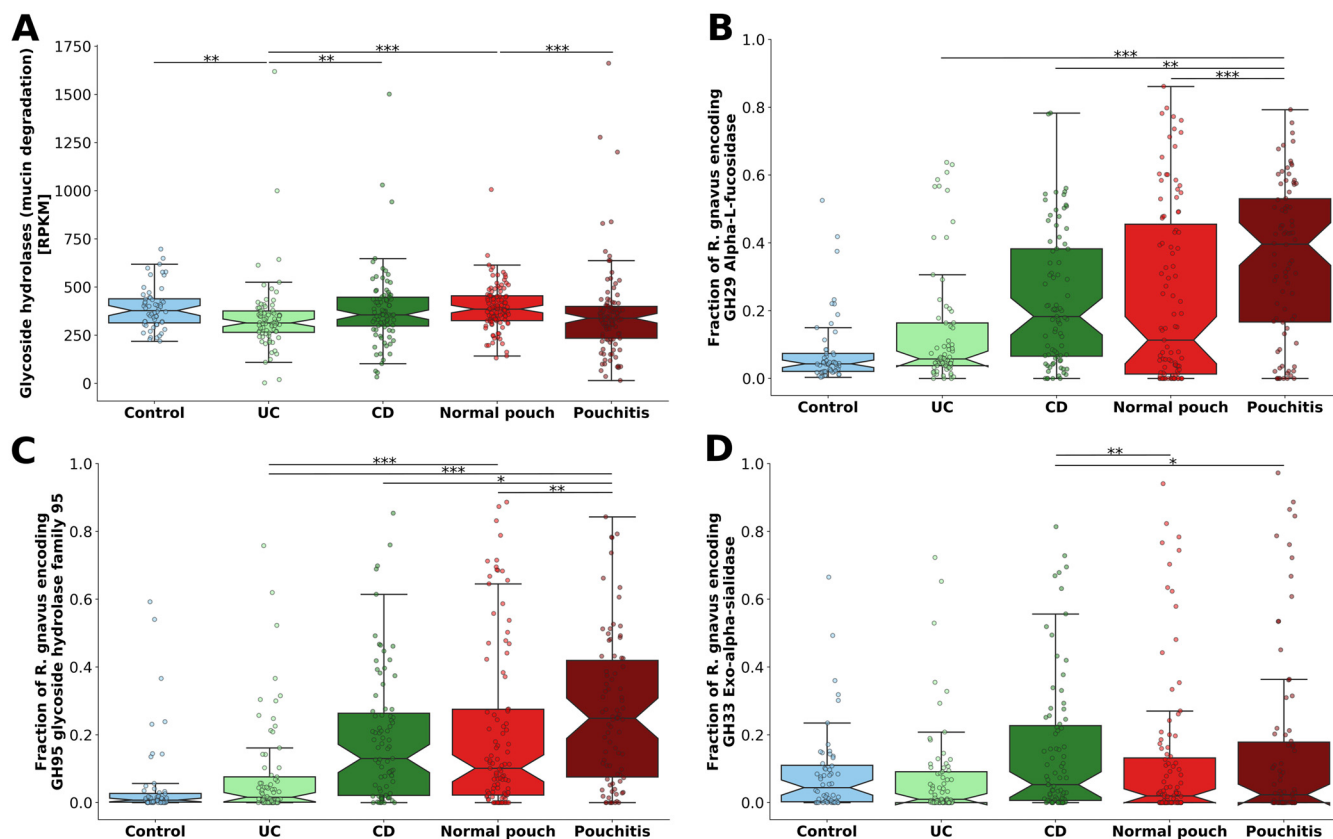
FIG 3 Mucin degradation potential of the microbiome of patients with IBD and healthy controls. (A) Bacterial community mucin degradation potential based on analysis of glycoside hydrolase genes (GH) that are involved in the breakdown and utilization of mucin-derived glycans. (B to D) The relative fraction of *Ruminococcus gnavus* strains encoding strain-specific glycoside hydrolases (normalized according to three single-copy housekeeping genes) involved in mucin degradation: (B) GH29 α-L-fucosidase, (C) GH95 glycoside hydrolase family 95, and (D) GH33 exo-alpha-sialidase. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Kruskal-Wallis, applying Dunn's multiple correction test. For a list of all pairwise P values, see Data Set S1, sheet 12. Box plot whiskers mark observations within the 1.5 interquartile range of the upper and lower quartiles.

subjects (median numbers of reads per kilobase of transcript per million mapped reads [RPKM], 386, 338, and 377, respectively; Fig. 3A). Patients with pouchitis had 12.3% lower community mucin degradation potential than patients with a normal pouch (Kruskal-Wallis, $P = 3.4 \times 10^{-4}$), while patients with UC had lower mucin degradation potential than the other phenotypes (Fig. 3A) (Kruskal-Wallis, $P < 0.01$). These results suggest that the mucin degradation potential of the microbiome does not increase in IBD compared to healthy subjects and even decreases in UC patients.

To better understand the specific role of *R. gnavus* in mucin degradation, we focused on its enzymes that are considered specific for host mucin utilization: GH29 (α-L-fucosidase), GH95 (glycoside hydrolase family 95), and GH33 (exo-alpha-sialidase). The former two can cleave fucose from host mucin glycans, while the latter can cleave terminal sialic acid from sialylated mucins. It was shown that the ability of *R. gnavus* to utilize host mucins is strain dependent and is attributed to a specific GH33 sialidase and to GH29 and GH95 fucosidases (26). Thus, we obtained sequences of GH29, GH95, and GH33 from the representative genome of *Ruminococcus gnavus* ATCC 29149 (JCM6515) and quantified the abundance of these genes and the fraction of *R. gnavus* strains that possess them in the metagenomes by read mapping (Fig. 3B to D). The fraction of *R. gnavus* strains encoding GH29 and GH95 fucosidases varied extensively across disease phenotypes (Fig. 3B and C) but was significantly higher in patients with pouchitis, with 39.6% and 25% (median) of the strains encoding these fucosidases, respectively (Kruskal-Wallis, $P < 0.01$). No significant difference was observed between patients with a normal pouch (11.3% and 10.1% [median]) and CD (18.3% and 13%

[median]) for GH29 and GH95, respectively. Less than 6% of *R. gnavus* strains in UC patients and in healthy subjects encoded these fucosidases. Interestingly, GH33 sialidase was encoded by a significantly smaller proportion of *R. gnavus* strains (Fig. 3D). Patients with a pouch had a similar fraction of *R. gnavus*-encoded GH33 sialidase (normal pouch, 2%, and pouchitis, 2.4% [medians]; Kruskal-Wallis, $P = 0.336$), which was comparable to values for UC patients and healthy subjects (Kruskal-Wallis, $P > 0.05$). Patients with CD had the highest abundance of *R. gnavus* strains encoding GH33 sialidase (5.3% [median]; Kruskal-Wallis, $P < 0.05$).

A study by Ruseler-van Embden et al. (27) established that the pH level in patients with a normal pouch (5.4) was significantly lower than in patients with pouchitis (6.5). Remarkably, this low pH in the normal pouch substantially inhibited mucin (including fucose residues) degradation. This might explain why we detected a markedly lower proportion of *R. gnavus* strains encoding fucosidases in patients with a normal pouch. Altogether, our findings demonstrate that as a consequence of strong dysbiosis in pouchitis, the overall bacterial community mucin degradation potential is lower, but there is a substantially higher number of mucin-degrading *R. gnavus* strains that might be proinflammatory. Mucin degradation in the normal pouch is similar to that of CD in terms of both microbial community- and *R. gnavus*-associated potential.

Butyrate synthesis potential of the microbiome is decreased in pouchitis. Following our finding that patients with a pouch had greatly reduced abundances of beneficial species, some of which are known SCFA producers, we aimed to directly quantify the genes encoding SCFA synthesis enzymes. SCFAs are the major metabolic products of anaerobic fermentation of nondigestible polysaccharides by the human colonic microbiota (28, 29). Here, we focused on butyrate, which is the preferred energy source for colonocytes, improves the gut barrier function, has immunomodulatory and anti-inflammatory properties, and reduces oxidative stress in the colon (28). Butyrate-producing species are found among many non-butyrate-producing species in two dominant families of *Firmicutes*, *Ruminococcaceae* and *Lachnospiraceae* (29). Four main pathways are known for butyrate production: the acetyl coenzyme A (acetyl-CoA), glutarate, 4-aminobutyrate, and lysine pathways. Acetyl-CoA is the dominant pathway for butyrate synthesis and is present in the majority of butyrate producers, and *but* is the most prevalent terminal gene (30).

We quantified and modeled the terminal genes of the four known butyrate production pathways—namely, *but*, *buk*, *ato*, and *4hbt*—in the metagenomes using an extensive and rigorously curated database for butyrate synthesis genes (31). Overall, patients with CD, normal pouch, and pouchitis had the lowest levels of butyrate synthesis pathways (Fig. 4A to D) ($P < 0.05$). Specifically, according to the *but* gene (Fig. 4A and E), the lowest abundance of these genes was in patients with pouchitis (median RPKM, 12), followed by significantly higher abundance in patients with a normal pouch (median RPKM, 78; Kruskal-Wallis, $P = 3.5 \times 10^{-6}$). No significant difference in butyrate synthesis genes between CD (median RPKM, 126) and UC (median RPKM, 136) was observed. Analysis of the *buk* homologs (Fig. 4B and F) revealed similar abundances between patients with CD (median RPKM, 6.6) and with normal pouches (median RPKM, 8.4), which were lower than those of patients with UC and healthy subjects (median RPKM, 14.7 and 25.3, respectively; Kruskal-Wallis, $P < 0.01$). Patients with pouchitis had the lowest levels (median RPKM, 4.75). The lysine (*ato*) and 4-aminobutyrate (*4hbt*) pathways were lower in patients with a pouch (Fig. 4C and D) and less abundant in general in all groups (Fig. 4G and H). Patients with a pouch had very low levels of both *ato* and *4hbt* genes (which were absent altogether in many samples), suggesting that the 4-aminobutyrate and lysine pathways for butyrate synthesis are exceptionally low in these patients. Similar trends in all four terminal genes were observed in pouch patients from the American cohort (Fig. S3E to H).

The taxonomic affiliation of butyrate-producing bacteria based on the similarity of the identified *but* and *buk* genes to those of known butyrate producers revealed a dominance of commensal *Firmicutes* such as *F. prausnitzii* and several species of
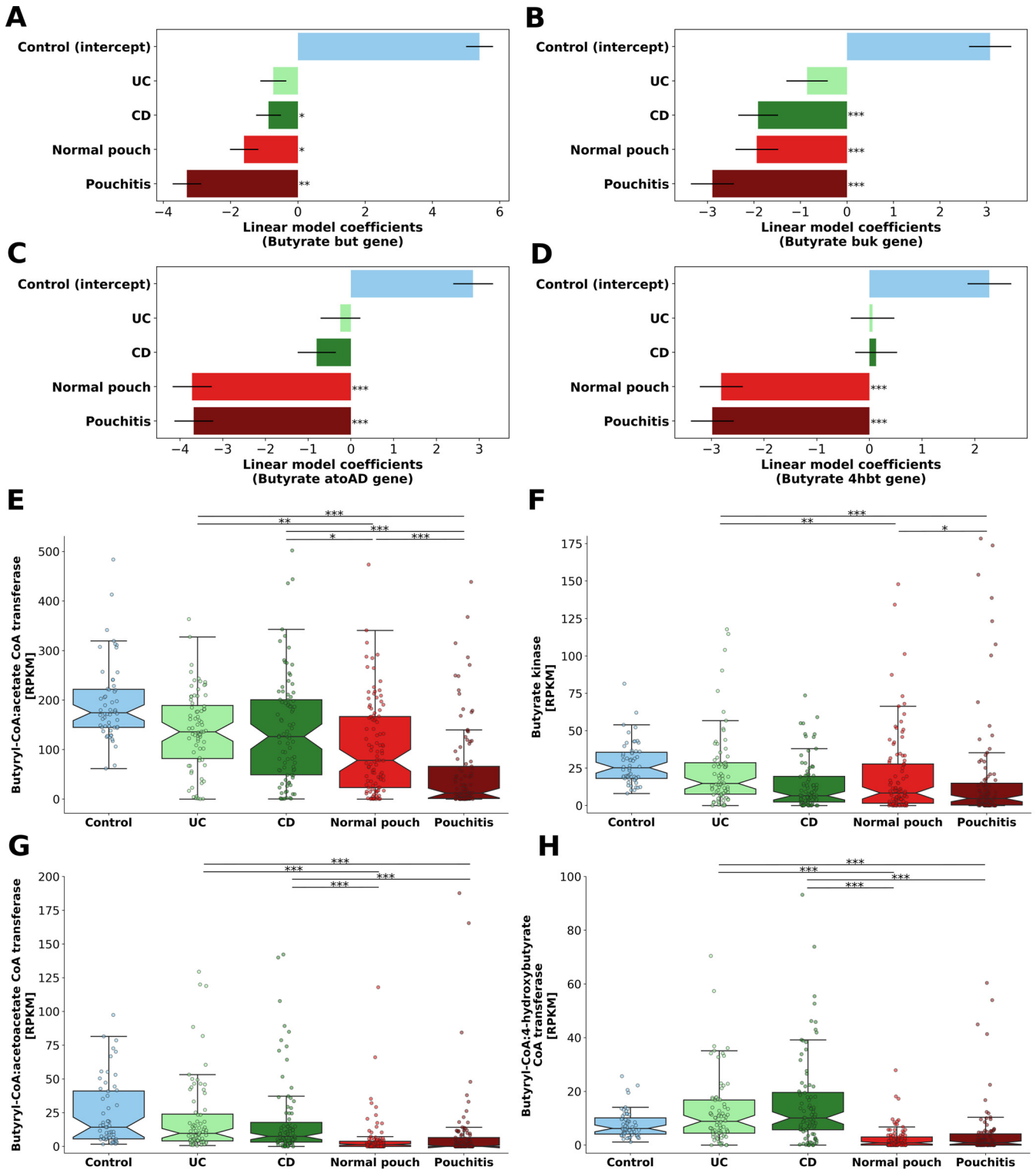
FIG 4 The main pathways of microbial butyrate synthesis in IBD and healthy subjects. Modeling and quantification of the terminal genes (*but*, *buk*, *ato*, and *4hbt*) for corresponding butyrate synthesis pathways in the metagenomes: (A and E) butyryl-CoA:acetate CoA transferase (*but*) pathway; (B and F) butyrate kinase (*buk*)–acetyl-CoA and glutarate pathways; (C and G) butyryl-CoA:acetoacetate CoA transferase (*ato*)–lysine pathway; (D and H) butyryl-CoA:4-hydroxybutyrate CoA transferase (*4hbt*)–4-aminobutyrate pathway. In panels A to D, linear model coefficients (effect sizes) are given for each phenotype for the modeled genes. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. In panels E to H, gene abundance is in RPKM units (reads per kilobase per million mapped reads), normalized according to the corresponding median gene length and total-sum scaled to the number of reads in each metagenome. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Kruskal-Wallis, applying Dunn's multiple correction test. For a list of all pairwise P values, see Data Set S1, sheet 12. Box plot whiskers mark observations within the 1.5 interquartile range of the upper and lower quartiles.

TABLE 1 Spearman correlation between dietary fiber intake and the abundance of butyrate genes (*but* and *buk*) in patients with pouch samples (*n* = 52)

Fiber source	Rho	*P*
Total	0.366	0.0076
Fruits	0.293	0.0346
Vegetables	0.275	0.0479
Potatoes	0.46	6×10^{-4}
Grains	−0.08	0.5729
Legumes and nuts	0.1	0.4847

Roseburia and *Eubacterium* (Fig. S5). Surprisingly, healthy subjects had a somewhat low relative abundance of some of these butyrate producers compared to subjects with disease phenotypes, which might be a result of the higher bacterial diversity in healthy individuals and consequently a lower "reliance" on a few species for butyrate production. *F. prausnitzii*, *Eubacterium hallii*, and *Subdoligranulum variabile* had comparable relative abundances in patients with normal pouches, UC patients, and healthy subjects, while patients with pouchitis and CD had low levels of these taxa. For a complete list of all the identified butyrate producers, see Data Set S1, sheet 5. Notably, a significant negative correlation was observed between the abundance of butyrate synthesis genes and fecal calprotectin (Fig. S6A to D). This was most pronounced for the *but* gene (Spearman $r = -0.322$; $P = 4.34 \times 10^{-10}$). This suggests that either a reduction in butyrate producers predisposes to intestinal inflammation or butyrate-producing bacteria are more sensitive to intestinal inflammation.

As butyrate production may be affected by patients' dietary habits, especially dietary fiber consumption, we correlated the average daily intake of dietary fiber derived from different food sources with butyrate synthesis genes in a subset of pouch patients (*n* = 52) that had dietary information (food frequency questionnaires [FFQs] obtained at the time of fecal sample collection) (Table 1). The abundances of *but* and *buk* genes were positively correlated with total dietary fiber intake (Spearman $r = 0.366$; $P = 0.0076$) and specifically with intake of fiber from fruit (Spearman $r = 0.293$; $P = 0.035$), vegetables (Spearman $r = 0.275$; $P = 0.048$), and potatoes (Spearman $r = 0.46$; $P = 6 \times 10^{-4}$). Higher fruit consumption was previously associated with increased abundance of known butyrate-producing species and lower recurrence of pouchitis (32). Nutritional intervention for patients with a pouch, such as the Mediterranean diet, was previously associated with decreased inflammatory markers and later onset of pouchitis and is characterized by a high intake of dietary fiber (33). Our results, therefore, highlight diet as an important factor that may affect butyrate synthesis and provide further support for dietary intervention as a plausible approach to modulate the microbiota for increased butyrate synthesis potential in the pouch.

Previous studies examining butyrate-producing bacteria, their gene content, or their metabolites in IBD reported various results. A study by Machiels et al. (34) reported no difference in fecal butyrate concentrations between healthy subjects and UC patients, while others reported decreased fecal butyrate in CD and (to a lesser degree) UC patients (35, 36). According to the quantification of *but* gene content, reduced butyrate synthesis was identified in patients with active and inactive CD but only in patients with active UC (37). For patients with a pouch, data about butyrate production are scarce, and butyrate synthesis genes had not been previously analyzed. It was suggested that bacterial dysbiosis in pouchitis causes SCFA deprivation (6), which might be related to its pathogenesis. Fecal levels of SCFA in patients with pouchitis were substantially lower than in patients with a normal pouch (38, 39), which may contribute to the higher pH observed in pouchitis (27). These past observations of butyrate and pH levels in the pouch are in line with our metagenomics-based findings and suggest that a shortage of butyrate (and its producers) is an important contributor to the pathogenesis of pouchitis. This may be clinically relevant, as SCFA enemas were
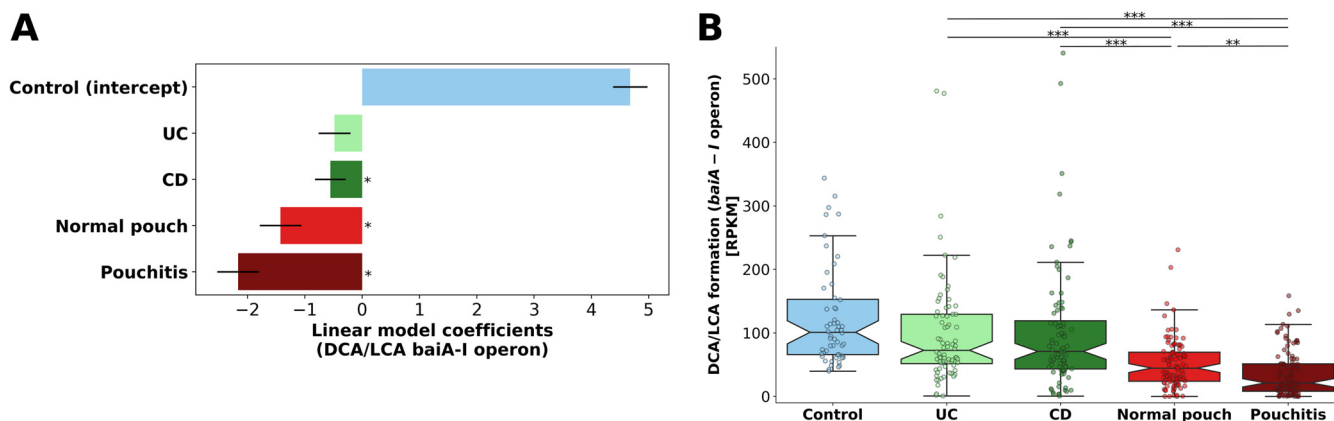
FIG 5 Secondary bile acid (DCA and LCA) formation potential in the metagenomes of patients with IBD and healthy subjects. The *bai* gene cluster (*baiA-bail* operon) was quantified in the metagenomes. (A) Linear model coefficients (effect sizes) are given for each phenotype for the modeled *bai* gene cluster. *, $P < 0.05$. (B) Gene abundance is in RPKM units (see the legend to Fig. 4). *, $P < 0.05$; **, $P < 0.01$; *, $P < 0.01$; Kruskal-Wallis, applying Dunn's multiple correction test. For a list of all pairwise P values, see Data Set S1, sheet 12. Box plot whiskers mark observations within the 1.5 interquartile range of the upper and lower quartiles.

successfully assessed for the treatment of diversion colitis (40) and for UC (41), yet their effect in pouchitis was rarely assessed in small uncontrolled trials and case studies (42). Our findings provide a mechanistic rationale for a reassessment of this therapy in pouchitis.

Secondary bile acid production potential is low in pouchitis. The secondary bile acids deoxycholic acid (DCA) and lithocholic acid (LCA) are formed in the colon from primary bile acid metabolism by the microbiota via a multistep 7α-dehydroxylation reaction. Secondary bile acids (DCA and LCA) have extensive effects on host metabolism and play both negative and positive roles in health and disease (43). The level of DCA in bile is thought to be controlled by two major factors: levels and activities of bile acid 7α-dehydroxylating gut microbes and colonic transit time (43). In IBD, impaired fecal bile acid metabolism was observed, with an increase in primary bile acids and a decrease in secondary bile acids (particularly during flares). *In vitro* and *in vivo* experimental models confirmed that DCA and LCA may exert anti-inflammatory effects (44, 45). Enzymes involved in 7α-dehydroxylation are encoded by bile acid-inducible (*bai*) gene clusters present in bacterial genomes from *Ruminococcaceae* and to a lesser extent from *Lachnospiraceae* and *Peptostreptococcaceae* (46). The *bai* gene cluster was found to be present and expressed in the gut microbiomes of most individuals but represented only a small fraction ($<1\%$) of total intestinal bacteria (47).

We quantified and modeled the *bai* gene cluster in the metagenomes of IBD and healthy subjects using an extensive and rigorously curated database for genes related to secondary bile acid formation (47). Patients with CD, normal pouch, and pouchitis were significantly associated with a decreased secondary bile acid formation potential (Fig. 5A) ($P < 0.05$). In line with the highest dysbiosis in patients with pouchitis, the *bai* gene cluster abundance was lowest in this group (median RPKM, 21.3) but higher in patients with a normal pouch (Fig. 5B) (median RPKM, 44.6; Kruskal-Wallis, $P = 0.002$). UC and CD patients had comparable levels of *bai* gene abundance (median RPKM, 72.6 and 71, respectively; Kruskal-Wallis, $P = 0.29$), and both were significantly higher than in patients with a pouch (Kruskal-Wallis, $P < 0.001$). A similar pattern in the *bai* gene cluster was observed in pouch patients from the American cohort (Fig. S3I).

Only a few dominant strains in the analyzed metagenomes were identified to carry the *bai* gene clusters (or parts thereof). They belonged to three families of *Firmicutes*: *Lachnospiraceae* (*Clostridium scindens*), *Peptostreptococcaceae* (*Clostridium sordellii* and *Proteocatella sphenisci*), and *Ruminococcaceae* (uncultured *Firmicutes* strain CAG:103 and *Clostridiales* strains UBA4701 and UBA1701). For a complete list of the identified secondary bile acid producers, see Data Set S1, sheet 6. Importantly, the abundance of

bai gene cluster was significantly negatively correlated with fecal calprotectin (Fig. S6E; Spearman $r = -0.312$; $P = 1.64 \times 10^{-9}$), implying that DCA- and LCA-producing bacteria are associated with lower levels of intestinal inflammation.

A reduction in secondary bile acid genes in patients with CD or UC (47, 48), as well as a reduction in DCA and LCA metabolites in highly dysbiotic patients with UC and CD, was previously described (13), supporting our current study. The key species carrying *bai* clusters in our metagenomes are concordant with those identified in metagenomes of a healthy population (46). Importantly, one of the prevalent secondary bile acids producers we detected, *C. scindens*, was associated with enhanced resistance to *C. difficile* infection in a secondary-bile-acid-dependent manner (49). The reduction of secondary-bile-acid-producing bacteria in IBD and particularly in pouchitis might leave these patients more exposed to pathogens or potential pathobionts. Interestingly, patients with an ileal pouch were previously shown to have markedly lower biliary and fecal DCA and LCA levels (compared to healthy controls), while fecal primary bile acid excretion was significantly higher, indicating attenuated bacterial conversion (50). These measurements complement our metagenomic findings and support our hypothesis that a reduction in 7α-dehydroxylation bacterial enzymes leads to a pouch microbiome with lower potential for producing secondary bile acids.

Enzymes related to protection against oxidative stress are highly prevalent in patients with a pouch. Previous studies suggested higher oxidative stress in the inflamed gut of patients with UC and CD (9, 25, 51). Here, we aimed to characterize and quantify genes that encode enzymes involved in oxidative stress response (oxygen-detoxifying enzymes) as a proxy for the levels of oxygen tolerance of microbes in the metagenomes of patients with IBD. We used the enzyme profiles annotated with HUMAnN2 to extract and model such oxygen-detoxifying genes. We identified seven enzymes known to be involved in oxidative stress (Data Set S1, sheet 7). However, only glutathione-disulfide reductase (EC 1.8.1.7), which serves to maintain the important antioxidant glutathione in the reduced form (52), was significantly higher in patients with a pouch (mean number of copies per million [CPM], 48.8 and 54.7 in normal pouch and pouchitis, respectively) than in those in UC and CD (mean CPM, 12.9 and 17.2, respectively).

The combined contribution of these seven oxidative stress response enzymes (combined oxidative stress response score) (Fig. S7) was the highest in patients with pouchitis (median RPKM, 401; Kruskal-Wallis, $P < 0.05$), followed by patients with a normal pouch (median RPKM, 298; Kruskal-Wallis, $P < 0.05$), and was significantly lower in the other phenotypes.

A clear difference was observed between the species that contributed highly to these antioxidative stress enzymes in pouch and those in nonpouch phenotypes (Fig. S8). In the former, the vast majority consisted of facultative anaerobes such as *E. coli* and to a lesser degree *Klebsiella* and few species of *Streptococcus* and *Lactobacillus*. In the latter, *E. coli* contribution was less pronounced and replaced by that of obligate anaerobes such as *Bacteroides* species, several *Clostridiales* species, *A. muciniphila*, *Bifidobacterium longum*, and *Prevotella copri*. A similar trend of increased facultative anaerobe abundance was previously observed in patients with CD or UC compared to healthy individuals (25). Our results suggest that microbes that are more oxygen tolerant and encode more oxygen-detoxifying enzymes have a fitness advantage and become more abundant in the highly oxidative environment of the pouch.

Metabolic pathway perturbation in patients with a pouch or CD. Our next goal was to use an untargeted approach to broadly analyze the functional potential of the microbiome and to identify metabolic pathways (MetaCyc annotations) that might be associated with IBD and with pouch phenotypes in particular. Such hypothesis-free analysis can reveal new functional pathways that may relate to the course of the disease and dysbiosis. We used HUMAnN2 to annotate the metagenomes for metabolic pathways and compared them across disease phenotypes using linear modeling (Data Set S1, sheet 8).

Metabolic pathways for the biosynthesis of the amino acids L-lysine (PWY-5097 and

PWY-2942) and L-ornithine (ARGININE-SYN4-PWY) were less abundant (false discovery rate [FDR] $P < 0.1$) in patients with a pouch compared to CD and UC patients. However, the abundances of other amino acid biosynthesis pathways were significantly enriched (FDR $P < 0.1$) in patients with a pouch and to some extent in those with CD: L-alanine (PWY0-1061), phenylalanine (PWY-6628), L-ornithine as an intermediate for L-arginine (GLUTORN-PWY), and L-arginine (ARGSYNBSUB-PWY). These pathways were encoded predominantly by members of the *Enterobacteriaceae* (*E. coli*, *Klebsiella* sp., and *Enterobacter cloacae*), *Enterococcus* (e.g., *Enterococcus faecium* and *Enterococcus faecalis*) and *Streptococcus* species, which were enriched in CD and in pouch phenotypes in particular (Fig. 2). In addition, we found that patients with CD, normal pouch, and pouchitis had a higher level (mean CPM, 10.3, 20.3, and 37.2, respectively) of molybdenum cofactor synthesis pathway (PWY-6823), encoded by *E. coli*, which is a crucial cofactor required for anaerobic respiration with nitrate as a terminal electron acceptor during intestinal inflammation (53). Interestingly, none of the above-mentioned amino acid pathways were significantly enriched in UC. It was previously demonstrated that a range of amino acids, including lysine, alanine, and phenylalanine were higher in the fecal metabolomes of patients with active CD and UC compared to healthy controls (54). Our findings are consistent with metagenomic and metabolomic studies of pediatric CD fecal samples that revealed positive associations of bacterial genes for nitrogen metabolism and amino acid synthesis from Proteobacteria (55), which also correlated with higher concentrations of fecal amino acids and their derivatives in CD patients (56). It is still unknown to what degree the higher levels of fecal amino acids are due to inflammation-driven host malabsorption or higher synthesis by the intestinal microbiota.

In addition to biosynthesis pathways, we observed a significant enrichment in the pathway for degradation of the amino acid ornithine (ORNDEG-PWY) in patients with a normal pouch (mean CPM, 16.6) and pouchitis (mean CPM, 30.9) and to a lesser degree in CD patients (mean CPM, 7.3). Moreover, the pathway for degradation of the aromatic organic compound phenylpropanoid (PWY-6690), which can be derived from the putrefaction of proteins, was significantly higher in patients with a normal pouch (mean CPM, 14.5), pouchitis (mean CPM, 27.4), and CD (mean CPM, 10.4). We identified two more pathways for aromatic compound degradation, 3-phenylpropanoate (PWY0-1277) and phenylacetate (PWY0-321), that followed a similar pattern of higher abundance in pouch and to lower extent in CD. These phenolic compounds are important intermediates in bacterial degradation of aromatic molecules. Lastly, galactarate sugar degradation (GALACTARDEG-PWY) was increased in patients with a normal pouch (mean CPM, 20.2), pouchitis (mean CPM, 36.6), and CD (mean CPM, 10.4). The degradation pathways we identified were encoded by *Enterobacteriaceae* species. It was reported that host-mediated galactarate oxidation following antibiotic treatment led to an expansion of *Salmonella enterica* and potentially other *Enterobacteriaceae* that used this sugar (57). Our results indicate that species of the *Enterobacteriaceae* may benefit from higher protein, aromatic compound, sugar, and galactarate availability, which can explain the relative expansion of this bacterial family in patients with CD (10) and with pouchitis (11, 12).

Classification models can distinguish between patients with a normal pouch and pouchitis based on species and functions. We observed multiple differences in species and functional genes between patients with a normal pouch and those with pouchitis, yet no bacterial feature on its own was highly discriminative. We, therefore, attempted to use species, metabolic pathway, and enzyme profiles in order to build classifiers to distinguish between these pouch phenotypes in the discovery cohort. We used a machine learning algorithm of gradient boosting trees (GBT) to train classifiers and evaluated its performance using 5-fold cross-validation. We designed our models based on the averaging of longitudinal data per subject, a classifier approach that was shown to be superior to the use of a single sample per subject (58). Our pouch cohort (discovery) consisted of 208 samples from 69 patients ($n = 35$ for normal pouch, $n = 34$ for pouchitis). To improve the model's predictive power, we used the feature importance scores that the algorithm assigns to each feature used during training; therefore,
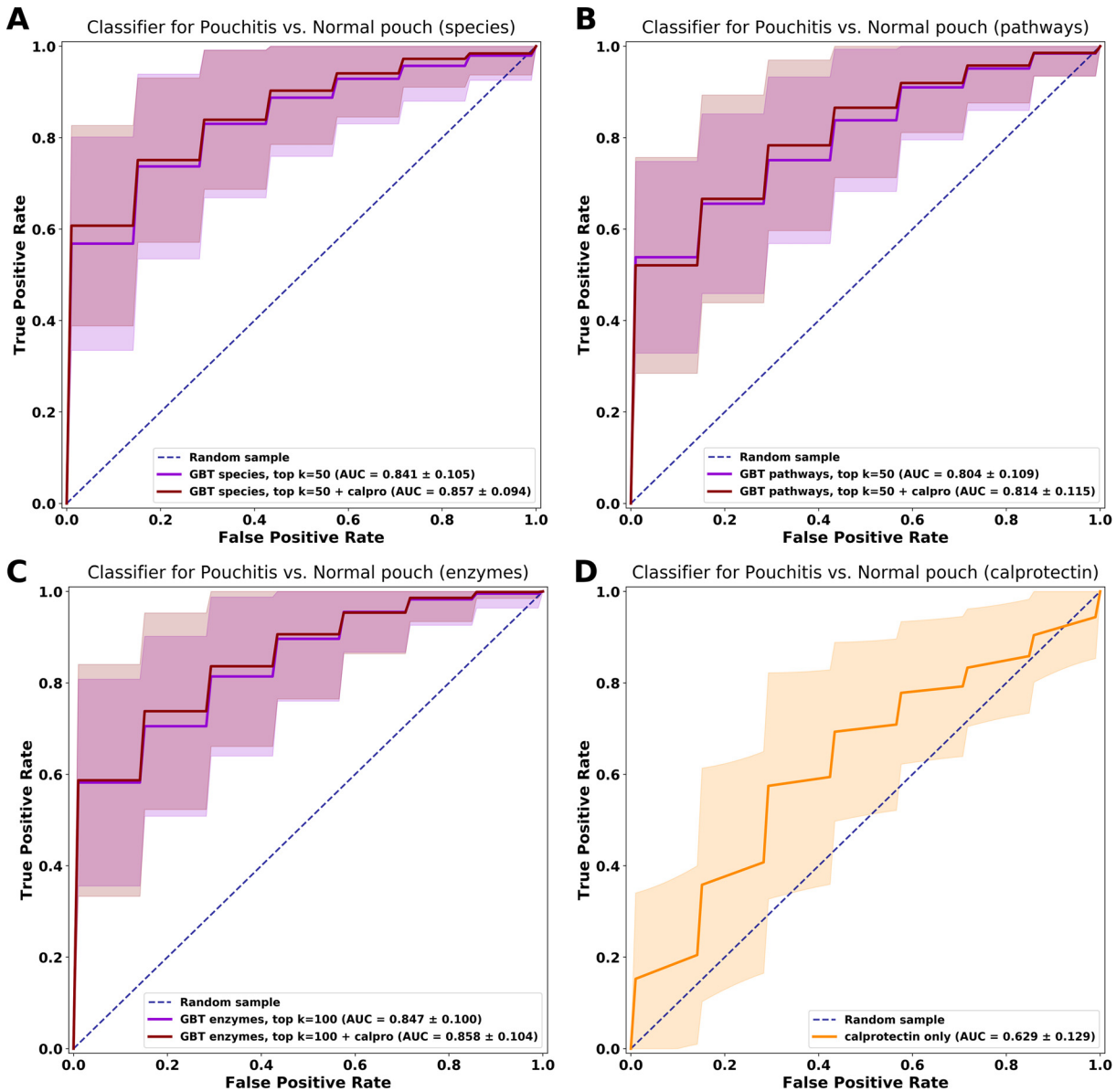
FIG 6 Classification models to distinguish between patients with a normal pouch and pouchitis. We trained and evaluated gradient boosting tree (GBT) classifiers on bacterial, metabolic pathway, and enzyme profiles from the metagenomes of patients with a pouch (discovery cohort) by using 5-fold cross-validation (randomly repeated 100 times). The area under the curve (AUC) (mean ± standard deviation) for each classifier's performance is presented. (A) Species-based classifier. (B) Metabolic pathway-based classifier. (C) Enzyme-based classifier. (D) Fecal calprotectin-based classifier. For each different classifier (species, pathway, and enzyme based, excepting calprotectin), two different models were built, with top-scoring features based on importance scores with and without fecal calprotectin as an additional predictor.

features with the highest scores are the most informative for correct classification. A GBT classifier trained on the 50 top-scoring features (species) achieved a mean area under the curve (AUC) of 0.841 ± 0.105 (Fig. 6A). Slightly improved outcomes were noticed when fecal calprotectin was incorporated into the model as a predictor (AUC = 0.857 ± 0.094). Next, we built GBT models with metabolic pathways (MetaCyc) and microbial enzymes (ECs) as features. The pathway-based GBT model with the 50 top-scoring features achieved an AUC of 0.804 ± 0.109, a worse performance than the species-based classifier, while the addition of calprotectin only slightly improved the model (AUC = 0.814 ± 0.115). Interestingly, the GBT models trained on the top-100-ranking bacterial enzymes achieved classification scores (without and with calprotectin,

AUC = 0.847 ± 0.1 and AUC = 0.858 ± 0.104, respectively) comparable to those of the species-based models, despite having a larger number of features (Fig. 6C). Furthermore, as the sole predictive feature, fecal calprotectin achieved a poor AUC of 0.625 ± 0.129 (Fig. 6D) in distinguishing normal pouch from pouchitis, emphasizing that intestinal inflammation alone cannot explain the observed differences between pouchitis and normal pouch phenotypes.

Next, we validated our model on 20 pouch patients from our center (not part of the discovery cohort) and on an American cohort of 15 patients with a pouch (17). The species- and enzyme-based models trained on the discovery cohort and tested on the validation cohort performed well, achieving AUC of 0.812 and 0.796, respectively (Fig. S9A and B). In terms of accuracy, the species model correctly classified 25/35 patients (71.4%), while the enzymes model correctly classified 27/35 patients (77.1%), suggesting that our classifier can generalize fairly well on unseen data (Fig. S9C and D).

We analyzed the top-ranking features from each model in order to identify which species and functions are the most discriminatory between the pouch phenotypes and thus might serve as potential predictors of pouchitis. Among the 50 species with the highest importance score, 15 were significantly different (Mann-Whitney, FDR $P < 0.1$) between normal pouch and pouchitis (Fig. 7A; Data Set S1, sheet 9). Commensal *Ruminococcaceae* (*Ruminococcus obeum* and *Ruminococcus flavefaciens*) and *Lachnospiraceae* (*Dorea formicigenerans* and *Dorea longicatena*) taxa and beneficial species such as *F. prausnitzii*, *E. rectale*, and *Roseburia inulinivorans* were more abundant in the normal pouch. Taxa that were enriched in the pouchitis group and might be inflammation associated were *E. coli*, an unclassified *Providencia* species, and *Yersinia* sp. (all members of the *Enterobacteriaceae*) and an unclassified *Acinetobacter* sp. (*Moraxellaceae*). Interestingly, *Providencia* species have been implicated in diarrhea, urinary tract infections, and sepsis (59), and members of the genus *Yersinia*, which includes enteric human pathogens, were found in ~7.7% of ileal biopsy specimens of both CD patients and controls (60). In a cohort of pediatric CD patients from a non-Western population, levels of *Acinetobacter* organisms were found to be significantly higher than in controls (61). We noted that three of the above-mentioned mucolytic bacteria appeared in the top-ranking species: *R. torques* was higher in normal pouch (Mann-Whitney, FDR $P < 0.1$), while *R. gnavus* and *A. muciniphila* were higher in pouchitis (not statistically significant).

Metabolic pathways with a high-ranking score in differentiating patients with a pouch (Fig. 7B; Data Set S1, sheet 10) included some general core metabolic pathways present in many organisms and thus merely reflected differences in species diversity (tricarboxylic acid [TCA] cycle, pentose phosphate pathway, and nucleoside and nucleotide biosynthesis). Many of the pathways that were enriched in pouchitis were encoded primarily by *E. coli* and several other *Enterobacteriaceae* that were more abundant in these patients. Potentially important degradation pathways of phytate, phenylacetate, ornithine, methylphosphonate, and sulfoquinovose were higher in pouchitis (Mann-Whitney, FDR $P < 0.05$). In addition, the pathway for biosynthesis of the lipid A (NAGLIPASYN-PWY) component of lipopolysaccharide (LPS) was significantly enriched in pouchitis. Lipopolysaccharide is a potent activator of innate immunity via Toll-like receptor 4, and patients with CD and UC, especially active ones, have been found to have increased serum levels of LPS (62). On the other hand, the biosynthesis of L-isoleucine and L-methionine was higher in patients with a normal pouch (Mann-Whitney, FDR $P < 0.05$). We also explored the enzymes that were highly ranked by the model (Fig. 7C; Data Set S1, sheet 11) and were differentially abundant between the phenotypes (Mann-Whitney, FDR $P < 0.05$). Levels of genes for two putative aminopeptidases (EC 3.4.11.1 and EC 3.4.11.10) as well as a gene for rhamnose sugar degradation (L-rhamnose mutarotase) were increased in pouchitis. Additional enzymes that were higher in pouchitis included adenosylmethionine decarboxylase, which is involved in the biosynthesis of spermidine, a polyamine that may play a role in survival and virulence of human bacterial pathogens (63), and D-sedoheptulose 7-phosphate isomerase, which is involved in the biosynthesis of D-mannoheptose, an LPS component. Notably,
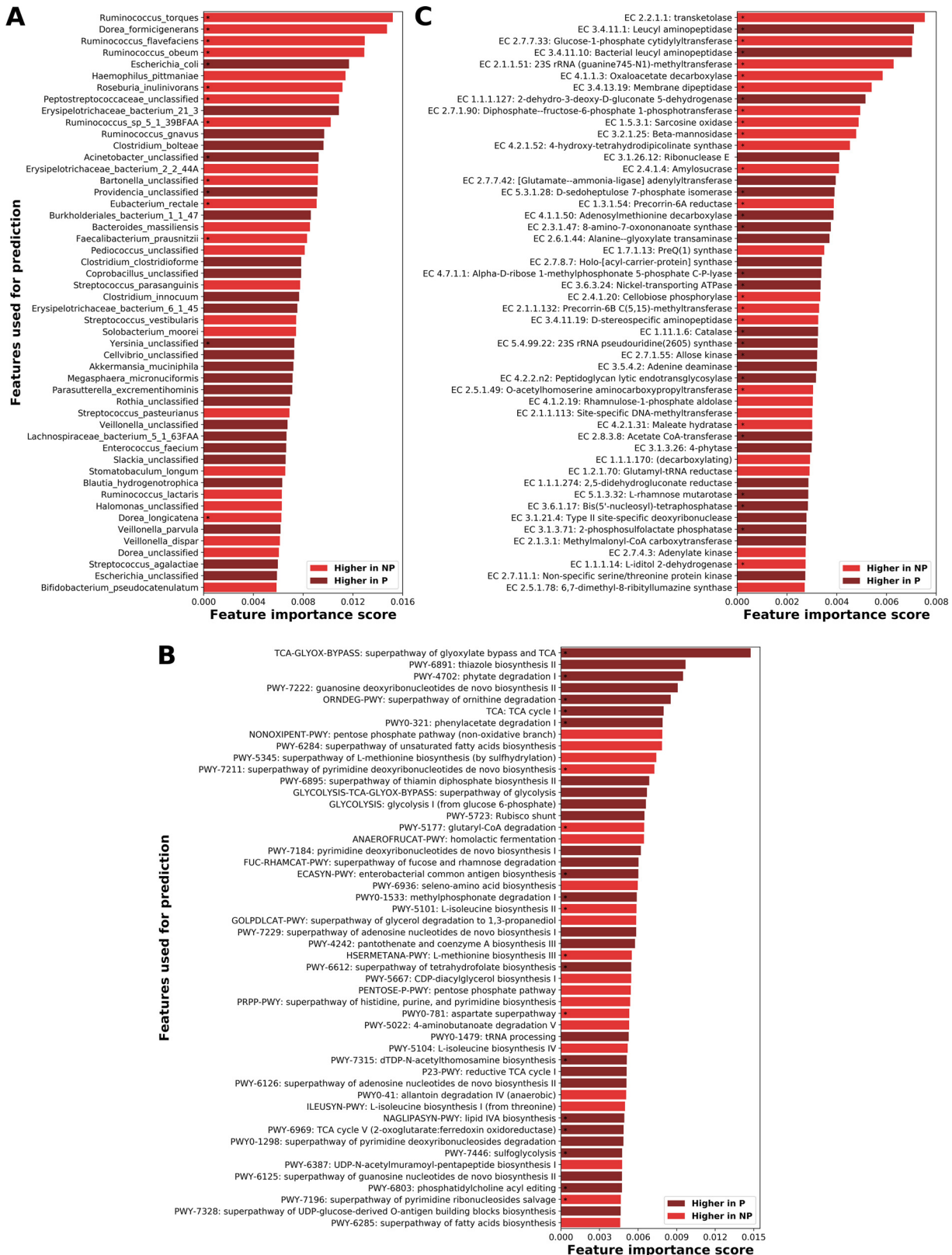
FIG 7 Highest-scoring microbiome features identified by the classification models as the most informative for distinguishing between patients with a normal pouch and pouchitis and thus possibly predictive of pouchitis. The feature importance scores (averaged across 5-fold cross-validation) for (A) species, (B) metabolic pathways, and (C) enzymes. The lighter red and darker red denote features higher in normal pouch (NP) and pouchitis (P) groups, respectively. For a list of all the importance scores for the features of each model, see Data Set S1, sheets 9 to 11.

catalase (EC 1.11.1.6), which detoxifies hydrogen peroxide, was significantly higher in pouchitis, indicating the presence of bacteria able to sustain higher levels of oxidative stress (facultative aerobes). In the normal pouch, genes related to vitamin B_{12} biosynthesis (EC 1.3.1.54, EC 2.1.1.132, and EC 2.1.1.133), starch, sucrose, and galactose metabolism (EC 2.4.1.4 and EC 3.2.1.10), sorbitol sugar degradation (EC 1.1.1.14), and homocysteine biosynthesis (EC 2.5.1.49) were enriched.

In summary, some of the enzymatic functions enriched in pouchitis result from the enrichment in Gram-negative bacteria (LPS) and in facultative aerobes that may benefit from intestinal inflammation. In contrast, normal pouch often has a broader diversity in metabolic capacities that accompanies the higher species diversity (Fig. 1D).

Conclusions. In this study, we explored species and functions in the microbiomes of a well-phenotyped cohort of patients with an ileal-anal pouch. We identified several metabolic pathways and enzymes that were altered in patients with a pouch and in CD. Most notably, patients with pouchitis had lower levels of genes for the synthesis of butyrate and secondary bile acids and harbored more facultative anaerobic bacteria, suggesting high oxidative stress during pouch inflammation. No excessive bacterial mucus utilization or degradation in IBD was found, but rather, a change in the mucolytic bacterial community composition, highlighted by an increased fraction of *R. gnavus* strains encoding fucosidases in patients with pouchitis, was seen. Patients with a pouch that later develop pouchitis are often considered a model for *de novo* inflammation in CD (14). Here, we show that the normal pouch is already dysbiotic (function- and species-wise), and based on multiple pathways, such as higher amino acid biosynthesis and degradation of aromatic compounds and sugars, encoded primarily by members of *Enterobacteriaceae*, pouch phenotypes are more similar to CD than to UC. Pouchitis thus represents a more extreme case of the various trends that constitute the dysbiosis observed in CD. Our machine learning models identified several species and genes that were predictive of pouchitis. Our analysis was limited by a relatively small number of patients and reliance on publicly available metagenomes from other cohorts for the nonpouch phenotypes, which introduces some country-specific biases. Nevertheless, we validated our analysis and models using a validation cohort and achieved a reasonably good classification. Our findings also delineate microbial and functional factors that underlie the pathogenesis of pouchitis and may suggest pathways that could be targeted for intervention in combination with dietary changes, such as higher intake of fiber, to attenuate small intestinal inflammation in pouchitis and CD. Future studies should generate higher-resolution clinical and nutritional data combined with microbial-based biomarkers which might be integrated into diagnostic tools to assist clinical decision-making.

MATERIALS AND METHODS

Study cohort description and metadata characteristics. Patients after pouch surgery were recruited at a dedicated pouch clinic in a tertiary IBD referral center in Israel (Rabin Medical Center [RMC]). The study was approved by the local institutional review board (0298–17) and the National Institutes of Health (NCT01266538). Demographic and clinical data were obtained during clinic visits. After providing written informed consent, patients were followed prospectively, and clinic visits, including fecal sample collection, were scheduled every 3 months or when patients had an episode of pouchitis. Fecal samples were collected in sterile cups and immediately frozen at −80°C until processing. Fecal calprotectin levels were measured. Pouch disease behavior (phenotype) was defined as normal pouch, acute/recurrent-acute pouchitis, chronic pouchitis, or Crohn's-like disease of the pouch (CLDP); see "Pouch disease behavior (pouch phenotype)" below for a detailed description. Disease activity was defined based on global physician assessment and on the clinical subscore of the pouchitis disease activity index (64).

For the exploratory analysis done in this study, the patients were divided into two groups (discovery cohort): normal pouch ($n = 35$; 103 fecal samples) and pouchitis ($n = 34$; 105 fecal samples) (Data Set S1, sheet 1). The normal pouch group included samples from patients with a normal pouch (with a UC or familial adenomatous polyposis [FAP] background), while the pouchitis group included samples from patients with chronic pouchitis and CLDP. For the validation cohort used for testing the machine learning models (see "Machine learning classifiers" below), we included 20 pouch patients from our center (not part of the discovery cohort) and 15 pouch patients from the American cohort (17) (Data Set S1, sheet 2).

We excluded samples collected during or within 1 month of antibiotic treatment. Of the pouch

samples in our data set, 84% were from patients who had been off antibiotics for 180 days or longer and the rest were from patients who had been off antibiotics for 1 to 6 months (Data Set S1, sheet 1). While patients with pouchitis tend to have a history of high antibiotic use, after 1 month off antibiotics, the microbiome diversity is significantly increased, and following 6 months postantibiotics, it recovers to pretreatment levels (12). Nevertheless, to account for any remaining confounding antibiotic effects, we also included antibiotic use (1 month or 6 months without antibiotics) as a covariate in the linear models.

Nonpouch IBD (patients with UC and CD) and healthy-adult metagenomes were obtained from previously published, publicly available data (18) and are part of the PRISM, LifeLines DEEP, and NLIBD cohorts. We analyzed these metagenomes with the same bioinformatic pipeline as the pouch metagenomes as described below, starting the analysis from raw reads.

Pouch disease behavior (pouch phenotype). Pouch disease behavior was defined as normal, acute/recurrent-acute pouchitis, chronic pouchitis, or CLDP, as previously defined (65). Briefly, a normal pouch was defined as one with no symptoms of pouchitis during the past 2 years and no antibiotic therapy. Acute/recurrent-acute pouchitis was defined as a flare of pouchitis responding to a short course (usually 2 weeks) of antibiotic therapy, or up to 4 flares/year, respectively. Chronic pouchitis was defined as >4 pouchitis flares/year or administration of antibiotics or IBD-specific anti-inflammatory therapy for more than 1 month. CLDP was defined as pouch-perianal disease, pouch strictures, or long segments of proximal small intestinal inflammation. Patients undergoing pouch surgery due to familial adenomatous polyposis (FAP) were recruited as well and had a normal pouch throughout follow-up.

Dietary information based on food frequency questionnaires. We obtained dietary-fiber information for a subset of 52 samples of patients with a pouch ($n = 36$). Intake of food groups and nutrients was assessed using food frequency questionnaires (FFQs) adapted to the Israeli population and administered in the MABAT Israeli nutrition and public health governmental study. For each of the 106 food items, a commonly used portion size was specified. Patients were asked how often on average they consumed each food item during the 6-month period prior to each visit, and daily intake of nutrients was calculated by multiplying the frequency of each food item by the nutrient content based on the Israeli TZAMERET food database and the United States Department of Agriculture food database (66).

DNA extraction and shotgun metagenomic sequencing of pouch samples. DNA was extracted from fecal samples as described in reference 11. Genomic libraries were prepared with the Nextera XT library preparation kit using approximately 1 ng of total DNA per sample (DNA concentrations verified by Qubit fluorometry). Metagenomic sequencing was done using Illumina (San Diego, CA) NextSeq500 paired-end (2 × 150-bp reads) at DNA Services Facility, University of Illinois, Chicago, IL. Sequencing reads were quality filtered with Trimmomatic v0.36 (67) using default parameters, thus removing low-quality and short reads and low-quality bases and clipping Illumina adaptors. Human host DNA reads were removed by mapping metagenomic reads against the GRCh38 human genome (build 38) with Bowtie2 version 2.2.9 (68) in –very-sensitive-local mode. After read trimming and human read removal, the mean sequence depth ($\pm$ standard deviation) was 0.93 ± 0.3 Gbp (6.22 ± 2 million reads) per sample.

Taxonomic and functional profiling of the metagenomic data sets. Taxonomic profiling was performed using MetaPhlAn2 classifier v2.6.0 (69), which classifies metagenomic reads by mapping to a database of clade-specific marker genes. MetaPhlAn2 was run with the following parameters changed from default: –tax_lev set to "s" (classify taxonomy to species level) and –ignore_virus and –ignore_eukaryotes to ignore viral and eukaryote reads, respectively. Species relative abundance tables from all samples were merged. Species present in less than 5% of the samples and at less than 0.1% relative abundance in any individual sample were removed. For all downstream statistical analysis, we used the bacterial species that are common to the pouch and nonpouch metagenomes (species intersection). This procedure can minimize potential confounders, such as species that appear solely in one data set due to country-specific differences. The only exception was for the Shannon diversity index calculation, where all taxa (union) from both data sets were used in order to obtain a more complete and accurate diversity trend.

Functional profiling was performed with HUMAnN2 v0.11.1 (70). Briefly, HUMAnN2 first maps metagenomic reads (nucleotide level with Bowtie2) against functionally annotated pangenomes of the species identified during taxonomic profiling. Reads that do not map to any identified pangenome are subjected to a translated search with DIAMOND (71) against the UniRef90 protein database, and hits are weighted according to alignment quality, sequence length, and coverage. Gene-level outputs are produced in reads per kilobase units and stratified according to known/unclassified microbiome contributions. Per-sample gene abundances were total-sum scaled and normalized (to account for the number of reads in each metagenome) to copies per million (CPM). The obtained gene abundances were regrouped to MetaCyc metabolic pathways and enzymes (using the Enzyme Commission number [EC]).

Screening the metagenomes for key bacterial functions with custom curated databases. For the identification and quantification of butyrate synthesis pathways, an extensive database from reference 31 was used, which encompassed hidden Markov models (HMM) on full-length proteins, considering the entire taxonomic diversity of butyrate producers, which contained 1,716 genomes. A modified version of this database was used so that only four terminal enzymes in the main pathways known for butyrate production from each of the genomes were considered. In both the acetyl-CoA and glutarate pathways, the terminal enzymes are butyryl-CoA:acetate CoA transferase (*but*) and butyrate kinase (*buk*). In the 4-

aminobutyrate pathway, the terminal enzyme is butyryl-CoA:4-hydroxybutyrate CoA transferase (*4hbt*), and in the lysine pathway, the terminal enzyme is butyryl-CoA:acetoacetate CoA transferase (*ato*). For the analysis of microbiome-derived secondary bile acids, we used the database from reference 47, which was built on the full-length HMM for the *bai* gene cluster (*baiA-baiI* operon), which consists of *baiA*, *baiB*, *baiCD*, *baiE*, *baiF*, *baiG*, *baiH*, and *baiI* genes, obtained from all genomes available in PATRIC.

Quality-filtered metagenomic reads were used as a query in a translated nucleotide search (BLASTX using DIAMOND v0.9.24.125 [71]) against the databases described above. Only the top hit per read was reported (–max-target-seqs 1), and the minimum E value threshold was set to 1×10^{-5} (–evalue). For butyrate synthesis genes, only reads with sequence similarity matches of ≥90% to a reference and an amino acid alignment length of ≥25 were retained and counted. For bacterial secondary bile acid production genes, the sequence similarity cutoff was lowered to ≥70% due to a less comprehensive reference database. Read counts for each of the key functions analyzed (butyrate and secondary bile acids genes) were normalized to RPKM units (reads per kilobase per million mapped reads), i.e., normalized according to the corresponding median gene length and total-sum scaled according to the number of reads in each metagenome. The key functional genes were assigned to bacterial taxa according to the taxonomic identity of the respective gene entries in the database.

Mucin degradation glycoside hydrolase analysis. For the analysis of glycoside hydrolases (GH) that are involved in mucin degradation, only GH that were functionally characterized by *in vitro* activity or transcriptomic assays (72) were considered. Thus, the following enzymes from HUMAnN2 EC4 annotations were considered mucin-related GH: protein *O*-GlcNAcase (EC 3.2.1.169), exo-alpha-sialidase (EC 3.2.1.18), β-galactosidase (EC 3.2.1.23), α-*N*-acetylgalactosaminidase (EC 3.2.1.49), α-*N*-acetylglucosaminidase (EC 3.2.1.50), α-L-fucosidase (EC 3.2.1.51), and β-*N*-acetylhexosaminidase (EC 3.2.1.52).

To analyze the specific GH of *Ruminococcus gnavus*, GH29 (α-L-fucosidase), GH95 (glycoside hydrolase family 95), and GH33 (exo-alpha-sialidase) protein sequences of *Ruminococcus gnavus* ATCC 29149 strain JCM6515 were downloaded from the Carbohydrate-Active enZYmes (CAZY) database (73). Three single-copy housekeeping genes (DNA gyrase subunit A [*gyrA*], recombinase A [*recA*], and 50S ribosomal protein L2 [*rplB*]) of the same *R. gnavus* strain were downloaded as well. These GH and the housekeeping genes were used to build a reference database for a similarity search, as mentioned above. For GH, sequence similarity matches of ≥90% and ≥25 amino acids, alignment length was used, and for housekeeping genes, sequence similarity matches were increased to ≥97% to count only closely related strains. Read counts were normalized by the corresponding gene length, and the percentage of *R. gnavus* strains encoding GH29, GH95, and GH33 were calculated by dividing each GH by the mean abundance of *gyrA*, *recA*, and *rplB*. As *gyrA* recruited the highest number of reads even after gene length normalization, we used only *gyrA* for the final calculation. Samples with fewer than 5 *gyrA* reads of *R. gnavus* were removed from the analysis.

Statistical analysis. The Kruskal-Wallis H test with Dunn's test for multiple-comparison correction (pairwise tests between phenotypes) was performed using kruskal.test and dunn.test functions in the R package dunn.test. Alpha-diversity (within samples) was calculated with the Shannon diversity index using the diversity function in the R package vegan. Beta-diversity (between samples) was calculated with Bray-Curtis dissimilarity (the vegdist function in the R package vegan) and visualized with the principal-coordinate analysis function cmdscale (classical multidimensional scaling) in R. Spearman rank correlation between the first axis of microbiome variation, Shannon diversity, and key functional genes and fecal calprotectin was performed using the cor.test function in R.

To calculate the proportion of variance explained by each tested variable on Bray-Curtis dissimilarity (for species, pathways, and enzymes), permutational multivariate analysis of variance (PERMANOVA) was performed using the adonis function in the R package vegan. The variables included disease phenotype, age, antibiotic use, fecal calprotectin, gender, disease activity, pouch phenotype, and intersubject variation (the last four variables were available only for the cohort of patients with a pouch and were fitted separately). The variance explained by each variable was calculated independently of other variables to avoid issues related to variable ordering.

To test for differently abundant microbiome features (species, enzyme, and pathway profiles) in different IBD phenotypes, generalized linear mixed-effects models from the R package MaAsLin2 v0.99.12 (http://huttenhower.sph.harvard.edu/maaslin) were used. Bacterial taxonomy and gene relative abundance data were log-transformed, and additive smoothing of minimum nonzero values was applied to zero values on a per-sample basis. The transformed abundances were modeled as a function of IBD phenotype (with healthy subjects as the reference point) while adjusting for age and antibiotic use (fixed effects). For antibiotic use, a time window of 1 month (recent use) or 6 months off antibiotics was used. To account for the longitudinal data set from pouch patients (repeated measurements), individual patients from whom a set of samples were derived were specified as a random effect (patient ID). To account for potential confounders due to the combination of several cohorts of patients, the cohort from which each sample was derived was specified as another random effect (cohort ID). All reported *P* values were adjusted for multiple-hypothesis testing using the Benjamini-Hochberg method. For differential abundance analysis of key bacterial genes (*but*, *buk*, *atoAD*, *4hbt*, and the *bai* gene cluster) in IBD phenotypes, a linear mixed-effects model was used with the same fixed and random effects as described above, with the lmer function from the R package lme4.

Machine learning classifiers. Profiles of bacterial species, metabolic pathways, and enzymes composition were used to build classification models to distinguish between pouch phenotypes (normal pouch versus pouchitis). The machine-learning algorithm of gradient boosting trees (GBT) was used, as implemented in the Python package XGBoost (eXtreme Gradient Boosting [74]) with the XGBClassifier

function. The best hyperparameters for each model were tuned empirically using a grid search matrix (GridSearchCV function in the Python scikit-learn package). Briefly, the learning rate was 0.1, the number of boosted trees was 250 (n_estimators), each tree was constructed randomly with 50% of the features for species and pathway models and with 10% for the enzyme model (colsample_bytree), trees were randomly trained on 50% of the samples (subsample), and maximum tree depth was 3 (max_depth). The following regularization parameters were used to reduce model complexity and overfitting: gamma = 0.1 (minimum loss reduction required to make a further partition on a leaf node), reg_alpha = 0.01 (L1 lasso), and reg_lambda = 0.5 (L2 ridge). The rest of the hyperparameters were left at the default settings. The GBT classifiers were based on the averaging of longitudinal microbiome data per subject (discovery cohorts of $n = 35$ for normal pouch and $n = 34$ for pouchitis classes). The classifiers were trained and evaluated on the discovery cohort by using 5-fold cross-validation (randomly repeated 100 times) with StratifiedKFold function (scikit-learn package), which preserves the same distribution of classes in each fold as in the complete training set. The classifiers were tested on the validation cohort, which consisted of 20 pouch patients from our center (not part of the discovery cohort), and on an American cohort of 15 patients with a pouch (17). The model performance was assessed by calculation area under the curve (AUC), sensitivity, specificity, and accuracy metrics.

To improve the model's predictive power (considering only the most informative features) and reduce complexity (reducing the feature space by removing uninformative and correlated features), feature importance scores were computed using the internal XGBoost function (feature_importances_), which assigns a score to each feature used in the decision tree during model training. The average feature importance scores for each model (species, pathways, and enzymes) were calculated by considering the mean score of each feature for each fold across the repeated ($\times 100$) 5-fold cross-validation.

Data sources. This study used our previously published sequence data from PRJNA524170. Metagenomes from patients with UC, patients with CD, and healthy controls were obtained from PRJNA400072. Additional pouch metagenomes from the American cohort were obtained from PRJNA600269.

Data availability. All the metagenomic sequence data generated and used in this study (patients with a pouch) have been deposited in NCBI SRA and are available under BioProject number PRJNA637365. Analysis scripts with code for the machine learning classifiers are available in the GitHub repository (https://github.com/VadimDu/pouchitis_classifiers).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
DATA SET S1, XLSX file, 0.9 MB.
FIG S1, TIF file, 1.4 MB.
FIG S2, TIF file, 1.5 MB.
FIG S3, TIF file, 2.2 MB.
FIG S4, JPG file, 2.8 MB.
FIG S5, TIF file, 1.6 MB.
FIG S6, TIF file, 1.5 MB.
FIG S7, TIF file, 0.4 MB.
FIG S8, JPG file, 2.9 MB.
FIG S9, TIF file, 0.8 MB.

analyzed and provided the clinical data; N.W. and I.D. enrolled and examined the patients; L.G. provided and analyzed the dietary information. V.D., L.R., U.G., and I.D. wrote the paper. All authors read, discussed, and approved the final manuscript.

REFERENCES

1. Lees CW, Barrett JC, Parkes M, Satsangi J. 2011. New IBD genetics: common pathways with other diseases. Gut 60:1739–1753. https://doi.org/10.1136/gut.2009.199679.

2. Ananthakrishnan AN, Bernstein CN, Iliopoulos D, Macpherson A, Neurath MF, Ali RAR, Vavricka SR, Fiocchi C. 2018. Environmental triggers in IBD: a review of progress and evidence. Nat Rev Gastroenterol Hepatol 15:39–49. https://doi.org/10.1038/nrgastro.2017.136.

3. Sartor RB, Wu GD. 2017. Roles for intestinal bacteria, viruses, and fungi in pathogenesis of inflammatory bowel diseases and therapeutic approaches. Gastroenterology 152:327–339.E4. https://doi.org/10.1053/j.gastro.2016.10.012.

4. Shen B, Fazio VW, Remzi FH, Lashner BA. 2005. Clinical approach to diseases of ileal pouch-anal anastomosis. Am J Gastroenterol 100:2796–2807. https://doi.org/10.1111/j.1572-0241.2005.00278.x.

5. Yanai H, Ben-Shachar S, Mlynarsky L, Godny L, Leshno M, Tulchinsky H, Dotan I. 2017. The outcome of ulcerative colitis patients undergoing pouch surgery is determined by pre-surgical factors. Aliment Pharmacol Ther 46:508–515. https://doi.org/10.1111/apt.14205.

6. Schieffer KM, Williams ED, Yochum GS, Koltun WA. 2016. The pathogenesis of pouchitis. Aliment Pharmacol Ther 44:817–835. https://doi.org/10.1111/apt.13780.

7. Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, Järnerot G, Tysk C, Jansson JK, Engstrand L. 2010. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology 139:1844–1854.E1. https://doi.org/10.1053/j.gastro.2010.08.049.

8. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran LG, Gratadoux J-J, Blugeon S, Bridonneau C, Furet J-P, Corthier G, Grangette C, Vasquez N, Pochart P, Trugnan G, Thomas G, Blottiere HM, Dore J, Marteau P, Seksik P, Langella P. 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A 105:16731–16736. https://doi.org/10.1073/pnas.0804812105.

9. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13:R79. https://doi.org/10.1186/gb-2012-13-9-r79.

10. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, González A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ. 2014. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 15:382–392. https://doi.org/10.1016/j.chom.2014.02.005.

11. Reshef L, Kovacs A, Ofer A, Yahav L, Maharshak N, Keren N, Konikoff FM, Tulchinsky H, Gophna U, Dotan I. 2015. Pouch inflammation is associated with a decrease in specific bacterial taxa. Gastroenterology 149:718–727. https://doi.org/10.1053/j.gastro.2015.05.041.

12. Dubinsky V, Reshef L, Bar N, Keizer D, Golan N, Rabinowitz K, Godny L, Yadgar K, Zonensain K, Tulchinsky H, Gophna U, Dotan I. 2020. Predominantly antibiotic-resistant intestinal microbiome persists in patients with pouchitis who respond to antibiotic therapy. Gastroenterology 158:610–624.e13. https://doi.org/10.1053/j.gastro.2019.10.001.

13. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vázquez-Baeza Y, White RA, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, Petrosino JF, Stappenbeck TS, Winter HS, Clish CB, Franzosa EA, Vlamakis H, Xavier RJ, Huttenhower C, IBDMDB Investigators. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569:655–662. https://doi.org/10.1038/s41586-019-1237-9.

14. Ben-Shachar S, Yanai H, Baram L, Elad H, Meirovithz E, Ofer A, Brazowski E, Tulchinsky H, Pasmanik-Chor M, Dotan I. 2013. Gene expression profiles of ileal inflammatory bowel disease correlate with disease phenotype and advance understanding of its immunopathogenesis. Inflamm Bowel Dis 19:2509–2521. https://doi.org/10.1097/01.MIB.0000437045.26036.00.

15. Ben-Shachar S, Yanai H, Sherman Horev H, Elad H, Baram L, Issakov O, Tulchinsky H, Pasmanik-Chor M, Shomron N, Dotan I. 2016. MicroRNAs expression in the ileal pouch of patients with ulcerative colitis is robustly up-regulated and correlates with disease phenotypes. PLoS One 11: e0159956. https://doi.org/10.1371/journal.pone.0159956.

16. Goren I, Yahav L, Tulchinsky H, Dotan I. 2015. Serology of patients with ulcerative colitis after pouch surgery is more comparable with that of patients with Crohn's disease. Inflamm Bowel Dis 21:2289–2295. https://doi.org/10.1097/MIB.0000000000000487.

17. Sinha SR, Haileselassie Y, Nguyen LP, Tropini C, Wang M, Becker LS, Sim D, Jarr K, Spear ET, Singh G, Namkoong H, Bittinger K, Fischbach MA, Sonnenburg JL, Habtezion A. 2020. Dysbiosis-induced secondary bile acid deficiency promotes intestinal inflammation. Cell Host Microbe 27:659–670.E5. https://doi.org/10.1016/j.chom.2020.01.021.

18. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, Vatanen T, Hall AB, Mallick H, McIver LJ, Sauk JS, Wilson RG, Stevens BW, Scott JM, Pierce K, Deik AA, Bullock K, Imhann F, Porter JA, Zhernakova A, Fu J, Weersma RK, Wijmenga C, Clish CB, Vlamakis H, Huttenhower C, Xavier RJ. 2018. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol 4:293–305. https://doi.org/10.1038/s41564-018-0306-4.

19. Szamosi JC, Forbes JD, Copeland JK, Knox NC, Shekarriz S, Rossi L, Graham M, Bonner C, Guttman DS, Van Domselaar G, Surette MG, Bernstein CN. 2020. Assessment of inter-laboratory variation in the characterization and analysis of the mucosal microbiota in Crohn's disease and ulcerative colitis. Front Microbiol 11:2028. https://doi.org/10.3389/fmicb.2020.02028.

20. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, Ananthakrishnan AN, Andrews E, Barron G, Lake K, Prasad M, Sauk J, Stevens B, Wilson RG, Braun J, Denson LA, Kugathasan S, McGovern DPB, Vlamakis H, Xavier RJ, Huttenhower C. 2018. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat Microbiol 3:337–346. https://doi.org/10.1038/s41564-017-0089-z.

21. Holmstrøm K, Collins MD, Møller T, Falsen E, Lawson PA. 2004. Subdoligranulum variabile gen. nov., sp. nov. from human feces. Anaerobe 10:197–203. https://doi.org/10.1016/j.anaerobe.2004.01.004.

22. Ze X, Duncan SH, Louis P, Flint HJ. 2012. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. ISME J 6:1535–1543. https://doi.org/10.1038/ismej.2012.4.

23. Png CW, Lindén SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin MA, Florin THJ. 2010. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. Am J Gastroenterol 105:2420–2428. https://doi.org/10.1038/ajg.2010.281.

24. Ganesh BP, Klopfleisch R, Loh G, Blaut M. 2013. Commensal Akkermansia muciniphila exacerbates gut inflammation in Salmonella Typhimurium-infected gnotobiotic mice. PLoS One 8:e74963. https://doi.org/10.1371/journal.pone.0074963.

25. Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, Lagoudas GK, Vatanen T, Fornelos N, Wilson R, Bertha M, Cohen M, Garber J, Khalili H, Gevers D, Ananthakrishnan AN, Kugathasan S, Lander ES, Blainey P, Vlamakis H, Xavier RJ, Huttenhower C. 2017. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome Med 9:103. https://doi.org/10.1186/s13073-017-0490-5.

26. Crost EH, Tailford LE, Le Gall G, Fons M, Henrissat B, Juge N. 2013. Utilisation of mucin glycans by the human gut symbiont Ruminococcus gnavus is strain-dependent. PLoS One 8:e76341. https://doi.org/10.1371/journal.pone.0076341.

27. Ruseler-van Embden JG, Schouten WR, van Lieshout LM. 1994. Pouchitis: result of microbial imbalance? Gut 35:658–664. https://doi.org/10.1136/gut.35.5.658.

28. Rivière A, Selak M, Lantin D, Leroy F, De Vuyst L. 2016. Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. Front Microbiol 7:979. https://doi.org/10.3389/fmicb.2016.00979.

29. Louis P, Flint HJ. 2017. Formation of propionate and butyrate by the human colonic microbiota: propionate and butyrate producing gut microbes. Environ Microbiol 19:29–41. https://doi.org/10.1111/1462-2920.13589.

30. Vital M, Howe AC, Tiedje JM. 2014. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. mBio 5:e00889-14. https://doi.org/10.1128/mBio.00889-14.

31. Vital M, Karch A, Pieper DH. 2017. Colonic butyrate-producing communities in humans: an overview using omics data. mSystems 2:e00130-17. https://doi.org/10.1128/mSystems.00130-17.

32. Godny L, Maharshak N, Reshef L, Goren I, Yahav L, Fliss-Isakov N, Gophna U, Tulchinsky H, Dotan I. 2019. Fruit consumption is associated with alterations in microbial composition and lower rates of pouchitis. J Crohns Colitis 13:1265–1272. https://doi.org/10.1093/ecco-jcc/jjz053.

33. Godny L, Reshef L, Pfeffer-Gik T, Goren I, Yanai H, Tulchinsky H, Gophna U, Dotan I. 2019. Adherence to the Mediterranean diet is associated with decreased fecal calprotectin in patients with ulcerative colitis after pouch surgery. Eur J Nutr 59:3183–3190. https://doi.org/10.1007/s00394-019-02158-3.

34. Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, Ballet V, Claes K, Van Immerseel F, Verbeke K, Ferrante M, Verhaegen J, Rutgeerts P, Vermeire S. 2014. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. Gut 63:1275–1283. https://doi.org/10.1136/gutjnl-2013-304833.

35. Marchesi JR, Holmes E, Khan F, Kochhar S, Scanlan P, Shanahan F, Wilson ID, Wang Y. 2007. Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. J Proteome Res 6:546–551. https://doi.org/10.1021/pr060470d.

36. Takaishi H, Matsuki T, Nakazawa A, Takada T, Kado S, Asahara T, Kamada N, Sakuraba A, Yajima T, Higuchi H, Inoue N, Ogata H, Iwao Y, Nomoto K, Tanaka R, Hibi T. 2008. Imbalance in intestinal microflora constitution could be involved in the pathogenesis of inflammatory bowel disease. Int J Med Microbiol 298:463–472. https://doi.org/10.1016/j.ijmm.2007.07.016.

37. Laserna-Mendieta EJ, Clooney AG, Carretero-Gomez JF, Moran C, Sheehan D, Nolan JA, Hill C, Gahan CGM, Joyce SA, Shanahan F, Claesson MJ. 2018. Determinants of reduced genetic capacity for butyrate synthesis by the gut microbiome in Crohn's disease and ulcerative colitis. J Crohns Colitis 12:204–216. https://doi.org/10.1093/ecco-jcc/jjx137.

38. Clausen MR, Tvede M, Mortensen PB. 1992. Short-chain fatty acids in pouch contents from patients with and without pouchitis after ileal pouch-anal anastomosis. Gastroenterology 103:1144–1153. https://doi.org/10.1016/0016-5085(92)91497-R.

39. Sagar PM, Taylor BA, Godwin P, Holdsworth PJ, Johnston D, Lewis W, Miller A, Quirke P, Williamson M. 1995. Acute pouchitis and deficiencies of fuel. Dis Colon Rectum 38:488–493. https://doi.org/10.1007/BF02148848.

40. Harig JM, Soergel KH, Komorowski RA, Wood CM. 1989. Treatment of diversion colitis with short-chain-fatty acid irrigation. N Engl J Med 320:23–28. https://doi.org/10.1056/NEJM198901053200105.

41. Scheppach W, Sommer H, Kirchner T, Paganelli G-M, Bartram P, Christl S, Richter F, Dusel G, Kasper H. 1992. Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. Gastroenterology 103:51–56. https://doi.org/10.1016/0016-5085(92)91094-K.

42. Wischmeyer P, Pemberton JH, Phillips SF. 1993. Chronic pouchitis after ileal pouch-anal anastomosis: responses to butyrate and glutamine suppositories in a pilot study. Mayo Clin Proc 68:978–981. https://doi.org/10.1016/S0025-6196(12)62270-8.

43. Ridlon JM, Harris SC, Bhowmik S, Kang D-J, Hylemon PB. 2016. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes 7:22–39. https://doi.org/10.1080/19490976.2015.1127483.

44. Duboc H, Rajca S, Rainteau D, Benarous D, Maubert M-A, Quervain E, Thomas G, Barbu V, Humbert L, Despras G, Bridonneau C, Dumetz F, Grill J-P, Masliah J, Beaugerie L, Cosnes J, Chazouillères O, Poupon R, Wolf C, Mallet J-M, Langella P, Trugnan G, Sokol H, Seksik P. 2013. Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. Gut 62:531–539. https://doi.org/10.1136/gutjnl-2012-302578.

45. Ward JBJ, Lajczak NK, Kelly OB, O'Dwyer AM, Giddam AK, Ní Gabhann J, Franco P, Tambuwala MM, Jefferies CA, Keely S, Roda A, Keely SJ. 2017. Ursodeoxycholic acid and lithocholic acid exert anti-inflammatory actions in the colon. Am J Physiol Gastrointest Liver Physiol 312:G550–G558. https://doi.org/10.1152/ajpgi.00256.2016.

46. Vital M, Rud T, Rath S, Pieper DH, Schlüter D. 2019. Diversity of bacteria exhibiting bile acid-inducible 7α-dehydroxylation genes in the human gut. Comput Struct Biotechnol J 17:1016–1019. https://doi.org/10.1016/j.csbj.2019.07.012.

47. Rath S, Rud T, Karch A, Pieper DH, Vital M. 2018. Pathogenic functions of host microbiota. Microbiome 6:174. https://doi.org/10.1186/s40168-018-0542-0.

48. Das P, Marcišauskas S, Ji B, Nielsen J. 2019. Metagenomic analysis of bile salt biotransformation in the human gut microbiome. BMC Genomics 20:517. https://doi.org/10.1186/s12864-019-5899-3.

49. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MRM, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature 517:205–208. https://doi.org/10.1038/nature13828.

50. Hakala K, Vuoristo M, Luukkonen P, Jarvinen HJ, Miettinen TA. 1997. Impaired absorption of cholesterol and bile acids in patients with an ileoanal anastomosis. Gut 41:771–777. https://doi.org/10.1136/gut.41.6.771.

51. Keshavarzian A, Banan A, Farhadi A, Komanduri S, Mutlu E, Zhang Y, Fields JZ. 2003. Increases in free radicals and cytoskeletal protein oxidation and nitration in the colon of patients with inflammatory bowel disease. Gut 52:720–728. https://doi.org/10.1136/gut.52.5.720.

52. Forman HJ, Zhang H, Rinna A. 2009. Glutathione: overview of its protective roles, measurement, and biosynthesis. Mol Aspects Med 30:1–12. https://doi.org/10.1016/j.mam.2008.08.006.

53. Winter SE, Lopez CA, Bäumler AJ. 2013. The dynamics of gut-associated microbial communities during inflammation. EMBO Rep 14:319–327. https://doi.org/10.1038/embor.2013.27.

54. Bjerrum JT, Wang Y, Hao F, Coskun M, Ludwig C, Günther U, Nielsen OH. 2015. Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals. Metabolomics 11:122–133. https://doi.org/10.1007/s11306-014-0677-3.

55. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, Bittinger K, Bailey A, Friedman ES, Hoffmann C, Albenberg L, Sinha R, Compher C, Gilroy E, Nessel L, Grant A, Chehoud C, Li H, Wu GD, Bushman FD. 2015. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. Cell Host Microbe 18:489–500. https://doi.org/10.1016/j.chom.2015.09.008.

56. Ni J, Shen T-CD, Chen EZ, Bittinger K, Bailey A, Roggiani M, Sirota-Madi A, Friedman ES, Chau L, Lin A, Nissim I, Scott J, Lauder A, Hoffmann C, Rivas G, Albenberg L, Baldassano RN, Braun J, Xavier RJ, Clish CB, Yudkoff M, Li H, Goulian M, Bushman FD, Lewis JD, Wu GD. 2017. A role for bacterial urease in gut dysbiosis and Crohn's disease. Sci Transl Med 9:eaah6888. https://doi.org/10.1126/scitranslmed.aah6888.

57. Faber F, Tran L, Byndloss MX, Lopez CA, Velazquez EM, Kerrinnes T, Nuccio S-P, Wangdi T, Fiehn O, Tsolis RM, Bäumler AJ. 2016. Host-mediated sugar oxidation promotes post-antibiotic pathogen expansion. Nature 534:697–699. https://doi.org/10.1038/nature18597.

58. Vázquez-Baeza Y, Gonzalez A, Xu ZZ, Washburne A, Herfarth HH, Sartor RB, Knight R. 2018. Guiding longitudinal sampling in IBD cohorts. Gut 67:1743–1745. https://doi.org/10.1136/gutjnl-2017-315352.

59. O'Hara CM, Brenner FW, Miller JM. 2000. Classification, identification, and clinical significance of Proteus, Providencia, and Morganella. Clin Microbiol Rev 13:534–546. https://doi.org/10.1128/cmr.13.4.534-546.2000.

60. Le Baut G, O'Brien C, Pavli P, Roy M, Seksik P, Tréton X, Nancey S, Barnich N, Bezault M, Auzolle C, Cazals-Hatem D, Viala J, Allez M, Hugot J-P, Dumay A, REMIND GROUP. 2018. Prevalence of Yersinia species in the ileum of Crohn's disease patients and controls. Front Cell Infect Microbiol 8:336. https://doi.org/10.3389/fcimb.2018.00336.

61. El Mouzan MI, Winter HS, Assiri AA, Korolev KS, Al Sarkhy AA, Dowd SE, Al Mofarreh MA, Menon R. 2018. Microbiota profile in new-onset pediatric Crohn's disease: data from a non-Western population. Gut Pathog 10:49. https://doi.org/10.1186/s13099-018-0276-3.

62. Rojo ÓP, Román ALS, Arbizu EA, de la Hera Martínez A, Sevillano ER, Martínez AA. 2007. Serum lipopolysaccharide-binding protein in endotoxemic patients with inflammatory bowel disease. Inflamm Bowel Dis 13:269–277. https://doi.org/10.1002/ibd.20019.

63. Shah P, Swiatlo E. 2008. A multifaceted role for polyamines in bacterial pathogens. Mol Microbiol 68:4–16. https://doi.org/10.1111/j.1365-2958.2008.06126.x.

64. Sandborn WJ, Tremaine WJ, Batts KP, Pemberton JH, Phillips SF. 1994. Pouchitis after ileal pouch-anal anastomosis: a pouchitis disease activity index. Mayo Clin Proc 69:409–415. https://doi.org/10.1016/S0025-6196(12)61634-6.

65. Tulchinsky H, Dotan I, Alper A, Brazowski E, Klausner JM, Halpern Z, Rabau M. 2008. Comprehensive pouch clinic concept for follow-up of patients after ileal pouch anal anastomosis: report of 3 years' experience in a tertiary referral center. Inflamm Bowel Dis 14:1125–1132. https://doi.org/10.1002/ibd.20430.

66. Montville JB, Ahuja JKC, Martin CL, Heendeniya KY, Omolewa-Tomobi G, Steinfeldt LC, Anand J, Adler ME, LaComb RP, Moshfegh A. 2013. USDA Food and Nutrient Database for Dietary Studies (FNDDS), 5.0. Procedia Food Sci 2:99–112. https://doi.org/10.1016/j.profoo.2013.04.016.

67. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

68. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

69. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 12:902–903. https://doi.org/10.1038/nmeth.3589.

70. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, Huttenhower C. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods 15:962–968. https://doi.org/10.1038/s41592-018-0176-y.

71. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

72. Tailford LE, Crost EH, Kavanaugh D, Juge N. 2015. Mucin glycan foraging in the human gut microbiome. Front Genet 6:81. https://doi.org/10.3389/fgene.2015.00081.

73. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.

74. Chen T, Guestrin C. 2016. XGBoost: a scalable tree boosting system, p 785–794. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '16 785–794. ACM Press, New York, NY. https://doi.org/10.1145/2939672.2939785.