The gut microbial profile in patients with primary sclerosing cholangitis is distinct from ulcerative colitis patients without biliary disease and healthy controls.

Martin Kummen[1,2,3,4], Kristian Holm[1,4], Jarl Andreas Anmarkrud[1,4], Ståle Nygård[5,6,7], Mette Vesterhus[1,8], Marte L. Høivik[9], Marius Trøseid[1,2,10], Hanns-Ulrich Marschall[11], Erik Schrumpf[1,3], Bjørn Moum[3,9], Helge Røsjø[7,12], Pål Aukrust[2,3,4,10], Tom H. Karlsen[1,2,3,4,13,14], Johannes R. Hov[1,2,3,4,14].

[1]Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; [2]K.G. Jebsen Inflammation Research Centre, University of Oslo, Oslo, Norway; [3]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [4]Research Institute of Internal Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; [5]Bioinformatics Core Facility, Institute for Medical Informatics, Oslo University Hospital, Oslo, Norway; [6]Institute for Experimental Medical Research, Oslo University Hospital and University of Oslo, Oslo, Norway; [7]K.G. Jebsen Cardiac Research Centre and Center for Heart Failure Research, University of Oslo, Oslo, Norway; [8]Department of Medicine, National Centre for Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway; [9]Department of Gastroenterology, Oslo University Hospital Ullevål, Oslo, Norway; [10]Section of Clinical Immunology and Infectious Diseases, Oslo University Hospital Rikshospitalet, Oslo, Norway; [11]Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [12]Division of Medicine, Akershus University Hospital, Lørenskog, Norway; [13]Department of Clinical Medicine, University of Bergen, Bergen, Norway; [14]Section of Gastroenterology, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway.

Corresponding author:
Dr. Johannes R. Hov
Norwegian PSC Research Center
Division for Cancer Medicine, Surgery and Transplantation
Oslo University Hospital Rikshospitalet
Pb 4950 Nydalen
N-0424 Oslo
e-mail: j.e.r.hov@medisin.uio.no
Telephone number: +47 23073627

Keywords:
Primary Sclerosing Cholangitis
Microbiota
Cross-Sectional Survey

Word count: 4214

Abbreviations: PSC, primary sclerosing cholangitis; IBD, inflammatory bowel disease; HC, healthy controls; UC, ulcerative colitis; PSC+IBD, primary sclerosing cholangitis with inflammatory bowel disease; PAMP, pathogen-associated molecular patterns; LPS, lipopolysaccharide; CD, Crohn's disease; BMI, body mass index; PPI, proton pump inhibitors; 5-ASA, 5-Aminosalicylic acid; INR, international normalised ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma-glutaryltransferase; P-ANCA, perinuclear anti-neutrophil cytoplasmic antibodies; FDR, false-discovery rate; OTU, operational taxonomic unit; Chao1, Chao1 bacterial richness estimate; AUC, area under the curve; ROC, receiver operating characteristics; AUROC, area under the receiver operating characteristics curve; NASH, nonalcoholic steatohepatitis; PBC, primary biliary cirrhosis; IBS, irritable bowel syndrome.

ABSTRACT

Objective: Gut microbiota could influence gut, as well as hepatic and biliary immune responses. We therefore thoroughly characterised the gut microbiota in primary sclerosing cholangitis (PSC) as compared with healthy controls (HC) and ulcerative colitis patients (UC) without liver disease.

Design: We prospectively collected 543 stool samples. After a stringent exclusion process, bacterial DNA was submitted for 16S rRNA gene sequencing. PSC and HC were randomised to an exploration or a validation panel, and only significant results ($P<0.05$, $Q_{FDR}<0.20$) in both panels were reported, followed by a combined comparison of all samples against UC.

Results: PSC patients (N=85) had markedly reduced bacterial diversity compared to HC (N=263, $P<0.0001$), and a different global microbial composition compared to both HC ($P<0.001$) and UC (N=36, $P<0.01$). The microbiota of PSC patients with and without inflammatory bowel disease (IBD) was similar. Twelve genera separated PSC and HC, out of which 11 were reduced in PSC. However, the *Veillonella* genus showed a marked increase in PSC compared to both HC ($P<0.0001$) and UC ($P<0.02$). Using ROC analysis, *Veillonella* abundance yielded an AUC of 0.64 to discriminate PSC from HC, while a combination of PSC associated genera yielded an AUC of 0.78.

Conclusion: PSC patients exhibited a gut microbial signature distinct from both HC and UC without liver disease, but similar in PSC with and without IBD. The *Veillonella* genus, which is also associated with other chronic inflammatory and fibrotic conditions, was enriched in PSC.

Summary box - Significance of this study

What is already known about this subject:

- The environmental factors influencing the risk of primary sclerosing cholangitis (PSC) are mostly unknown.
- PSC has a striking association with inflammatory bowel disease (IBD), with up to 80% of patients being affected.
- Several studies have found a distinct gut microbiota in inflammatory and immune-mediated diseases.
- One small study of the ileocecal mucosal microbiota exists in PSC, showing a slightly altered gut microbiota.

What are the new findings?

- PSC patients showed a distinct gut microbial profile, separate from healthy controls and ulcerative colitis patients without liver disease, with marked enrichment of the *Veillonella* genus.
- The intra-individual bacterial diversity of PSC patients was markedly reduced compared to healthy controls.
- The gut microbiota was similar in PSC with and without IBD.
- Ursodeoxycholic acid treatment was not associated with an altered gut microbiota in PSC.

How might it impact on clinical practice in the foreseeable future?

- Our findings provide a rationale for studies of the gut microbiota as a biomarker of disease and as a new treatment target in PSC, which could have clinical relevance.

INTRODUCTION

Primary sclerosing cholangitis (PSC) is a chronic, cholestatic liver disease of unknown aetiology, characterised by inflammation and fibrosis of the biliary tree.[1] Up to 80% of PSC patients have concomitant inflammatory bowel disease (IBD), most often categorised as ulcerative colitis (UC).[1] Multiple genetic risk factors have been identified in PSC;[2] however, established risk genes in PSC collectively explain only a small fraction of the disease liability,[2] suggesting that environmental factors are important in PSC development. The known environmental risk factors are limited to smoking and coffee, both protective against PSC,[3,4] highlighting the need for further research.

Early treatment-trials in PSC engaged antibiotics,[5] and more recently both metronidazole (in combination with ursodeoxycholic acid) and vancomycin have been shown to reduce alkaline phosphatase in PSC patients.[6,7] So far no studies have shown a long-term benefit of antibiotics on hard end-points like liver transplantation or death,[8] but collectively, the data suggest that manipulation of the gut microbes could potentially influence the disease process. In addition, several observations suggest that the microbial contents of the gut, i.e. the gut microbiota, could be directly involved in the pathogenesis of PSC. Evidence for this includes animal models of small bowel bacterial overgrowth that show PSC-like changes in the liver, which can be counter-acted by antibiotics,[9] and cultured cholangiocytes from PSC patients that seem hypersensitive to pathogen-associated molecular patterns (PAMPs) such as lipopolysaccharide (LPS).[10] LPS also seems to accumulate in cholangiocytes from PSC patients [11] and cross-reactivity between the most common autoantibodies in PSC (anti-neutrophil antibodies) and bacterial proteins has previously been reported.[12]

The gut microbiota represents a metabolically highly active human “organ”.[13] Changes in the diversity and structure of the gut microbiota have lately been implicated in the pathogenesis of several metabolic and inflammatory conditions including gastrointestinal disorders like IBD, but also systemic disorders like diabetes, obesity and atherosclerosis.[13–18] In IBD, gut microbial profiles have been reported in both UC and Crohn's disease (CD) as being distinctly different from healthy controls (HC).[17,19] However, although PSC is strongly associated with IBD, our knowledge of the gut microbiota in PSC is limited.[20]

We therefore hypothesised that the faecal gut microbiota of PSC differs from that of healthy controls and ulcerative colitis patients without liver disease, and sought to investigate this by conducting a cross sectional cohort study, surveying the gut microbiota using high-throughput sequencing in a robust two-stage design.

MATERIALS AND METHODS

Participants

We performed a cross-sectional collection of stool samples from non-transplanted PSC patients at the Norwegian PSC Research Center (NoPSC) biobank at Oslo University Hospital Rikshospitalet, a tertiary care centre. The diagnosis of PSC was made according to clinical guidelines and typical findings on cholangiography or liver biopsy, and all PSC patients had undergone screening for IBD.[21] IBD diagnosis was based on findings at colonoscopy and histology and accepted criteria.[22] Routine biochemical parameters were retrieved from hospital databases, including platelets, creatinine, total bilirubin, albumin, international normalised ratio (INR), aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, gamma-glutaryltransferase, together with perinuclear anti-neutrophil cytoplasmic antibody (P-ANCA) status. Mayo risk scores were calculated using the algorithm for the revised Mayo risk score.[23] UC

patients without a medical history of liver disease, and in clinical remission, were recruited in an outpatient setting from Oslo University Hospital Ullevål (Oslo, Norway), a secondary care centre. HC were randomly selected from donors registered in the national Norwegian Bone Marrow Donor Registry (Oslo, Norway).

Ethics

The study was performed in accordance with the declaration of Helsinki. Written informed consent was obtained from all study participants. Ethical approval was obtained from Regional Committee for Medical and Health Research Ethics in South-Eastern Norway (reference number 2012/286b).

Sample collection, exclusion criteria and DNA extraction

Demographic data, medical history, diet and medication were assessed from a questionnaire (Supplementary information). For PSC and UC this was supplemented with, and controlled against patient records.

Stool samples were collected using Stool Collection Tubes with Stool DNA Stabilizer (Stratec Molecular GmbH, Berlin, Germany), and a standardised collection device was used by all participants after voiding.[24] Samples were then sent by mail to the study centre and frozen at minimum -20°C on arrival according to the recommendation of the manufacturer.

Samples with >72 hours from collection to freezer were excluded (time limit according to the manufacturer). All participants exposed to antibiotics the preceding four weeks, and participants with previous bowel resection, gastrointestinal stoma or on specific diets (e.g. vegan, vegetarian, gluten free and milk free diets) were also excluded (Supplementary table 1).

DNA was extracted using the PSP Spin Stool DNA Kit (Stratec Molecular GmbH, Berlin, Germany), according to the manufacturer instructions.

Library preparation and sequencing

Library preparations were performed in accordance with a well established protocol.[25] In short, libraries were constructed from PCR amplicons of the V3-V4 region of the 16S rRNA gene generated using unique dual-index primers for each sample and Accuprime Pfx SuperMix (Thermo Fisher Scientific, Waltham, MA, USA). Amplicons were then cleaned and normalised using the SequalPrep Normalization Plate Kit (Life Technologies, Carlsbad, CA, USA). Subsequent quality control was performed on a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) using the Agilent High Sensitive DNA Kit (Agilent Technologies, Santa Clara, CA, USA). In addition, the libraries were quantified using the KAPA Library Quantification Kit (Kapa Biosystems Ltd., London, UK). Libraries were submitted to the Norwegian Sequencing Centre (Oslo, Norway) for Illumina MiSeq sequencing using the v3 kit (San Diego, CA, USA).

Sequence processing and data quality control

Paired-end reads were overlapped and merged using FLASH (version 1.2.10).[26] Quality control, trimming and closed reference operational taxonomic unit (OTU) mapping to the Greengenes database (version 13.8, OTUs with 97% sequence similarity) were performed using the Quantitative Insights Into Microbial Ecology (QIIME) platform (version 1.8.0).[27,28] OTUs with a number of sequences <0.005% of the total number of sequences were discarded as recommended.[29] Samples with <8,000 reads were also discarded.

Statistical analysis

Comparison of categorical variables was performed using the Chi-square test, or Fisher's exact test where appropriate. Mann-Whitney U test was applied for continuous variables. For correlation analyses, Spearman's rank correlation test was utilised. False-discovery rate (FDR) was calculated according to Benjamini-Hochberg, FDR-corrected P-values were denoted Q_{FDR} and was used when preforming all untargeted screening analyses of different taxa. Calculations of rarefied alpha diversity (Chao1 bacterial richness estimate [Chao1], Shannon diversity index and Phylogenetic diversity) and beta diversity were performed in QIIME. Fold-change in relative abundance was calculated by dividing the mean relative abundance in each category.

All regression analyses of relative taxa abundances and under the under the receiver operating characteristics curves (AUROC) analyses were performed in the statistical programming language R (version 3.1.2). Analogous to previous studies,[30] the relative abundances were arcsine square root transformed before regression analyses. Area under the under curves (AUC) were calculated to evaluate the performance of the fitted logistic regression models. The AUCs were based on predicted probability of PSC for each individual, using the multivariate logistic regression coefficient estimates together with the individual's transformed relative abundances for each bacterial taxa included in the analysis. Differences between AUCs were compared according to the method of DeLong. Linear regression analyses of alpha diversity were performed in SPSS (version 22; IBM, NY, USA). The Linear discriminant analysis effect size tool (LEfSe, version 1.0) and Multivariate Association with Linear Models framework (MaAsLin, version 1.0.1, revision 13:4033a2ee4558) were accessed online from http://huttenhower.sph.harvard.edu/galaxy/ (The Huttenhower Lab, Department of Biostatistics, Harvard School of Public Health, MA, USA) using standard

parameters.[31,32] For MaAsLin; age, sex, smoking status, body mass index (BMI) and number of prescriptions of antibiotics the last 12 months before inclusion were used as covariates. Unless otherwise specified, all other calculations were performed in SPSS.

Before post-sequencing data quality control, PSC and HC were randomly assigned to either an exploration or a validation panel. During the statistical analyses, only significant findings in the exploration panel ($P<0.05$, $Q_{FDR}<0.20$) were repeated in the verification panel. Finally, all samples were joined in a combined panel for comparison against UC (Figure 1). When doing AUROC analyses in the two-stage design, we used coefficients based on the exploration panel when calculating AUC for the validation panel, thus avoiding over-optimistic AUCs due to overfitting.

RESULTS

In total, we collected samples from 144 PSC patients, 51 UC patients and 348 HC, of which 85 PSC, 36 UC and 263 HC samples were included in the final analyses after exclusions and data quality control (Figure 1 and Supplementary table 1). Fifty-five of the PSC patients (64.7%) had concomitant IBD (44 [51.8%] with UC and 11 [12.9%] with CD, Supplementary table 2), and 23 (27.1%) had concomitant autoimmune disease. As shown in Table 1, PSC and HC were comparable in both the exploration and the validation panel, except a higher proportion of males in PSC in the validation panel and lower frequency of smokers in PSC in the exploration panel. PSC in the validation panel also had a slight increase in platelet count, compared to PSC in the exploration panel.

	Exploration panel					Validation panel					PSC E vs V	HC E vs V
	PSC (N=41)		HC (N=124)		P-value	PSC (N=44)		HC (N=139)		P-value	P-value	P-value
Age, median years (min-max)	50.0	(23-82)	46.0	(31-61)	0.12	48.0	(21-69)	47.0	(30-61)	0.15	0.94	0.40
Sex, male (%)	24	(58.5)	51	(41.1)	0.07	29	(65.9)	57	(41.0)	<0.01	0.51	1.00
BMI, median kg/m^2 (min-max)	25.5	(18-37)	25.9	(18-39)	0.86	24.4	(18-38)	25.5	(19-43)	0.17	0.80	0.61
Smoking, yes (%)	1	(2.4)	17	(13.7)	<0.05	1	(2.3)	13	(9.4)	0.19	1.00	0.33
Sample time in RT, median hours (min-max)	27.0	(15-63)	32.0	(18-72)	0.51	27.5	(13-59)	34.0	(16-71)	0.28	1.00	0.65
Courses of AB < 12 months, median (min-max)	0	(0-7)	0	(0-5)	0.18	0.0	(0-10)	0	(0-3)	0.79	0.15	0.82
Ulcerative colitis, n (%)	19	(46.3)				25	(56.8)				0.39	
Crohn's disease, n (%)	8	(19.5)				3	(6.8)				0.11	
Medication, yes, n (%)												
PPI	4	(9.8)	12	(9.7)	1.00	2	(4.5)	6	(4.3)	1.00	0.42	0.09
Antihistamines	2	(4.9)	8	(6.5)	1.00	4	(9.1)	6	(4.3)	0.26	0.68	0.58
Statins	3	(7.3)	5	(4.0)	0.41	4	(9.1)	6	(4.3)	0.26	1.00	1.00
Ursodeoxycholic acid	13	(31.7)				12	(27.3)				0.81	
Prednisolon	4	(9.8)				9	(20.5)				0.23	
5-ASA	19	(46.3)				16	(36.4)				0.39	
Infliximab						1	(2.3)				1.00	
Azathioprine	4	(9.8)				8	(18.2)				0.36	
Budesonide	2	(4.9)				1	(2.3)				0.61	
Disease specific variables												Available for
PSC Disease duration, median years (min-max)	10.6	(1-32)				6.6	(1-26)				0.12	N=85
IBD duration, median years (min-max)	10.3	(0-41)				11.9	(2-45)				0.95	N=63
Other autoimmune disease, yes (%)	12	(29.3)				11	(25.0)				0.63	N=80
Mayo risk score, median (min-max)	-0.04	(-1.5-3.2)				-0.06	(-1.9-3.3)				0.98	N=64
P-ANCA status, positive (%)	21	(51.2)				23	(52.3)				0.65	N=77
Platelet count, 10^9/L, median (min-max)	210	(57-578)				256	(62-432)				<0.05	N=77
Creatinine, µmol/L, median (min-max)	69	(42-106)				68	(42-100)				0.98	N=78
Bilirubin, µmol/L, median (min-max)	13	(5-101)				13	(3-114)				0.61	N=77
Albumin, g/L, median (min-max)	43	(24-47)				43	(16-47)				0.26	N=72
INR, median (min-max)	1.0	(0.9-1.3)				1.0	(0.9-1.3)				0.44	N=68
AST, U/L, median (min-max)	59	(19-197)				43	(18-172)				0.49	N=72
ALT, U/L, median (min-max)	59	(14-331)				48	(14-320)				0.66	N=78
ALP, U/L, median (min-max)	178	(50-598)				142	(30-589)				0.40	N=77
GGT, U/L, median (min-max)	201	(10-1576)				175	(12-1091)				0.86	N=76

Table 1: Demographics for the exploration and the validation panel. PSC, primary sclerosing cholangitis; HC, healthy controls; E, exploration panel; V, validation panel; BMI, body mass index; RT, room temperature; AB, antibiotics; IBD, inflammatory bowel disease; PPI, proton pump inhibitors; 5-ASA, 5-Aminosalicylic acid; INR, international normalised ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma-glutaryltransferase.

Reduced intra-individual bacterial diversity in PSC

The gut microbiota in PSC was significantly less diverse compared to HC in the exploration panel, as measured by Shannon diversity index (Figure 2A). This was also the case for the other diversity measurements (Chao1: 385 vs 472, $P<0.0001$ and Phylogenetic diversity: 21.6 vs 25.5, $P<0.0001$, Supplementary figure 1). These findings were all confirmed in the validation panel ($P<0.0001$, Figure 2B and Supplementary figure 1). In the combined panel, PSC and UC showed similar reduced diversity compared with HC (Figure 2C and Supplementary figure 1).

All findings were still significant after adjusting for age, sex, smoking status, BMI and number of prescriptions for antibiotic the last 12 months before inclusion (combined panel, Supplementary table 3). In HC, the use of antibiotics the last 12 months (N=44, 16.7%) was associated with reduced alpha diversity; Shannon diversity index: 6.0 vs 6.2, Chao1: 418 vs 472 and Phylogenetic diversity: 23.5 vs 25.6, all $P<0.02$, while there was no effect on diversity by antibiotic use in PSC (Supplementary figure 2A) or UC. The 74 PSC patients (87.1%) without antibiotics use the last 12 months showed a reduction in diversity also when compared only with the subgroup of HC who had used antibiotics the last year, and similar diversity compared with the other PSC patients (Supplementary figure 2A).

The *Veillonella* genus is increased in PSC

In total we identified 160 different taxa in the included samples. At the genus level, 18 taxa were different when comparing PSC and HC in the exploration panel ($P<0.05$ and $Q_{FDR}<0.20$), and we were able to replicate 12 of these in the validation panel (Figure 3). Eleven of these genera were reduced in PSC compared to HC. However, the last genus, *Veillonella* showed a 4.8-fold increase in PSC compared to HC ($P<0.0001$) and a 7.8-

fold increase compared to UC ($P<0.01$) in the combined panel. Eleven of the 12 genera separating PSC and HC (including *Veillonella*) were still significantly different after adjusting for age, sex, smoking status, BMI and number of prescriptions for antibiotics in the last 12 months before inclusion (Supplementary table 4). We were further able to trace 93% of the sequences in the *Veillonella* genus to two distinct species, *Veillonella dispar* (82%) and *Veillonella parvula* (11%), both increased in PSC compared to HC and UC ($P<0.0001$ and $P<0.01$, respectively in the combined panel).

In addition to the *Veillonella*, when performing an exploratory comparison of all taxa between PSC and UC, six other genera had different abundances in these phenotypes in the combined panel; *Akkermansia*, *Clostridium* and an unidentified genus in the Ruminococcaceae family were enriched in PSC, while *Dorea*, *Oscillospira* and *Citrobacter* were enriched in UC (all: $P<0.05$ and $Q_{FDR}<0.20$).

To further validate our analytical strategy we applied the standard microbiota analysis tools LEfSe and MaAsLin on the combined dataset. Together they confirmed nine of the 12 genera differing between PSC and HC in the initial analyses, including enrichment of *Veillonella* in PSC compared with both HC and UC (Supplementary table 4). MaAsLin also reported a negative association between BMI and the relative abundance of the Christensenellaceae family (coefficient: -0.001, $P<0.01$, q-value<0.02).

Global microbiota composition in PSC differs from HC and UC

The over-all microbial composition of PSC patients showed a clear shift compared to HC in both panels when analysing measures of beta diversity (Supplementary figure 3A&B). In concordance, this was also the case in the combined panel as shown in Figure 4A (unweighted UniFrac, PERMANOVA: pseudo-F-statistic: 12.2, $P<0.001$). We were able to identify alpha diversity as one of the factors driving the differences

along principal component 1 (Figure 4B). PSC patients also showed global differences compared to UC (unweighted UniFrac, PERMANOVA: pseudo-F-statistic: 2.6, P<0.01).

To evaluate the potential of using gut microbiota profiles to separate clinical phenotypes, we first performed AUROC analysis using the relative abundance of the *Veillonella* genus to distinguish PSC from HC, giving an AUC of 0.61 (95% CI: 0.51-0.72, P<0.05) in the exploration panel and 0.67 (0.58-0.76, P<0.001) in the validation panel, and an AUC of 0.64 (0.58-0.71, P<0.0001) in the combined panel, as shown in Figure 5A. Using only the nine genera that were significantly different between HC and PSC, and that were confirmed by all validation methods (linear regression, LEfSe and MaAsLin, Supplementary table 4) increased the AUC to 0.78 in all panels (95% CI: 0.70-0.87, 0.72-0.87 and 0.73-0.84 in the exploration, validation and combined panel, respectively, all P<0.0001 [Figure 5A and Supplementary figure 4A]). When performing 1000 permutations that randomly assigned samples to the exploration or validation panel, we found similar results as in the two cohorts separately (Supplementary figure 4B). Using the same strategy comparing PSC and UC, the *Veillonella* genus alone gave an AUC of 0.65 (0.54-0.75, P<0.05, Figure 5B), and the 7 genera that differed between PSC and UC in the explorative analyses gave an AUC of 0.82 (CI: 0.73-0.90, P<0.0001, Figure 5B).

The effect of disease severity, IBD status and drugs on the gut microbiota in PSC

We then analysed all included PSC patients separately (N=85) to explore subphenotypes. Only three of the PSC patients (3.5%) in the cohort were classified as small duct PSC and these did not deviate significantly from large duct PSC for any of the parameters studied (Supplementary table 5). Also, excluding small duct PSC from the dataset did not change any of the results in the study in a way affecting the interpretation. There were no differences between PSC patients with and without IBD in

regard to alpha diversity (Figure 2D and Supplementary table 5) or beta diversity (Supplementary figure 3C). Adjusting for age, sex, smoking status, BMI, number of prescriptions for antibiotic the last 12 months before inclusion, duration of PSC and duration of IBD did not alter the results (Supplementary table 3). All alpha diversity measurements were also similar in PSC patients with IBD irrespective of subtype (UC or CD, Figure 2E and Supplementary table 5), and IBD subtypes were also similar in regard to beta diversity (unweighted UniFrac). Male and female PSC patients showed no difference in regard to alpha or beta diversity (Supplementary figure 2B&C). We were also unable to identify any taxa that differed between PSC with and without IBD, PSC with and without concomitant autoimmune disease, or male and female PSC patients. We found no associations between alpha diversity and duration of PSC disease, age at diagnosis, liver biochemistry, P-ANCA status or Mayo risk score, but identified a negative correlation between a cluster of unknown genera in the Clostridiaceae family and duration of PSC disease (r: −0.41, Q_{FDR}<0.05). There was a positive correlation between the *Veillonella* genus and Mayo risk score (r: 0.25, uncorrected P<0.05). Six of the PSC patients (7.0%) had undergone liver transplantation after inclusion in the study, and these patients had higher abundance of *Veillonella* compared to other PSC patients (uncorrected P<0.02, Supplementary figure 2D). PSC patients without any medication (N=12) had similar alpha diversity compared to other PSC patients (Supplementary table 5), decreased alpha diversity compared to HC (all measurements: P<0.01), and they did not separate from other PSC patients in the beta diversity plot (Supplementary figure 3D). None of the individual medications registered in the PSC group, including ursodeoxycholic acid, had a significant impact on alpha diversity (Supplementary table 3 and 5).

DISCUSSION

In this cross-sectional cohort we have identified large differences in the gut microbiota of PSC patients compared to controls, with reduced alpha diversity and different abundance of several bacterial taxa. PSC and UC without liver disease were more similar, but still significantly different when measuring global overlap, and one bacterial genus, *Veillonella*, was highly enriched in PSC compared to both HC and UC. In addition, the presence of IBD did not influence the gut microbiota in PSC, suggesting that the associations are primarily accounting for the liver and bile duct affections.

The most striking feature of the PSC gut microbiota compared with HC was the reduced intra-individual bacterial diversity, and importantly, this was not related to the use of antibiotics during the last 12 months. Reduced bacterial diversity in stool has been observed in several other inflammatory and metabolic conditions, like IBD, type 1 diabetes, arthritis and obesity.[13,15,16,33] Our data are in line with reduced diversity observed in a Dutch study of the microbiota of gut mucosal biopsies in 12 PSC patients.[20] Besides often being associated with the healthy state, high diversity has been described as a driving force for evolution of the immune system, allowing the host to accommodate environmental antigens and possibly self-antigens;[34] however, reduced bacterial diversity does not seem to be the case for liver disease in general, e.g. nonalcoholic steatohepatitis (NASH) or cirrhosis,[35,36] suggesting that the observed differences are related to the characteristics of PSC. The concept that PSC is associated with a certain gut microbiota profile is further supported by the results that a combination of the most robust associated bacterial taxa gave an AUC of 0.78 when separating PSC and HC, and 0.82 separating PSC and UC. This is in line with findings in treatment-naïve Crohn's disease with AUC of 0.66-0.85,[19] and an AUC of 0.71-0.83 in type 2 diabetes.[37,38]

Altered abundance of several genera contributed to the unique gut microbial signature found in PSC in our study, the most prominent being the marked enrichment of the *Veillonella* genus, an obligate anaerobic, gram-negative coccus, sensitive to metronidazole, but not vancomycin.[39] It has been associated with other inflammatory and progressive fibrotic conditions like in pulmonary cystic fibrosis and idiopathic pulmonary fibrosis (lung microbiota), and recurrence of disease in CD patients undergoing ileocecal resection (mucosal biopsies).[40–42] While suggesting that *Veillonella* has a role in the aetiology of PSC would be speculative, the robust association with PSC and links to fibrosis in other phenotypes warrant further study.

Although limited information is available regarding the effects on the microbiota of cholestasis and cholangiopathies in humans, in experimental cholestasis in animals (using bile duct ligation), increased bacterial translocation and systemic endotoxin levels have been observed, but with a minor contribution of the microbiota.[43] The multidrug resistance 2 knockout (*mdr2*$^{-/-}$) mouse model of PSC on the other hand, showed marked exacerbation of their hepatobiliary disease when raised in a germ free environment.[44] *In vitro* data suggested that this in part could be explained by the absence of commensal microbial metabolites,[44] e.g. secondary bile acids. Secondary bile acids have anti-inflammatory properties *in vitro,*[45] and decreased levels of these secondary bile acids compared to healthy controls has been reported in PSC patients, and in IBD patients (both UC and CD) during disease flares.[44–46] On the other hand, secondary bile acids are not reduced in primary biliary cirrhosis (PBC), another human cholestatic disease.[46] Together this highlights the need for future studies to assess the involvement of the gut microbiota in cholestasis.

While the gut microbiota of UC patients without PSC was different from HC in our study, as in other studies,[17] it was highly similar in PSC irrespective of IBD status.

This could suggest that the dysbiosis in PSC patients is related to the liver disease, and not IBD. This is supported by the observation that IBD subtype in PSC patients did not influence the gut microbial profile, which contrasts the differences between UC and CD in patients without PSC.[16] It is also possible that very subtle IBD phenotypes in the PSC patients not discovered by endoscopy screening could be present and influence the gut microbiota, or the gut microbiota profile in PSC may influence the accompanying IBD. In this regard it should be noted that IBD in PSC has several characteristics that differ from that of IBD patients without liver disease.[1,47–50]

A crucial question, which can only be speculated upon, is whether disease or its treatment is causing changes in the microbiota; do the microbiota alterations represent an actual link between the gut and the liver in PSC, or are they secondary to advanced liver disease? There were no correlations between PSC duration or biochemical parameters and diversity in the present study, speaking against the latter, although a link between particularly high prevalence of *Veillonella* and more severe liver disease cannot be excluded. Data on pre-clinical microbiota profile in inflammatory diseases are scarce, but data from type 1 diabetes and CD suggest that gut microbiota changes could precede at least the clinical onset of disease.[15,19] Overall, it is therefore possible that the observed microbiota alterations in PSC are involved in disease development.

Our knowledge of the effects of pharmacological agents on the microbiota is limited. We were unable to detect any effect of medication on the gut microbiota of our PSC patients, including the use of ursodeoxycholic acid. 5-Aminosalicylic acid (5-ASA) was used by 41% of the PSC patients in our study, and it has previously been shown to cause a decrease in microbial diversity in irritable bowel syndrome (IBS),[51] but did not show any effect on diversity in our PSC patients. Neither was there any association between antibiotics the previous year and decreased diversity in the PSC, contrasting our

findings in the HC group. This could suggest that antibiotics, to a lesser extent than the disease in itself, are affecting the microbial diversity; however, long lasting effects of antibiotics in the PSC group may be present.[52,53]

The major strengths of this study are the inclusion of a large number of PSC patients and controls, the use of a standardised collection procedure and state-of-the art library preparation and sequencing methods. Power calculations for gut microbiota studies are challenging and not well developed, in part because little is known about the effect sizes to be expected and a large number of different bacterial taxa present. The number of available samples therefore determined the study size. It will be important to increase statistical power in further studies to explore the role of the gut microbiota in subphenotypes in PSC.

To reduce the risk of false positive results, we applied a conservative two-stage design in our analysis when comparing PSC and HC, performed multivariate linear regressions and validated the analyses by applying other published tools. In regard of validation of data quality, we were able to reproduce key features of the UC gut microbiota (reduced diversity and depletion of the *Akkermansia* genus),[54,55] and the association between the abundance of Christensenellaceae and BMI.[56] A question to be resolved is the importance of missing detailed dietary history in this and other microbiota studies, but no standardised method for adjusting microbiota data using diet exists.[57] In the present study this was handled by removing all individuals reporting to have structured adjustments in their diet, e.g. vegetarians and gluten-free, but it cannot be excluded that subtle, undetected dietary factors could have influenced our results.

In conclusion, PSC patients showed a distinct gut microbial signature, clearly separate from both HC and UC without liver disease, but similar in PSC with and without IBD. The *Veillonella* genus, which is also associated with other chronic inflammatory and

fibrotic conditions, was highly enriched in PSC. Overall, this study provides a basis and rationale for further studies of the microbiota both related to pathophysiology and clinical utility in PSC, with the potential to improve patient care.

Acknowledgements

The authors would like to thank Turid Bua, study nurse at the Department of Gastroenterology, Oslo University Hospital Ullevål, for help with recruiting the ulcerative colitis controls, Kristian Alfsnes, Eva Kristin Schjelderup and Mette Nyberg for contributing to DNA extractions, Hege Dahlen Sollid, Mona Bjørnstad, Liv Wenche Torbjørnsen for help at the Norwegian PSC Research Center, Fredrik Bäckhed for helpful discussions and the Norwegian Sequencing Centre for help with sequencing. An Illumina MiSeq was covered by a grant from Fougner-Hartmann's family fund. Benedicte A. Lie and the Norwegian Bone Marrow Donor Registry are acknowledged for providing access to healthy controls.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: Dr. Hov reports grants from Norwegian Research Council; Dr. Karlsen reports grants from Western Norway Regional Health Authority, during the conduct of the study. All other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Funding

THK was supported by a grant from Western Norway Regional Health Authority (grant number 911802). JRH is funded by the Norwegian PSC Research Center and the Norwegian Research Council (project number 240787/F20).

References

1 Hirschfield GM, Karlsen TH, Lindor KD, et al. Primary sclerosing cholangitis. *Lancet* 2013;382:1587–99 doi:10.1016/S0140-6736(13)60096-3 [published Online First: 28 June 2013].

2 Henriksen EKK, Melum E, Karlsen TH. Update on primary sclerosing cholangitis genetics. *Curr Opin Gastroenterol* 2014;30:310–9 doi:10.1097/MOG.0000000000000052.

3 Andersen IM, Tengesdal G, Lie BA, et al. Effects of coffee consumption, smoking, and hormones on risk for primary sclerosing cholangitis. *Clin Gastroenterol Hepatol* 2014;12:1019–28 doi:10.1016/j.cgh.2013.09.024 [published Online First: 25 September 2014].

4 Loftus EV, Sandborn WJ, Tremaine WJ, et al. Primary sclerosing cholangitis is associated with nonsmoking: a case-control study. *Gastroenterology* 1996;110:1496–502 doi:10.1053/gast.1996.v110.pm8613055.

5 Rankin JG, Boden RW, Goulston SJ, et al. The liver in ulcerative colitis; treatment of pericholangitis with tetracycline. *Lancet* 1959;274:1110–2 doi:10.1016/S0140-6736(59)90098-4.

6 Tabibian JH, Weeding E, Jorgensen RA, et al. Randomised clinical trial: vancomycin or metronidazole in patients with primary sclerosing cholangitis - a pilot study. *Aliment Pharmacol Ther* 2013;37:604–12 doi:10.1111/apt.12232 [published Online First: 5 February 2013].

7 Färkkilä M, Karvonen A-L, Nurmi H, et al. Metronidazole and ursodeoxycholic acid for primary sclerosing cholangitis: a randomized placebo-controlled trial. *Hepatology* 2004;40:1379–86 doi:10.1002/hep.20457.

8 Tabibian JH, Talwalkar JA, Lindor KD. Role of the microbiota and antibiotics in primary sclerosing cholangitis. *Biomed Res Int* 2013;2013:389537. doi:10.1155/2013/389537 [accessed 17 August 2014].

9 Lichtman SN, Keku J, Clark RL, et al. Biliary tract disease in rats with experimental small bowel bacterial overgrowth. *Hepatology* 1991;13:766–72 doi:10.1002/hep.1840130425.

10 Mueller T, Beutler C, Picó AH, et al. Enhanced innate immune responsiveness and intolerance to intestinal endotoxins in human biliary epithelial cells contributes to chronic cholangitis. *Liver Int* 2011;31:1574–88 doi:10.1111/j.1478-3231.2011.02635.x [published Online First: 15 September 2011].

11 Sasatomi K, Noguchi K, Sakisaka S, et al. Abnormal accumulation of endotoxin in biliary epithelial cells in primary biliary cirrhosis and primary sclerosing cholangitis. *J Hepatol* 1998;29:409–16 doi:10.1016/S0168-8278(98)80058-5.

12 Terjung B, Söhne J, Lechtenberg B, et al. p-ANCAs in autoimmune liver disorders recognise human beta-tubulin isotype 5 and cross-react with microbial protein FtsZ. *Gut* 2010;59:808–16 doi:10.1136/gut.2008.157818 [published Online First: 1 December 2009].

13 Turnbaugh PJ, Hamady M, Yatsunenko T, et al. A core gut microbiome in obese and lean twins. *Nature* 2009;457:480–4 doi:10.1038/nature07540 [published Online First: 30 November 2008].

14 Tyler AD, Smith MI, Silverberg MS. Analyzing the human microbiome: a “how to” guide for physicians. *Am J Gastroenterol* 2014;109:983–93 doi:10.1038/ajg.2014.73 [published Online First: 22 July 2014].

15 Giongo A, Gano KA, Crabb DB, et al. Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J* 2011;5:82–91 doi:10.1038/ismej.2010.92 [published Online First: 8 July 2010].

16 Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65 doi:10.1038/nature08821.

17 Machiels K, Joossens M, Sabino J, et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* 2014;63:1275–83 doi:10.1136/gutjnl-2013-304833 [published Online First: 10 September 2013].

18 Tang WHW, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 2013;368:1575–84 doi:10.1056/NEJMoa1109400 [published Online First: 25 April 2013].

19 Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn’s disease. *Cell Host Microbe* 2014;15:382–92 doi:10.1016/j.chom.2014.02.005 [published Online First: 12 March 2014].

20 Rossen NG, Fuentes S, Boonstra K, et al. The mucosa-associated microbiota of PSC patients is characterized by low diversity and low abundance of uncultured Clostridiales II. *J Crohns Colitis* 2015;9:342–8 doi:10.1093/ecco-jcc/jju023 [published Online First: 28 December 2014].

21 European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. *J Hepatol* 2009;51:237–67 doi:10.1016/j.jhep.2009.04.009 [published Online First: 6 June 2009].

22 Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol* 1989;24(Suppl 170):2–6 doi:10.3109/00365528909091339.

23 Kim WR, Therneau TM, Wiesner RH, et al. A revised natural history model for primary sclerosing cholangitis. *Mayo Clin Proc* 2000;75:688–94 doi:10.4065/75.7.688.

24 Ahlquist DA, Schwartz S, Isaacson J, et al. A stool collection device: the first step in occult blood testing. *Ann Intern Med* 1988;108:609–12 doi:10.7326/0003-4819-108-4-609.

25 Kozich JJ, Westcott SL, Baxter NT, et al. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013;79:5112–20 doi:10.1128/AEM.01043-13 [published Online First: 21 June 2013].

26 Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011;27:2957–63 doi:10.1093/bioinformatics/btr507 [published Online First: 7 September 2011].

27 Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–6 doi:10.1038/nmeth.f.303 [published Online First: 11 April 2010].

28 DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;72:5069–72 doi:10.1128/AEM.03006-05.

29 Navas-Molina JA, Peralta-Sánchez JM, González A, et al. Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol* 2013;531:371–444 doi:10.1016/B978-0-12-407863-5.00019-8.

30 The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14 doi:10.1038/nature11234 [published Online First: 13 June 2012].

31 Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13:R79. doi:10.1186/gb-2012-13-9-r79 [accessed 5 May 2013].

32 Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011;12:R60. doi:10.1186/gb-2011-12-6-r60 [accessed 15 March 2013].

33 Scher JU, Ubeda C, Artacho A, et al. Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. *Arthritis Rheumatol* 2015;67:128–39 doi:10.1002/art.38892 [published Online First: 27 December 2014].

34 Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 2006;124:837–48 doi:10.1016/j.cell.2006.02.017.

35 Jiang W, Wu N, Wang X, et al. Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease. *Sci Rep* 2015;5:8096. doi:10.1038/srep08096 [accessed 13 March 2015].

36 Chen Y, Yang F, Lu H, et al. Characterization of fecal microbial communities in patients with liver cirrhosis. *Hepatology* 2011;54:562–72 doi:10.1002/hep.24423 [published Online First: 26 June 2011].

37 Karlsson FH, Tremaroli V, Nookaew I, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 2013;498:99–103 doi:10.1038/nature12198 [published Online First: 29 May 2013].

38 Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 2012;490:55–60 doi:10.1038/nature11450 [published Online First: 26 September 2012].

39 Pérez-Jacoiste Asín MA, Fernández-Ruiz M, Serrano-Navarro I, et al. Polymicrobial endocarditis involving Veillonella parvula in an intravenous drug user: case report and literature review of Veillonella endocarditis. *Infection* 2013;41:591–4 doi:10.1007/s15010-012-0398-3 [published Online First: 1 March 2013].

40 Fodor AA, Klem ER, Gilpin DF, et al. The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. *PLoS One* 2012;7:e45001. doi:10.1371/journal.pone.0045001 [accessed 9 December 2014].

41 Molyneaux PL, Cox MJ, Willis-Owen SAG, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014;190:906–13 doi:10.1164/rccm.201403-0541OC [published Online First: 3 September 2014].

42 De Cruz P, Kang S, Wagner J, et al. Association between specific mucosa-associated microbiota in Crohn's disease at the time of resection and subsequent disease recurrence: a pilot study. *J Gastroenterol Hepatol* 2015;30:268–78 doi:10.1111/jgh.12694 [published Online First: 24 January 2015].

43 Fouts DE, Torralba M, Nelson KE, et al. Bacterial translocation and changes in the intestinal microbiome in mouse models of liver disease. *J Hepatol* 2012;56:1283–92 doi:10.1016/j.jhep.2012.01.019 [published Online First: 9 February 2012].

44 Tabibian JH, O'Hara SP, Trussoni CE, et al. Absence of the intestinal microbiota exacerbates hepatobiliary disease in a murine model of primary sclerosing cholangitis. *Hepatology* 2016;63:185-96 doi:10.1002/hep.27927 [published Online First: 10 August 2015].

45 Duboc H, Rajca S, Rainteau D, et al. Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. *Gut* 2013;62:531–9 doi:10.1136/gutjnl-2012-302578 [published Online First: 19 September 2012].

46 Trottier J, Białek A, Caron P, et al. Metabolomic profiling of 17 bile acids in serum from patients with primary biliary cirrhosis and primary sclerosing

cholangitis: a pilot study. *Dig Liver Dis* 2012;44:303–10 doi:10.1016/j.dld.2011.10.025 [published Online First: 9 December 2011].

47 Jørgensen KK, Grzyb K, Lundin KEA, et al. Inflammatory bowel disease in patients with primary sclerosing cholangitis: clinical characterization in liver transplanted and nontransplanted patients. *Inflamm Bowel Dis* 2012;18:536–45 doi:10.1002/ibd.21699 [published Online First: 31 March 2011].

48 Sinakos E, Samuel S, Enders F, et al. Inflammatory bowel disease in primary sclerosing cholangitis: a robust yet changing relationship. *Inflamm Bowel Dis* 2013;19:1004–9 doi:10.1097/MIB.0b013e3182802893 [published Online First: 14 March 2013].

49 Loftus EV, Harewood GC, Loftus CG, et al. PSC-IBD: a unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. *Gut* 2005;54:91–6 doi:10.1136/gut.2004.046615.

50 Fausa O, Schrumpf E, Elgjo K. Relationship of inflammatory bowel disease and primary sclerosing cholangitis. *Semin Liver Dis* 1991;11:31–9 doi:10.1055/s-2008-1040420.

51 Andrews CN, Griffiths TA, Kaufman J, et al. Mesalazine (5-aminosalicylic acid) alters faecal bacterial profiles, but not mucosal proteolytic activity in diarrhoea-predominant irritable bowel syndrome. *Aliment Pharmacol Ther* 2011;34:374–83 doi:10.1111/j.1365-2036.2011.04732.x [published Online First: 14 June 2011].

52 Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl*

Acad Sci U S A 2011;108(Suppl):4554–61 doi:10.1073/pnas.1000087107 [published Online First: 16 September 2010].

53 Dethlefsen L, Huse S, Sogin ML, et al. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* 2008;6:e280. doi:10.1371/journal.pbio.0060280 [accessed 17 June 2014].

54 Rajilić-Stojanović M, Shanahan F, Guarner F, et al. Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm Bowel Dis* 2013;19:481–8 doi:10.1097/MIB.0b013e31827fec6d [published Online First: 4 February 2013].

55 Png CW, Lindén SK, Gilshenan KS, et al. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol* 2010;105:2420–8 doi:10.1038/ajg.2010.281 [published Online First: 20 July 2010].

56 Goodrich JK, Waters JL, Poole AC, et al. Human genetics shape the gut microbiome. *Cell* 2014;159:789–99 doi:10.1016/j.cell.2014.09.053 [published Online First: 6 November 2014].

57 Eaton JE, Juran BD, Atkinson EJ, et al. A comprehensive assessment of environmental exposures among 1000 North American patients with primary sclerosing cholangitis, with and without inflammatory bowel disease. *Aliment Pharmacol Ther* 2015;41:980–90 doi:10.1111/apt.13154 [published Online First: 17 March 2015].

FIGURE LEGENDS

Figure 1: Study design. After a rigorous exclusion process, samples were submitted to sequencing. Before post-sequencing quality control, PSC and HC were randomly assigned to either an exploration or a validation panel. Samples that failed sequencing, or producing <8,000 reads were discarded. During the statistical analyses, only significant findings in the exploration panel ($P<0.05$ and $Q_{FDR}<0.20$) were repeated in the verification panel. Lastly, all samples were joined in a combined panel for comparison with UC. PSC, primary sclerosing cholangitis; UC, ulcerative colitis; QC, quality control.

Figure 2: (A and B) Alpha diversity, here illustrated by the Shannon diversity index, was consistently reduced in PSC patients compared with healthy controls across all panels, (C) and were similar in PSC and UC patients in the combined panel. (D) PSC patients without IBD showed similar bacterial diversity as PSC with IBD (PSC+IBD), (E) and all IBD subgroups in PSC had reduced diversity compared to healthy controls. PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls; IBD, inflammatory bowel disease; CD, Crohn's disease. Data shown as IQR + min, max. *P<0.05, ****P<0.0001.

Figure 3: The 12 genera confirmed in the validation panel, differing between PSC and HC, illustrated by the ratio between PSC and HC on a logarithmic scale. Ratio between UC and HC included for comparison. One genus, *Veillonella,* showed significant increase in PSC compared to both HC and UC (framed box). The ML615J-28, RF32 and YS2 order are part of the Tenericutes, Proteobacteria and Cyanobacteria phyla, respectively. All data in the chart is based on the combined panel. The ratios were calculated by dividing the mean relative abundance in the PSC/UC patients by the mean

relative abundance in HC. Data in the framed box is shown as median and inter-quartile range. PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls. *Q_{FDR}<0.05, **Q_{FDR}<0.01, ***Q_{FDR}<0.001, ****Q_{FDR}<0.0001.

Figure 4: (A) Beta diversity plot from the combined panel showing a clear separation of PSC from HC (unweighted UniFrac, PERMANOVA: pseudo-F-statistic: 12.2, P<0.001) and a more subtle separation of PSC and UC samples (unweighted UniFrac, PERMANOVA: pseudo-F-statistic: 2.6, P<0.01). (B) Same plot as in (A), but samples coloured according to their Shannon diversity index, showing diversity as an important factor along principle component 1 (PC1). PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls; PC, principal component.

Figure 5: (A) Gut microbiota-based PSC classification using AUROC analysis (in the combined panel) using the relative abundance of the *Veillonella* genus (AUC=0.64, 95% CI: 0.58-0.71, P<0.0001). When using each individual`s arcsine square root transformed abundance of the nine genera differing between PSC and HC (and validated by all secondary analyses, denoted PSC_{HC}), together with the coefficients from multivariate logistic regression, the AUC increased to 0.78 (0.73-0.84, P<0.0001). (B) Same analysis repeated for PSC and UC. The *Veillonella* genus alone gave an AUC of 0.65 (0.54-0.75, P<0.05), and the seven genera differing between PSC and UC (denoted PSC_{UC}) gave an AUC of 0.82 (0.73-0.90, P<0.0001). PSC, primary sclerosing cholangitis; HC, healthy controls; UC, ulcerative colitis; AUROC, area under the receiver operator characteristics curve; AUC, area under the curve.

FIGURES

Figure 1:

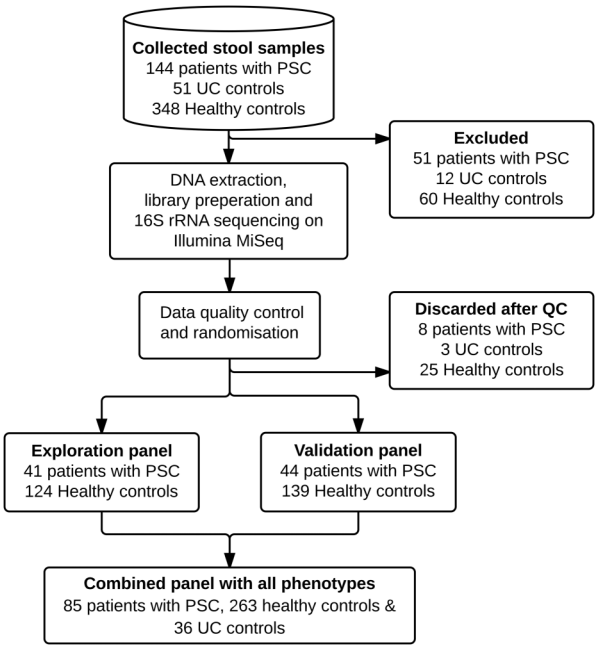


Figure 2:

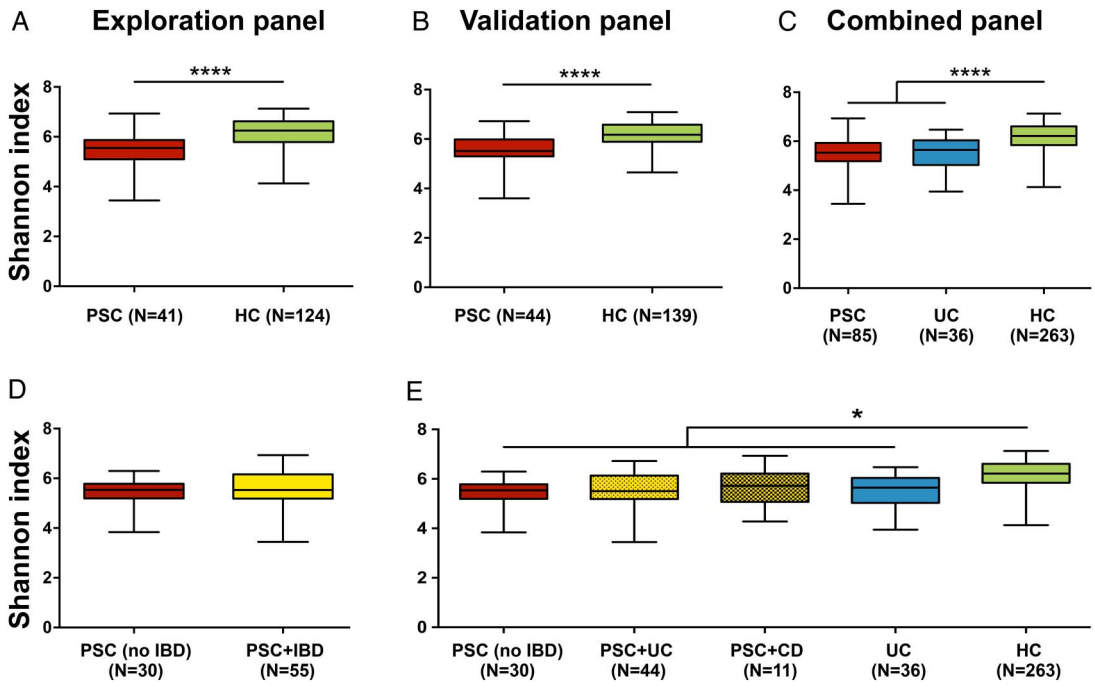


Figure 3:

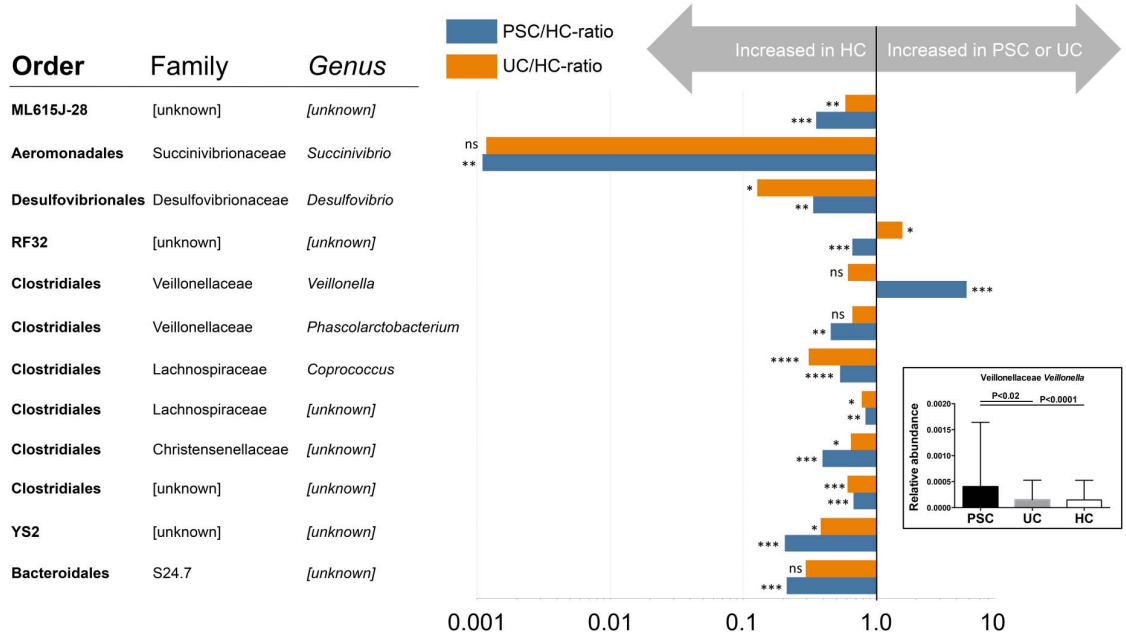


Figure 4:

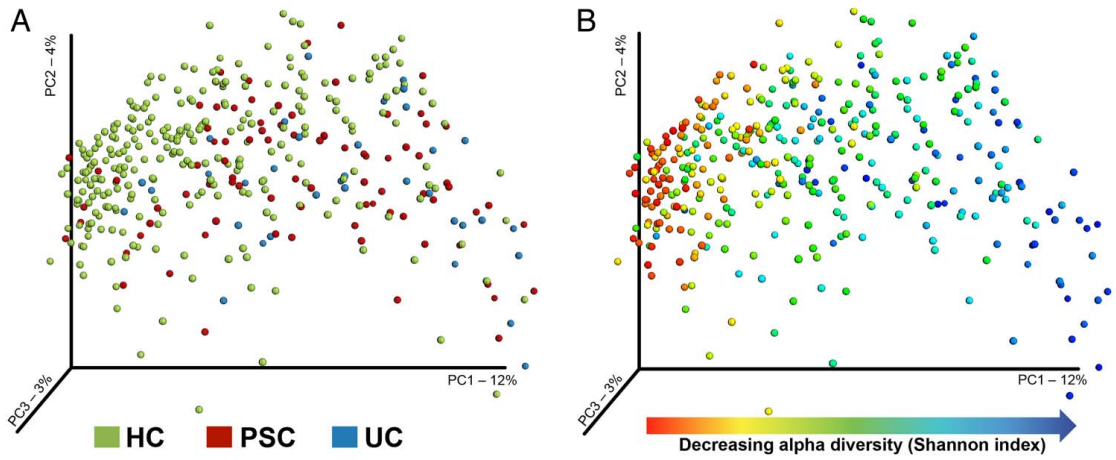
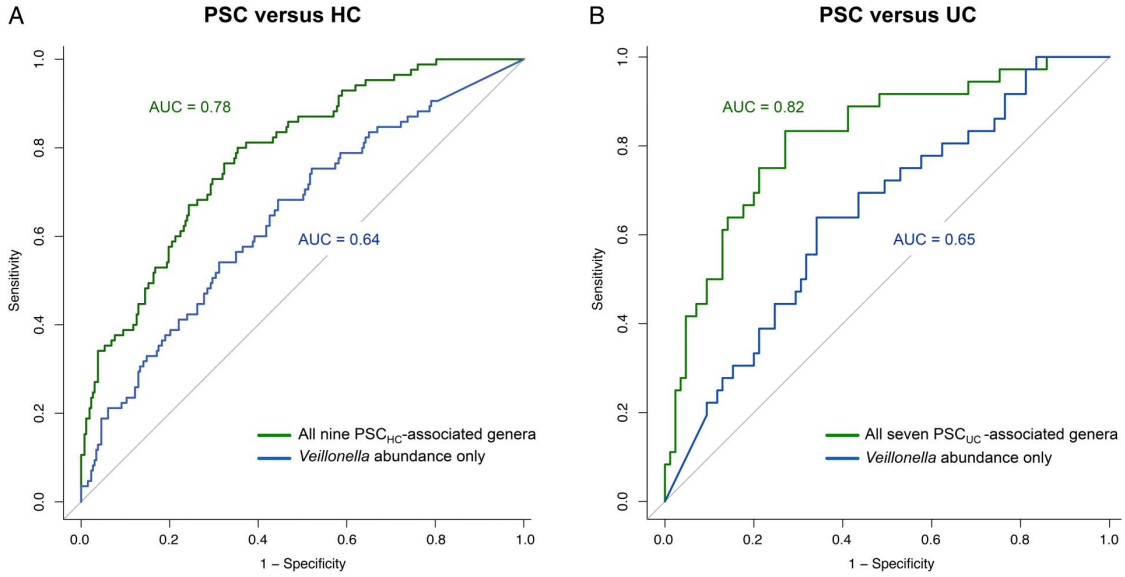


Figure 5:


A
PSC versus HC
AUC = 0.78
AUC = 0.64
All nine PSC$_{HC}$-associated genera
Veillonella abundance only
Sensitivity
1 – Specificity
0.0
0.2
0.4
0.6
0.8
1.0
B
PSC versus UC
AUC = 0.82
AUC = 0.65
All seven PSC$_{UC}$ -associated genera
Veillonella abundance only
Sensitivity
1 – Specificity

SUPPLEMENTARY INFORMATION AND FIGURES

The gut microbial profile in patients with primary sclerosing cholangitis is distinct from ulcerative colitis patients without biliary disease and healthy controls.

Martin Kummen[1,2,3,4], Kristian Holm[1,4], Jarl Andreas Anmarkrud[1,4], Ståle Nygård[5,6,7], Mette Vesterhus[1,8], Marte L. Høivik[9], Marius Trøseid[1,2,10], Hanns-Ulrich Marschall[11], Erik Schrumpf[1,3], Bjørn Moum[3,9], Helge Røsjø[7,12], Pål Aukrust[2,3,4,10], Tom H. Karlsen[1,2,3,4,13,14], Johannes R. Hov[1,2,3,4,14].

[1]Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; [2]K.G. Jebsen Inflammation Research Centre, University of Oslo, Oslo, Norway; [3]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [4]Research Institute of Internal Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; [5]Bioinformatics Core Facility, Institute for Medical Informatics, Oslo University Hospital, Oslo, Norway; [6]Institute for Experimental Medical Research, Oslo University Hospital and University of Oslo, Oslo, Norway; [7]K.G. Jebsen Cardiac Research Centre and Center for Heart Failure Research, University of Oslo, Oslo, Norway; [8]Department of Medicine, National Centre for Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway; [9]Department of Gastroenterology, Oslo University Hospital Ullevål, Oslo, Norway; [10]Section of Clinical Immunology and Infectious Diseases, Oslo University Hospital Rikshospitalet, Oslo, Norway; [11]Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [12]Division of Medicine, Akershus University Hospital, Lørenskog, Norway; [13]Department of Clinical Medicine, University of Bergen, Bergen, Norway; [14]Section of Gastroenterology, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway.

Table of Contents

1. Questionnaire used in the study (translated from Norwegian)

Time of sampling: Date: _____ Time: _____

1. Do you follow a special diet? Yes No

If yes, what diet (vegetarian, gluten free, lactose free, low carbohydrate etc.), specify shortly: _____

2. Have you used any antibiotics the last 4 weeks? Yes No

If yes, what type of antibiotic did you use and what dose?

_____

3. How many courses of antibiotic therapy have you taken last year (number and names of any kind)?

Quantity: Type:

4. Do you have pets? Yes No

If yes, specify in short (e.g. dog): _____

5. Have you smoked daily during the last month? Yes No

6. Have you ever removed parts of your intestines and/or do you have a stoma?

7. What is your height and weight?

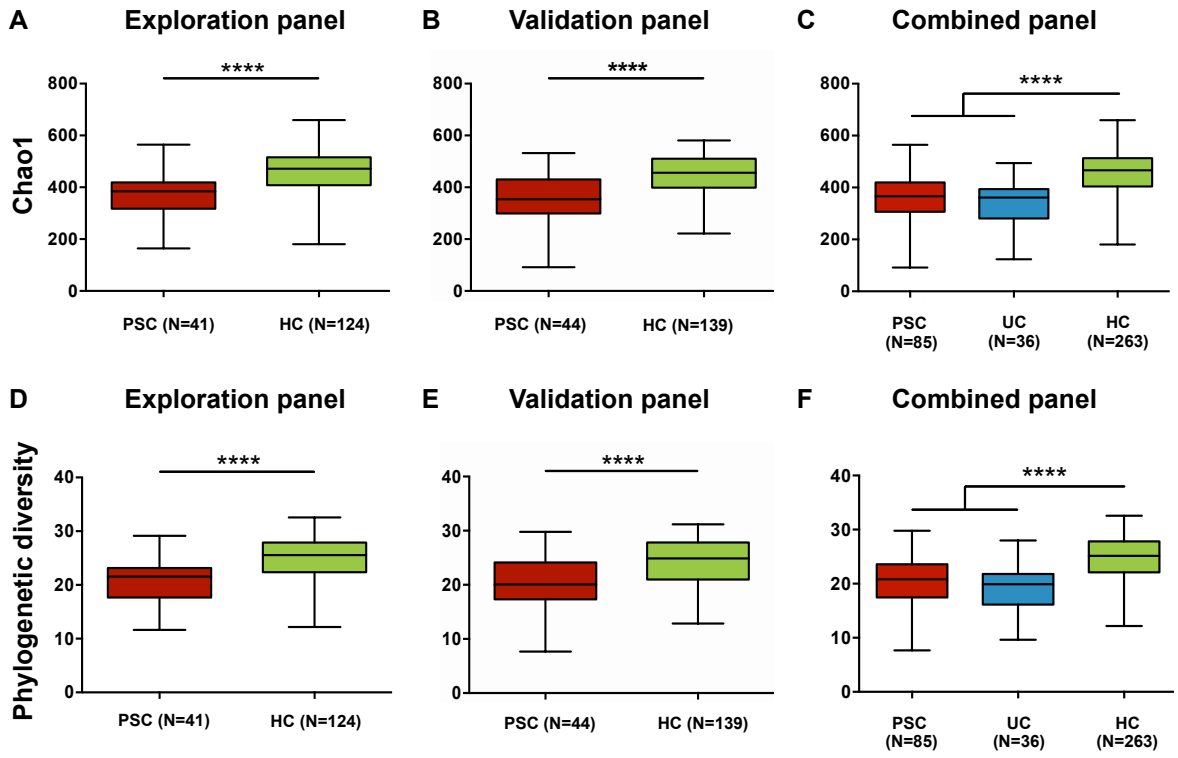
Height (cm): _____ Weight (kg): _____

Write down all your regular medication (dose is not necessary, just a trade name):

Comments:

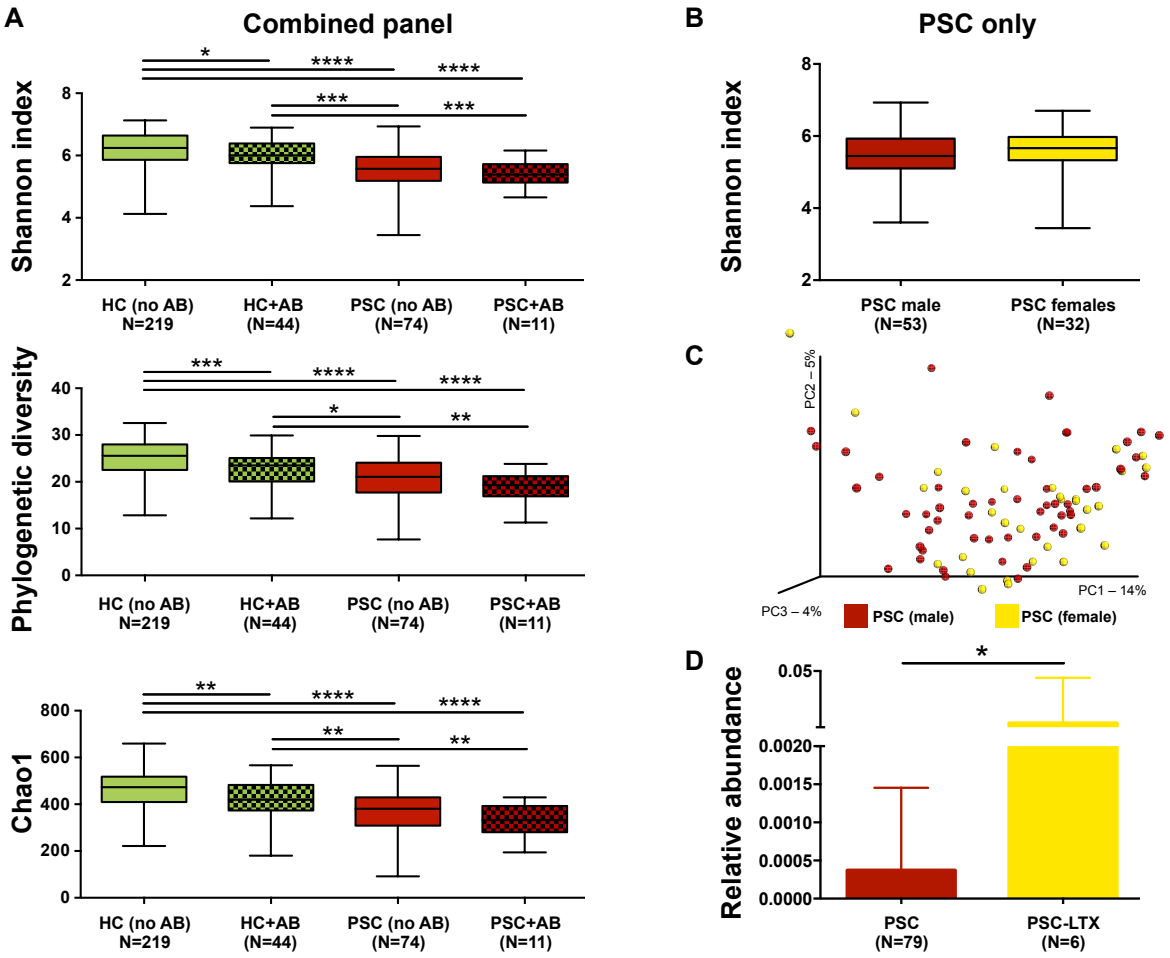
2. Supplementary figures

Supplementary figure S1



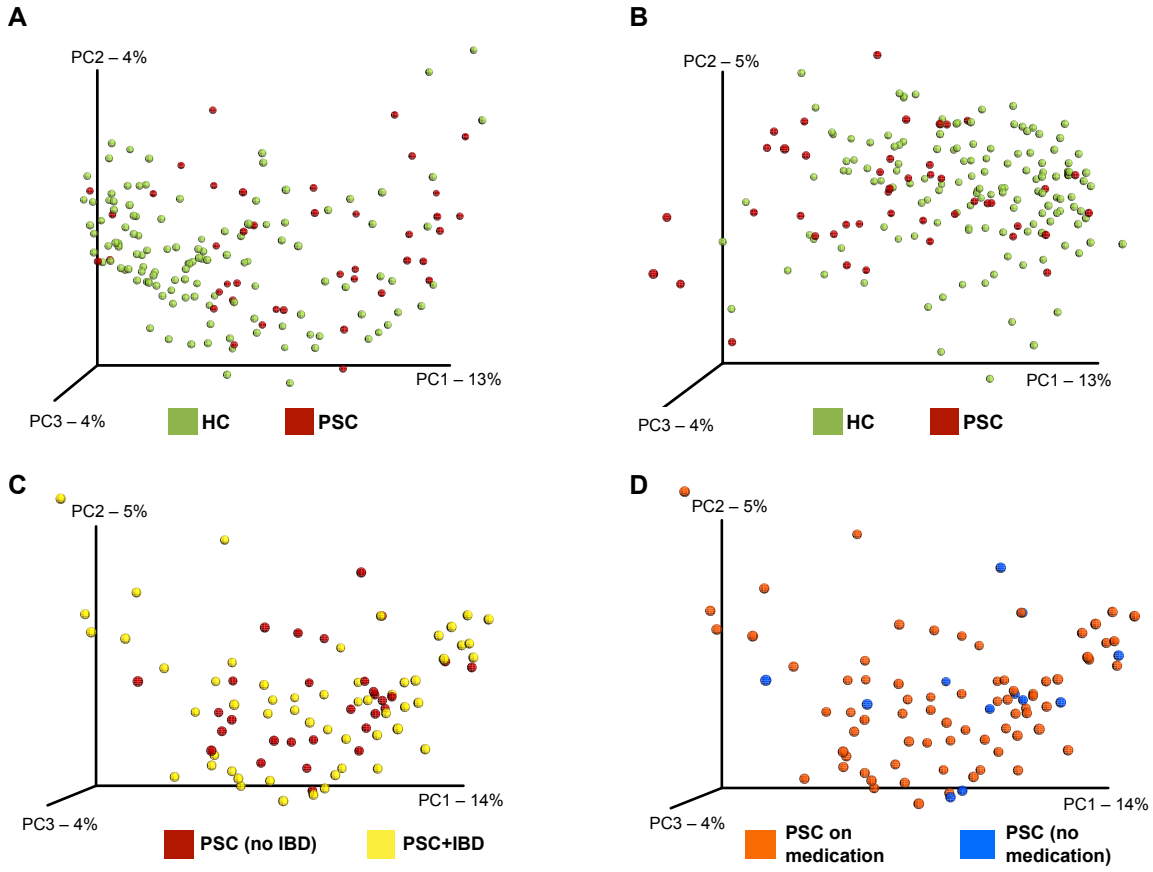
Alpha diversity: (A&B) Chao1 and (D&E) Phylogenetic diversity was consistently reduced in PSC patients compared with HC across all panels, (C&F) and were similar in PSC and UC patients in he combined panel. PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls; Chao1, Chao1 bacterial richness estimate. Panels show IQR + min, max. ****P<0.0001.

Supplementary figure S2



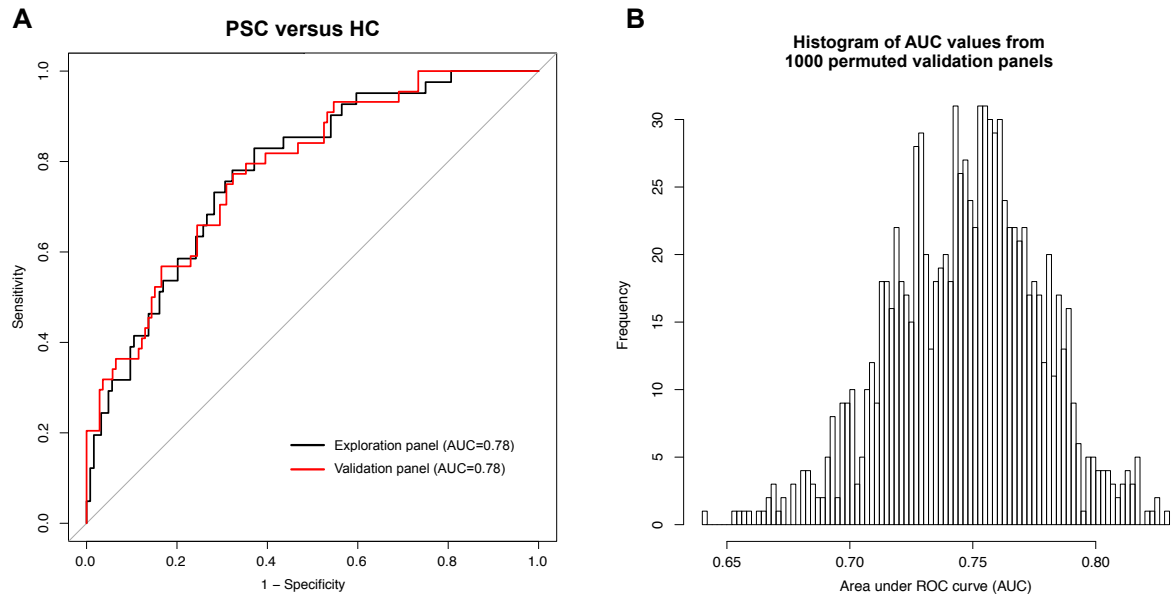
Alpha diversity: (A) Diversity and antibiotics the last 12 months before inclusion (participants who used antibiotics the last 4 weeks before inclusion were excluded all together). All diversity measurements were reduced in PSC irrespective of antibiotics, compared to HC with and without history of antibiotics use the last year. PSC patients had similar alpha diversity, irrespective of antibiotics use the last year. (B) Male and female PSC patients show similar alpha diversity (here illustrated by Shannon diversity index), (C) and beta diversity (unweighted UniFrac). (D) The relative abundance of the *Veillonella* genus was higher in PSC patients who underwent liver transplantation after study inclusion compared to other PSC patients (P<0.05). PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls; AB, antibiotics; Chao1, Chao1 bacterial richness estimate; PSC-LTX, PSC patients with liver transplantation after inclusion in the study. Panels A&B show IQR + min, max, panel D shows median + IQR. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Supplementary figure S3



Beta diversity (unweighted UniFrac): (A) Shift in global bacterial composition between PSC and HC in both the exploration panel (unweighted UniFrac, PERMANOVA: pseudo-F-statistic: 5.8, P<0.001) and (B) the validation panel (unweighted UniFrac, PERMANOVA: pseudo-F-statistic: 7.3, P<0.001). (C) Only looking at PSC patients, in the combined panel, no difference could be observed between PSC patients according to their IBD status. (D) All over medication use did also not show any particular clustering among the PSC patients. PSC, primary sclerosing cholangitis; HC, healthy controls; IBD, inflammatory bowel disease; PC, principal component.

Supplementary figure S4



Gut microbiota-based PSC classification using AUROC analysis: ROC-curves calculated by using coefficients from multiple logistic regression from the nine genera differing between HC and PSC, and that were confirmed by all validation methods (linear regression, LEfSe and MaAsLin), together with transformed relative abundances for each bacterial taxa included, (A) in the exploration panel (AUC=0.78, 95% CI: 0.70-0.87, P<0.0001) and for the validation panel, but using coefficients from the exploration panel (AUC=0.78, 95% CI: 0.72-0.87, P<0.0001). AUC values were similar, indicating that over fitting has been avoided in the analyses. (B) Histogram of AUC values from the validation panel when using the same strategy as above and performing 1000 permutations, randomly assigning samples to explorations or validation panel. PSC, primary sclerosing cholangitis; HC, healthy controls; UC, ulcerative colitis; AUROC, area under the receiver operator characteristics curve; AUC, area under the curve; ROC, receiver operator characteristics.

SUPPLEMENTARY TABLES

The gut microbial profile in patients with primary sclerosing cholangitis is distinct from ulcerative colitis patients without biliary disease and healthy controls.

Martin Kummen[1,2,3,4], Kristian Holm[1,4], Jarl Andreas Anmarkrud[1,4], Ståle Nygård[5,6,7], Mette Vesterhus[1,8], Marte L. Høivik[9], Marius Trøseid[1,2,10], Hanns-Ulrich Marschall[11], Erik Schrumpf[1,3], Bjørn Moum[3,9], Helge Røsjø[7,12], Pål Aukrust[2,3,4,10], Tom H. Karlsen[1,2,3,4,13,14], Johannes R. Hov[1,2,3,4,14].

[1]Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; [2]K.G. Jebsen Inflammation Research Centre, University of Oslo, Oslo, Norway; [3]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [4]Research Institute of Internal Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; [5]Bioinformatics Core Facility, Institute for Medical Informatics, Oslo University Hospital, Oslo, Norway; [6]Institute for Experimental Medical Research, Oslo University Hospital and University of Oslo, Oslo, Norway; [7]K.G. Jebsen Cardiac Research Centre and Center for Heart Failure Research, University of Oslo, Oslo, Norway; [8]Department of Medicine, National Centre for Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway; [9]Department of Gastroenterology, Oslo University Hospital Ullevål, Oslo, Norway; [10]Section of Clinical Immunology and Infectious Diseases, Oslo University Hospital Rikshospitalet, Oslo, Norway; [11]Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [12]Division of Medicine, Akershus University Hospital, Lørenskog, Norway; [13]Department of Clinical Medicine, University of Bergen, Bergen, Norway; [14]Section of Gastroenterology, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway.

Table of Contents

Supplementary table S1

Overview of excluded samples by phenotype and total.

	PSC		UC		HC		Total	
	n	%	n	%	n	%	n	%
Exclusion criteria	**144**	**100**	**51**	**100**	**348**	**100**	**543**	**100**
Missing sample information	2	1.4			16	4.6	18	3.3
Sampling error	1	0.7	2	3.9	7	2.0	10	1.8
Bowel resection / stoma	30	20.8	1	2.0			31	5.7
Diet	7	4.8	4	7.8	12	3.4	23	4.2
Antibiotics last 4 weeks	8	5.4	3	5.9	5	1.4	16	2.9
Room temperature > 72 hours	3	2.0	2	3.9	20	5.7	25	4.6
Excluded before sequencing	**51**	**35.1**	**12**	**23.5**	**60**	**17.2**	**123**	**22.7**
Excluded after sequencing (failed sequencing/low read count)	8	5.4	3	5.9	25	7.2	36	6.6
Applicable for analysis	**85**	**59.4**	**36**	**70.6**	**263**	**75.6**	**384**	**70.7**

PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls.

Supplementary table S2

Demographics for the combined panel.

	PSC (n=85)		Healthy controls (n=263)		Ulcerative colitis (n=36)		P-value (PSC vs HC)	P-value (PSC vs UC)	P-value (UC vs HC)
Age, median years (min-max)	49.0	(21-82)	46.0	(30-61)	40.0	(22-69)	<0.05	<0.01	<0.05
Sex, male (%)	53	(62.4)	108	(41.1)	16	(44.4)	<0.01	0.08	0.72
BMI, median kg/m2 (min-max)	25.4	(18-38)	25.6	(18-43)	24.4	(18-34)	0.26	0.15	<0.05
Smoking, yes (%)	2	(2.4)	30	(11.4)	0		<0.01		
Sample time in RT, median hours (min-max)	27.0	(13-63)	33.0	(16-72)	26.0	(19-71)	0.21	0.76	0.22
Courses of AB < 12 months, median (min-max)	0	(0-10)	0	(0-5)	0	(0-4)	0.47	0.16	0.24
Ulcerative colitis, n (%)	44	(51.8)			36	(100.0)		<0.01	
Crohn's disease, n (%)	11	(12.9)							
Small duct PSC	3	(3.5)							
Mediacation, yes, n (%)									
PPI	6	(7.1)	18	(6.8)	2	(5.6)	1.00	1.00	1.00
Antihistamines	6	(7.1)	14	(5.3)			0.59		
Statins	7	(8.2)	11	(4.2)			0.16		
Ursodeoxycholic acid	25	(29.4)							
Prednisolon	13	(15.3)			5	(13.9)		1.00	
5-ASA	35	(41.2)			29	(80.6)		<0.01	
Infliximab	1	(1.2)			14	(38.9)		<0.01	
Adalimumab					4	(11.1)			
Azathioprine	12	(14.1)			10	(27.8)		0.12	
Budesonide	3	(3.5)			2	(5.6)		0.63	
PSC specific variables								Available for	
PSC Disease duration, median years (min-max)	9.1	(1-32)						N=85	
IBD duration, median years (min-max)	11.8	(0-45)						N=63	
Other autoimmune disease, yes (%)	23	(27.1)						N=80	
Mayo risk score, median (min-max)	-0.06	(-1.9-3.3)						N=64	
P-ANCA status, positive (%)	44	(51.8)						N=77	
Platelet count, 109/L, median (min-max)	223	(57-578)						N=77	
Creatinine, µmol/L, median (min-max)	68	(42-106)						N=78	
Bilirubin, µmol/L, median (min-max)	13	(3-114)						N=77	
Albumin, g/L, median (min-max)	43	(16-47)						N=72	
INR, median (min-max)	1.0	(0.9-1.3)						N=68	
AST, U/L, median (min-max)	45	(18-197)						N=72	
ALT, U/L, median (min-max)	56	(14-331)						N=78	
ALP, U/L, median (min-max)	170	(30-598)						N=77	
GGT, U/L, median (min-max)	184	(10-1576)						N=76	

PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls; BMI, body mass index; RT, room temperature; AB, antibiotics; PPI, proton pump inhibitors; 5-ASA, 5-Aminosalicylic acid; IBD, inflammatory bowel disease; INR, international normalised ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma-glutaryltransferase.

Supplementary table S3

Linear regression analyses of alpha diversity (Shannon diversity index, Chao1 bacterial richness estimate [Chao1] and Phylogenetic diversity) in the combined panel, for (A) PSC versus HC, (B) PSC versus UC and (C) PSC with and without IBD. For the multivariate analyses sex, smoking, age and BMI were considered obligate as covariates and all other variables that showed an effect in the univariate analyses with a P-value<0.10 were also included as covariates. In the PSC-subgroup analyses use of antibiotics, duration of PSC and duration of IBD were also considered obligate as covariates. Bilirubin, ALP, AST, ALT, and GGT were transformed by the natural logarithm prior to regression analyses due to a right-skewed distribution. Significant P-values (<0.05) in bold.

PSC, primary sclerosing cholangitis; HC, healthy controls; UC, ulcerative colitis; BMI, body mass index; AB, antibiotics; RT, room temperature; IBD, inflammatory bowel disease; PPI, proton pump inhibitors; PSC+IBD, PSC with inflammatory bowel disease; 5-ASA, 5-Aminosalicylic acid; INR, international normalised ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma-glutaryltransferase; P-ANCA, perinuclear anti-neutrophil cytoplasmic antibodies.

A)

Shannon diversity index - PSC vs HC

	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC	-0.654	(-0.8, -0.5)	**0.000**	-0.67	(-0.8, -0.5)	**0.000**
Sex (male)	0.129	(-0.01, 0.27)	0.076	0.04	(-0.09, 0.18)	0.505
Smoking (yes)	0.061	(-0.19, 0.31)	0.626	-0.08	(-0.3, 0.1)	0.500
Age	0.005	(-0.005, 0.01)	0.279	0.01	(0.001, 0.02)	**0.028**
BMI	-0.011	(-0.03, 0.01)	0.231	-0.01	(-0.03, 0.002)	0.087
Courses of AB < 12 months	-0.096	(-0.17, -0.02)	**0.017**	-0.06	(-0.13, 0.02)	0.121
Sample time in RT	0.004	(-0.002, 0.01)	0.239			
Statins (yes)	-0.228	(-0.55, 0.09)	0.163			
PPI (yes)	-0.003	(-0.3, 0.3)	0.986			
Antihistamines (yes)	-0.054	(-0.4, 0.3)	0.728			

Phylogenetic diversity - PSC vs HC

	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC	-4.215	(-5.2, -3.2)	**0.000**	-4.28	(-5.3, -3.2)	**0.000**
Sex (male)	0.763	(-0.2, 1.7)	0.121	0.28	(-0.6, 1.2)	0.539
Smoking (yes)	0.381	(-1.3, 2.1)	0.654	-0.50	(-2.0, 1.0)	0.523
Age	0.013	(-0.04, 0.07)	0.642	0.03	(-0.02, 0.09)	0.190
BMI	-0.110	(-0.2, 0.01)	0.074	-0.12	(-0.2, -0.01)	**0.037**
Courses of AB < 12 months	-0.914	(-1.4, -0.4)	**0.001**	-0.67	(-1.2, -0.2)	**0.008**
Sample time in RT	0.034	(-0.01, 0.07)	0.104			
Statins (yes)	-0.722	(-2.9, 1.5)	0.516			
PPI (yes)	-0.426	(-2.3, 1.5)	0.661			
Antihistamines (yes)	-0.175	(-2.3, 1.9)	0.868			

Chao1 - PSC vs HC

	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC	-90.071	(-110.4, -69.7)	**0.000**	-91.43	(-112.4, -70.4)	**0.000**
Sex (male)	12.997	(-6.3, 32.3)	0.187	1.62	(-16.2, 19.4)	0.858
Smoking (yes)	8.575	(-24.9, 42.0)	0.614	-9.62	(-40.0, 20.8)	0.534
Age	0.115	(-1.0, 1.2)	0.840	0.56	(-0.5, 1.6)	0.280
BMI	-1.491	(-3.9, 0.9)	0.224	-1.64	(-3.9, 0.6)	0.148
Courses of AB < 12 months	-18.089	(-28.7, -7.5)	**0.001**	-13.76	(-23.6, -3.9)	**0.007**
Sample time in RT	0.580	(-0.2, 1.4)	0.161			
Statins (yes)	-9.694	(-53.3, 33.9)	0.662			
PPI (yes)	-13.591	(-51.7, 24.5)	0.483			
Antihistamines (yes)	-1.754	(-43.3, 39.8)	0.934			

B)

Shannon diversity index - PSC vs UC

	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC	-0.034	(-0.3, 0.2)	0.806	0.02	(-0.3, 0.3)	0.911
Sex (male)	0.179	(-0.1, 0.4)	0.156	0.16	(-0.1, 0.4)	0.189
Smoking (yes)	-0.230	(-1.2, 0.7)	0.641	-0.03	(-1.0, 0.9)	0.944
Age	0.003	(-0.01, 0.01)	0.512	0.005	(-0.01, 0.01)	0.301
BMI	-0.053	(-0.09, 0.02)	**0.001**	-0.05	(-0.09, -0.02)	**0.001**
Courses of AB < 12 months	-0.065	(-0.16, 0.03)	0.198			
Sample time in RT	0.004	(-0.01, 0.01)	0.448			
Statins (yes)	0.082	(-0.5, 0.6)	0.761			
PPI (yes)	-0.159	(-0.7, 0.3)	0.530			
Antihistamines (yes)	-0.373	(-0.9, 0.2)	0.196			

Phylogenetic diversity - PSC vs UC

	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC	1.335	(-0.4, 3.1)	0.135	1.53	(-0.3, 3.3)	0.097
Sex (male)	1.164	(-0.5, 2.8)	0.158	1.42	(-0.2, 3.1)	0.088
Smoking (yes)	-2.657	(-9.0, 3.7)	0.408	-0.39	(-7.0, 6.2)	0.906
Age	0.045	(-0.02, 0.11)	0.144	0.04	(-0.03, 0.1)	0.265
BMI	-0.189	(-0.40, 0.03)	0.083	-0.19	(-0.41, 0.04)	0.099
Courses of AB < 12 months	-0.602	(-1.2, 0.04)	0.065	-0.49	(-1.2, 0.2)	0.167
Sample time in RT	0.040	(-0.03, 0.11)	0.248			
Statins (yes)	1.801	(-1.7, 5.3)	0.304			
PPI (yes)	-1.722	(-5.0, 1.5)	0.296			
Antihistamines (yes)	-0.935	(-4.7, 2.8)	0.621			

Chao1 - PSC vs UC

	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC	27.804	(-7.43, 63.0)	0.121	32.03	(-4.6, 68.6)	0.086
Sex (male)	20.624	(-12.0, 53.3)	0.214	26.77	(-6.2, 59.7)	0.110
Smoking (yes)	-50.666	(-178.0, 76.6)	0.432	-3.32	(-136.8, 130.1)	0.961
Age	0.745	(-0.48, 1.97)	0.231	0.50	(-0.8, 1.8)	0.436
BMI	-3.189	(-7.5, 1.1)	0.147	-2.89	(-7.3, 1.6)	0.201
Courses of AB < 12 months	-13.139	(-26.0, -0.30)	**0.045**	-11.61	(-25.6, 2.4)	0.103
Sample time in RT	0.732	(-0.6, 2.1)	0.288			
Statins (yes)	39.156	(-30.2, 108.5)	0.266			
PPI (yes)	-48.520	(-113.4, 16.4)	0.141			
Antihistamines (yes)	-24.561	(-99.4, 50.3)	0.517			

C)

Shannon diversity index - PSC only - PSC+IBD vs PSC without IBD						
	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC+IBD	0.097	(-0.23, 0.42)	0.553	0.056	(-0.5, 0.6)	0.850
Sex (male)	0.029	(-0.3, 0.3)	0.858	-0.003	(-0.5, 0.5)	0.992
Smoking (yes)	-0.221	(-1.2, 0.8)	0.668	0.025	(-1.2, 1.3)	0.968
Age	0.006	(-0.01, 0.02)	0.311	0.007	(-0.02, 0.03)	0.527
BMI	-0.046	(-0.09, 0.01)	**0.020**	-0.025	(-0.09, 0.04)	0.421
Courses of AB < 12 months	-0.057	(-0.2, 0.1)	0.303	-0.035	(-0.2, 0.1)	0.640
Sample time in RT	0.007	(-0.01, 0.02)	0.287			
Statins (yes)	0.095	(-0.5, 0.7)	0.738			
PPI (yes)	0.061	(-0.5, 0.7)	0.841			
Antihistamines (yes)	-0.371	(-1.0, 0.2)	0.222			
Ursodeoxycholic acid (yes)	-0.182	(-0.5, 0.2)	0.286			
Cholestyramine (yes)	-0.119	(-0.8, 0.6)	0.747			
Prednisolon (yes)	0.098	(-0.3, 0.5)	0.650			
5-ASA (yes)	-0.176	(-0.5, 0.1)	0.266			
Infliximab (yes)	-0.316	(-1.8, 1.1)	0.662			
Azathioprine (yes)	-0.083	(-0.5, 0.4)	0.712			
Budesonide (yes)	0.317	(-0.5, 1.2)	0.453			
Platelets	0.000	(-0.002, 0.002)	0.768			
Creatinine	0.001	(-0.01, 0.01)	0.851			
Total bilirubin	0.060	(-0.2, 0.3)	0.625			
Albumin	-0.017	(-0.05, 0.02)	0.316			
INR	2.082	(0.3, 3.9)	**0.025**	2.047	(-0.4, 4.5)	0.095
AST	0.010	(-0.3, 0.3)	0.946			
ALT	0.005	(-0.2, 0.2)	0.959			
ALP	-0.111	(-0.3, 0.1)	0.336			
GGT	0.001	(-0.1, 0.1)	0.983			
P-ANCA (positive)	-0.034	(-0.4, 0.3)	0.835			
Other autimmune disease (yes)	0.228	(-0.1, 0.6)	0.202			
MAYO risk score	0.041	(-0.1, 0.2)	0.664			
Time from PSC diagnosis	0.000	(-0.002, 0.001)	0.631	-0.001	(-0.003, 0.002)	0.666
Time from IBD diagnosis	-0.001	(-0.002, 0.001)	0.296	-0.001	(-0.003, 0.001)	0.175

	Phylogenetic diversity - PSC only - PSC without IBD					
	Unadjusted			Adjusted		
	Beta	(95% CI)	P-value	Beta	(95% CI)	P-value
PSC+IBD	0.026	(-2.0, 2.1)	0.980	0.02	(-3.7, 3.8)	0.992
Sex (male)	1.051	(-1.0, 3.1)	0.305	1.07	(-2.3, 4.4)	0.520
Smoking (yes)	-3.082	(-9.6, 3.4)	0.347	-1.68	(-9.5, 6.2)	0.668
Age	0.048	(-0.03, 0.13)	0.236	0.02	(-0.1, 0.2)	0.792
BMI	-0.227	(-0.48, 0.03)	0.076	-0.21	(-0.6, 0.2)	0.297
Courses of AB < 12 months	-0.500	(-1.2, 0.2)	0.158	-0.14	(-1.1, 0.8)	0.776
Sample time in RT	0.059	(-0.03, 0.15)	0.188			
Statins (yes)	1.417	(-2.2, 5.0)	0.434			
PPI (yes)	-0.192	(-4.1, 3.7)	0.921			
Antihistamines (yes)	-1.383	(-5.2, 2.5)	0.477			
Ursodeoxycholic acid (yes)	-0.853	(-3.0, 1.3)	0.435			
Cholestyramine (yes)	-0.298	(-5.0, 4.4)	0.899			
Prednisolon (yes)	0.624	(2.1, 3.4)	0.652			
5-ASA (yes)	-1.427	(-3.4, 0.6)	0.156			
Infliximab (yes)	-5.214	(-14.3, 3.9)	0.257			
Azathioprine (yes)	-1.415	(-4.2, 1.4)	0.321			
Budesonide (yes)	3.562	(-1.7, 8.9)	0.185			
Platelets	-0.004	(-0.02, 0.01)	0.514			
Creatinine	0.000	(-0.08, 0.08)	0.996			
Total bilirubin	0.106	(-1.5, 1.7)	0.894			
Albumin	-0.104	(-0.3, 0.1)	0.345			
INR	11.169	(-0.3, 22.7)	0.057	10.28	(-5.1, 25.7)	0.184
AST	0.331	(-1.4, 2.1)	0.709			
ALT	0.445	(-0.01, 0.02)	0.485			
ALP	-0.090	(-1.6, 1.4)	0.903			
GGT	0.241	(-0.6, 1.1)	0.584			
P-ANCA (positive)	-0.060	(-2.2, 2.0)	0.955			
Other autimmune disease (yes)	1.291	(-1.0, 3.5)	0.255			
MAYO risk score	0.039	(-1.1, 1.2)	0.946			
Time from PSC diagnosis	0.002	(-0.01, 0.01)	0.725	0.01	(-0.01, 0.02)	0.449
Time from IBD diagnosis	-0.004	(-0.01, 0.004)	0.282	-0.01	(-0.02, 0.003)	0.161

	Chao1 - PSC only - PSC+IBD vs PSC without IBD					
	Unadjusted			Adjusted		
	Beta	**(95% CI)**	**P-value**	**Beta**	**(95% CI)**	**P-value**
PSC+IBD	-9.064	(-50.7, 32.5)	0.666	-11.97	(-93.6, 69.7)	0.769
Sex (male)	22.926	(-17.8, 63.7)	0.267	16.86	(-52.8, 86.6)	0.628
Smoking (yes)	-59.501	(-190.2, 71.2)	0.368	-30.27	(-187.6, 127.1)	0.699
Age	0.775	(-0.8, 2.4)	0.343	0.28	(-2.7, 3.2)	0.847
BMI	-3.869	(-9.0, 1.2)	0.135	-2.37	(-11.3, 6.5)	0.592
Courses of AB < 12 months	-10.912	(-25.0, 3.1)	0.126	-6.07	(-26.1, 14.0)	0.544
Sample time in RT	1.022	(-0.7, 2.8)	0.254			
Statins (yes)	31.186	(-40.9, 103.3)	0.392			
PPI (yes)	-20.183	(-97.8, 57.4)	0.606			
Antihistamines (yes)	-34.017	(-111.4, 43.3)	0.384			
Ursodeoxycholic acid (yes)	-5.382	(-49.0, 38.3)	0.807			
Cholestyramine (yes)	0.565	(-93.4, 94.6)	0.990			
Prednisolon (yes)	8.309	(-47.0, 63.6)	0.766			
5-ASA (yes)	-33.938	(-73.7, 5.8)	0.093	-21.40	(-92.1, 49.3)	0.544
Infliximab (yes)	-76.213	(-260.0, 107.6)	0.412			
Azathioprine (yes)	-30.016	(-86.8, 26.8)	0.296			
Budesonide (yes)	62.775	(-44.2, 169.8)	0.247			
Platelets	-0.069	(-0.3, 0.2)	0.563			
Creatinine	-0.493	(-2.0, 1.0)	0.523			
Total bilirubin	7.121	(-24.8, 39.1)	0.658			
Albumin	-2.040	(-6.5, 2.5)	0.369			
INR	197.338	(-38.5, 433.2)	0.100	238.94	(-67.1, 544.9)	0.122
AST	6.429	(-29.4, 42.3)	0.722			
ALT	8.113	(-17.6, 33.8)	0.532			
ALP	-0.753	(-30.7, 29.2)	0.960			
GGT	4.430	(-13.3, 22.2)	0.620			
P-ANCA (positive)	2.858	(-40.2, 45.9)	0.895			
Other autimmune disease (yes)	23.305	(-22.1, 68.7)	0.310			
MAYO risk score	1.890	(-22.0, 25.8)	0.875			
Time from PSC diagnosis	0.088	(-0.1, 0.3)	0.433	0.18	(-0.1, 0.5)	0.275
Time from IBD diagnosis	-0.129	(-0.3, 0.03)	0.108	-0.19	(-0.4, 0.04)	0.099

Supplementary table S4

The final 12 genera confirmed in the validation panel that differed between PSC and HC. Results from Mann-Whitney, linear regression, LEfSe and MaAsLin, together with the median relative abundances of each bacterial taxa in each phenotype. In the linear regression analyses age, sex, smoking status, BMI and number of prescriptions for antibiotics the last 12 months before inclusion were used as covariates. MaAsLin and linear regression analyses did not reveal any significant difference between PSC and UC.

PSC, primary sclerosing cholangitis; UC, ulcerative colitis; HC, healthy controls; LDA, linear discriminant analysis score; DC, discriminatory class; ns., nonsignificant.

			Exploration panel - Mann-Whitney (PSC vs HC)			Validation panel - Mann-Whitney (PSC vs HC)		
Order	Family	*Genus*	Increased in	P-value	Q-FDR	Increased in	P-value	Q-FDR
Bacteroidales	S24.7	*[unknown]*	HC	0.010	0.081	HC	0.002	0.013
YS2	[unknown]	*[unknown]*	HC	0.003	0.043	HC	0.008	0.033
Clostridiales	[unknown]	*[unknown]*	HC	0.002	0.043	HC	0.002	0.013
Clostridiales	Christensenellaceae	*[unknown]*	HC	0.002	0.043	HC	0.001	0.010
Clostridiales	Lachnospiraceae	*[unknown]*	HC	0.019	0.105	HC	0.015	0.062
Clostridiales	Lachnospiraceae	*Coprococcus*	HC	0.000	0.015	HC	0.001	0.010
Clostridiales	Veillonellaceae	*Phascolarctobacterium*	HC	0.013	0.102	HC	0.034	0.116
Clostridiales	Veillonellaceae	*Veillonella*	PSC	0.029	0.140	PSC	0.001	0.010
RF32	[unknown]	*[unknown]*	HC	0.017	0.102	HC	0.001	0.013
Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*	HC	0.018	0.102	HC	0.018	0.070
Aeromonadales	Succinivibrionaceae	*Succinivibrio*	HC	0.016	0.102	HC	0.010	0.041
ML615J-28	[unknown]	*[unknown]*	HC	0.033	0.140	HC	0.000	0.004

*Covariates: age, sex, smoking status, BMI and number of prescriptions for antibiotics the last 12 months before inclusion.

LEfSe reference: Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;12:R60. doi:10.1186/gb-2011-12-6-r60 [published Online First: 24 June

MaAsLin reference: Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012;13:R79. doi:10.1186/gb-2012-13-9-r79 [published Online First: 26 September 2012].

			Combined panel															
			Mann-Whitney (PSC vs UC)			Linear regression (PSC vs HC)*			LEfSe (PSC vs HC)			LEfSe (PSC vs UC)			MaAsLin (PSC vs HC)*		
Order	Family	*Genus*	Increased in	P-value	Q-FDR	P-value	Q-FDR	Beta	DC	LDA	P-value	DC	LDA	P-value	P-value	q-value	Coefficient
Bacteroidales	S24.7	*[unknown]*		ns.		0.005	0.020	-0.024	HC	2.74	0.000			ns.	0.004	0.025	-0.025
YS2	[unknown]	*[unknown]*		ns.		0.000	0.003	-0.016	HC	2.83	0.000			ns.	0.000	0.001	-0.014
Clostridiales	[unknown]	*[unknown]*		ns.		0.000	0.000	-0.058	HC	3.37	0.000			ns.	0.000	0.000	-0.065
Clostridiales	Christensenellaceae	*[unknown]*		ns.		0.001	0.006	-0.017	HC	2.35	0.000			ns.	0.001	0.008	-0.014
Clostridiales	Lachnospiraceae	*[unknown]*		ns.		0.001	0.005	-0.032	HC	3.19	0.001			ns.	0.000	0.004	-0.034
Clostridiales	Lachnospiraceae	*Coprococcus*		ns.		0.000	0.000	-0.046	HC	3.01	0.000			ns.	0.000	0.000	-0.047
Clostridiales	Veillonellaceae	*Phascolarctobacterium*		ns.		0.005	0.018	-0.031	HC	2.80	0.001			ns.	0.002	0.018	-0.032
Clostridiales	Veillonellaceae	*Veillonella*	PSC	0.0107	0.176	0.000	0.004	0.017	PSC	2.50	0.000	PSC	2.63	0.0076	0.007	0.034	0.005
RF32	[unknown]	*[unknown]*		ns.		0.001	0.006	-0.027	HC	2.45	0.000			ns.	0.000	0.004	-0.026
Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*		ns.		0.007	0.026	-0.011			ns.			ns.	0.009	0.041	-0.009
Aeromonadales	Succinivibrionaceae	*Succinivibrio*		ns.		ns.			HC	2.46	0.000			ns.		ns.	
ML615J-28	[unknown]	*[unknown]*		ns.		0.003	0.014	-0.008			ns.			ns.	0.000	0.004	-0.006

			Median relative abundances						
			Exploration panel		Validation panel		Combined panel		
Order	Family	*Genus*	HC	PSC	HC	PSC	HC	UC	PSC
Bacteroidales	S24.7	*[unknown]*	0	0	0	0	0	0	0
YS2	[unknown]	*[unknown]*	0.00004474	0	0	0	0.00003841	0	0
Clostridiales	[unknown]	*[unknown]*	0.06138230	0.03457256	0.05893320	0.03366515	0.06025947	0.02122473	0.03457256
Clostridiales	Christensenellaceae	*[unknown]*	0.00082704	0.00022743	0.00100334	0.00015451	0.00087941	0.00019155	0.00018593
Clostridiales	Lachnospiraceae	*[unknown]*	0.08141825	0.05955697	0.08254610	0.06698648	0.08194081	0.06426477	0.06318431
Clostridiales	Lachnospiraceae	*Coprococcus*	0.01578138	0.00431092	0.00960404	0.00460299	0.01224219	0.00456231	0.00438825
Clostridiales	Veillonellaceae	*Phascolarctobacterium*	0.00025520	0	0.00040772	0.00004317	0.00028934	0.00003648	0.00001607
Clostridiales	Veillonellaceae	*Veillonella*	0.00010660	0.00031689	0.00019144	0.00050188	0.00014794	0.00015086	0.00040215
RF32	[unknown]	*[unknown]*	0.00138110	0.00005391	0.00205392	0.00001472	0.00174456	0	0.00002944
Desulfovibrionales	Desulfovibrionaceae	*Desulfovibrio*	0	0	0	0	0	0	0
Aeromonadales	Succinivibrionaceae	*Succinivibrio*	0	0	0	0	0	0	0
ML615J-28	[unknown]	*[unknown]*	0	0	0.00002102	0	0	0	0

Supplementary table S5

Alpha diversity. Comparison in the combined panel of alpha diversity (Shannon diversity index, Chao1 bacterial richness estimate [Chao1] and Phylogenetic diversity), in: (A) IBD-subphenotypes in PSC, (B) medication in PSC and (C) small duct PSC, large duct PSC and HC. Significant P-values (<0.05) in bold.

PSC, primary sclerosing cholangitis; IBD, inflammatory bowel disease, PSC+IBD, PSC with inflammatory bowel disease; UC, ulcerative colitis; CD, Crohn's disease; HC, healthy controls.

A)

	PSC		PSC (no IBD)				PSC+IBD					PSC-UC				
	n	%	n	%			n	%				n	%			
Of total:	85	22.1	30	8.2			55	14.9				44	11.5			
Within PSC:	85	100.0	30	35.6			55	64.7				44	51.8			
Within PSC+IBD:	-		-				55	100.0				44	80.0			
Alpha diversity measure	median	range	median	range	P-value vs HC	P-value vs UC	median	range	P-value vs PSC (no IBD)	P-value vs HC	P-value vs UC	median	range	P-value vs PSC (no IBD)	P-value vs HC	P-value vs UC
Phylogenetic diversity	20.8	22.1	20.2	15.1	**4.59E-08**	0.23	21.1	22.1	0.79	**1.09E-08**	0.16	20.0	20.0	0.80	**3.79E-08**	0.38
Shannon diversity index	5.5	3.0	5.5	2.5	**3.73E-11**	0.44	5.5	3.5	0.37	**1.10E-08**	0.78	5.5	3.3	0.46	**6.52E-08**	0.90
Chao1	366.2	472.7	357.4	338.3	**8.24E-08**	0.21	379.5	472.7	0.94	**3.04E-11**	0.14	356.4	400.4	0.57	**1.24E-10**	0.42

	PSC-CD						UC			HC	
	n	%					n	%		n	%
Of total:	11	2.9					36	9.4		263	68.5
Within PSC:	11	12.9					-			-	
Within PSC+IBD:	11	20.0					-			-	
Alpha diversity measure	median	range	P-value vs PSC (no IBD)	P-value vs HC	P-value vs UC	P-value vs PSC-UC	median	range	P-value vs HC	median	range
Phylogenetic diversity	21.9	21.3	0.12	**0.03**	**0.04**	0.33	19.9	18.3	**1.10E-11**	25.2	20.4
Shannon diversity index	5.7	2.7	0.38	**0.02**	0.61	0.07	5.6	2.5	**2.92E-08**	6.2	3.0
Chao1	401.4	425.4	0.20	**0.02**	**0.01**	0.15	361.7	370.6	**5.58E-14**	466.3	478.8

B)

Alpha diversity measure	5-Aminosalicylic acid: Yes n	%	5-Aminosalicylic acid: No n	%	P-value	Ursodeoxycholic acid: Yes n	%	Ursodeoxycholic acid: No n	%	P-value	Prednisolon: Yes n	%	Prednisolon: No n	%	P-value
	35	41.2	50	58.8		25	29.4	60	70.6		13	15.3	72	84.7	
	median	range	median	range		median	range	median	range		median	range	median	range	
Phylogenetic diversity	19.3	19.7	21.1	22.1	0.16	19.9	15.1	21.2	22.1	0.50	19.8	18.2	20.9	21.5	0.88
Shannon diversity index	5.4	3.5	5.6	2.9	0.34	5.4	2.7	5.6	3.5	0.32	5.5	3.3	5.5	3.4	0.81
Chao1	349.0	359.9	383.6	472.7	0.09	359.8	323.4	381.7	472.7	0.84	353.9	367.5	372.8	472.7	0.96

Alpha diversity measure	Azathioprine: Yes n	%	Azathioprine: No n	%	P-value	Proton pump inhibitors: Yes n	%	Proton pump inhibitors: No n	%	P-value	Statins: Yes n	%	Statins: No n	%	P-value
	12	14.1	73	85.9		6	7.1	79	92.9		7	8.2	78	91.8	
	median	range	median	range		median	range	median	range		median	range	median	range	
Phylogenetic diversity	19.6	22.1	21.0	19.3	0.42	22.3	10.8	20.7	22.1	0.89	22.2	12.8	20.7	22.1	0.47
Shannon diversity index	5.5	3.3	5.5	3.4	0.80	5.7	1.3	5.5	3.5	0.90	5.7	1.7	5.5	3.5	0.68
Chao1	366.3	439.7	366.2	399.2	0.55	382.1	237.5	366.2	472.7	0.85	413.0	233.8	364.7	472.7	0.40

Alpha diversity measure	Antihistamins: Yes n	%	Antihistamins: No n	%	P-value	Cholestyramine: Yes n	%	Cholestyramine: No n	%	P-value	Medication (any): Yes n	%	Medication (any): No n	%	P-value
	6	7.1	79	92.9		4	4.7	81	95.3		12	14.1	73	85.9	
	median	range	median	range		median	range	median	range		median	range	median	range	
Phylogenetic diversity	18.9	4.7	21.1	22.1	0.35	20.8	9.1	20.8	22.1	0.87	20.7	22.1	21.6	15.2	0.43
Shannon diversity index	5.3	2.0	5.6	3.5	0.23	5.4	1.5	5.5	3.5	0.76	5.5	3.5	5.7	2.3	0.31
Chao2	331.0	102.3	379.5	472.7	0.22	395.1	141.1	363.2	472.7	0.98	366.2	465.3	370.7	345.7	0.61

C)

	N	**Phylogenetic diversity** Median	P-value*	**Chao1** Median	P-value*	**Shannon diversity Index** Median	P-value*
Healthy controls	263	25.15	0.910	466.28	0.673	6.22	0.421
Large duct PSC	82	20.09	0.121	355.06	0.139	5.51	0.464
Small duct PSC	3	22.68		401.05		5.56	

*P-value versus small duct PSC