SUPPLEMENTARY INFORMATION

Supplementary methods

DNA isolation and sequencing

DNA isolation of fecal samples was performed in batches of 11 or 23 by repeated bead beating in combination with the PSP spin stool kit (Stratec Molecular, Berlin, Germany) as described previously (1). For each DNA isolation batch, one additional isolation was performed on PCR-grade water as a negative control.

Amplicon library preparation and sequencing was performed according to a previous published protocol (2). The 515f/806r primer pair was used to amplify the V4 region of the 16S rRNA gene. PCR reactions were performed using 25 µL NEB Phusion High-Fidelity Master Mix (New England Biolabs, Ipswich, USA), 4 µl 515f/806r primer mix and 30 ng metagenomics DNA under the following conditions: denaturation at 98^{o}C for 3 minutes, followed by 30 cycles of denaturation at 98^{o}C for 45 seconds, annealing at 55^{o}C for 45 seconds and extension at 72^{o}C for 45 seconds. The final elongation step was at 72^{o}C for 7 minutes. Amplicons were purified using the AMPure XP purification (Agencourt, Massachusetts, USA) according to the manufacturer's instructions. Amplicons were mixed in equimolar concentrations and sequenced on an Illumina MiSeq instrument.

Sequence Analysis

Quality control of the sequencing data were performed using FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) using default settings. Data demultiplexing, length and quality filtering and clustering of reads into Operational Taxonomic Units (OTUs) at 97% sequence identity was done using the online Integrated Microbial Next Generation Sequencing (IMNGS) platform (3) using default settings except for minimum and maximum length for amplicons, which were set at 100 and 500 bp respectively.

After quality filtering and binning and removing unassigned reads, a total of 2,829,437 sequences with an average of 19,649 paired sequences per sample (range: 11,744-26,641 sequences/sample) remained for downstream analysis and were clustered in 473 OTUs.

Data normalization, alpha indices, taxonomical binning and unsupervised clustering were performed using *Rhea* (4).

In order not to discard informative data, normalization in Rhea is performed by dividing OTU counts per sample for their total count (sample depth) and then multiplying the obtained relative abundance for the lowest sample depth (11744 reads/sample).

References

1. Tedjo DI, Jonkers DMAE, Savelkoul PH, et al. The Effect of Sampling and Storage on the Fecal Microbiota Composition in Healthy and Diseased Subjects. PLOS ONE 2015;10:e0126685.
2. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 2012;6:1621-1624.
3. Lagkouvardos I, Joseph D, Kapfhammer M, et al. IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci Rep 2016;6:33721.
4. Lagkouvardos I, Fischer S, Kumar N, et al. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. PeerJ 2017;5:e2836.

SUPPLEMENTARY TABLES

Supplementary Table 1 CD patients characteristics

	Remission-remission (n=35)	**Remission-active (n=22)**	**Healthy controls (n=15)**
Age at inclusion (median, IQR)	43 (33-53)	38 (26-59)	25 (23-30)
Male (%)	10 (28.6)	10 (45.5)	8 (53.3)
Smoking (%)	8 (22.9)	2 (9.1)	0 (0)
Age at diagnosis[1] (%)			
A1 (<16 year)	1 (2.9)	1 (4.5)	Na
A2 (17-40 year)	30 (85.7)	13 (59.1)	
A3 (>40 year)	4 (11.4)	8 (36.4)	
Disease localization[1] (%)			
L1 (ileal)	12 (34.3)	7 (31.8)	Na
L2 (colonic)	8 (22.9)	7 (31.8)	
L3 (ileocolonic)	15 (42.9)	8 (36.4)	
Phenotype at inclusion[1] (%)			
B1 (nonstricturing, non penetrating)	26 (74.3)	12 (54.5)	Na
B2 (stricturing)	6 (17.1)	5 (22.7)	

B3 (penetrating)	3 (8.6)	5 (22.7)	
Abdominal Surgery[2]	8 (22.9)	4 (18.2)	0 (0)

[1] According to Montreal Classification

[2] includes (hemi)colectomy and ileocecal resection

Supplementary Table 2 Medication use and time between sampling moments for active and remission samples

	RR		RA	
	Remission (n=35)	**Remission (n=35)**	**Remission (n=22)**	**Active (n=22)**
Medication (%)[‡]				
Mesalazine	5 (14.3)	5 (14.3)	4 (18.2)	5 (22.7)
Thiopurines	11 (31.4)	11 (31.4)	9 (40.9)	7 (31.8)
Biologicals	18 (51.4)	19 (54.3)	13 (59.1)	15 (68.2)
Corticosteroids	1 (2.9)	0 (0)	1 (4.5)	1 (4.5)
Proton Pump Inhibitors	7 (20)	7 (20)	8 (36.4)	8 (36.4)
Antibiotics[#]	1 (2.9)	0 (0)	1 (4.5.)	0 (0)
Time between sampling moments (week, median, IQR)		14 (11-21)[1]		20 (8-36)[2]

[‡]Six RR and five RA patients had a medication change between consecutive samples during the study period. In the RR group, mesalazine was stopped by 1 patients, prednisone by 1 patient, and biologicals in 2 patients, while 1 patient started mesalazine and 1 patient with started biologicals. In the RA group, 2 patients started with biologicals, 2 patients stopped with thiopurines and 1 patient started with mesalazine.

[1] Time between first remission and second remission samples

[2] Time between first remission and first active samples

[#] Ciprofloxacin and cotrimoxazol were used two and one month prior to sample collection, respectively.

Supplementary Table 3 z statistics and p-value resulting from Generalized Linear Model on clinical history 1 year before the study period and enterotype clusters

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.53063	0.398527	-1.33147	0.183033
E2	0.443617	0.577119	0.768675	0.442086
E3	0.81831	0.861485	0.949883	0.342172

Supplementary Table 4 z statistics and p-value resulting from Generalized Linear Model on clinical course during the5 years following the study period and enterotype clusters

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.223144	0.387298	0.576154	0.564511
E2	0.405465	0.584523	0.693669	0.48789
E3	0.064539	0.856349	0.075365	0.939925

Supplementary Table 5 F statistics and p-value resulting from PERMANOVA on microbiota composition using the clinical history 1 year before the study period as explanatory variable

	Df	SumOfSqs	R2	F	Pr(>F)
1 year before	2	0.393114439	0.036689	1.028328	0.399
Residual	54	10.32169678	0.963311	NA	NA
Total	56	10.71481121	1	NA	NA

Supplementary Table 6 F statistics and p-value resulting from PERMANOVA on microbiota composition using the clinical course during the 5 year following the study period as explanatory variable

	Df	SumOfSqs	R2	F	Pr(>F)
rec.5.Y.after	2	0.361603	0.033748	0.943019	0.5063
Residual	54	10.35321	0.966252	NA	NA
Total	56	10.71481	1	NA	NA

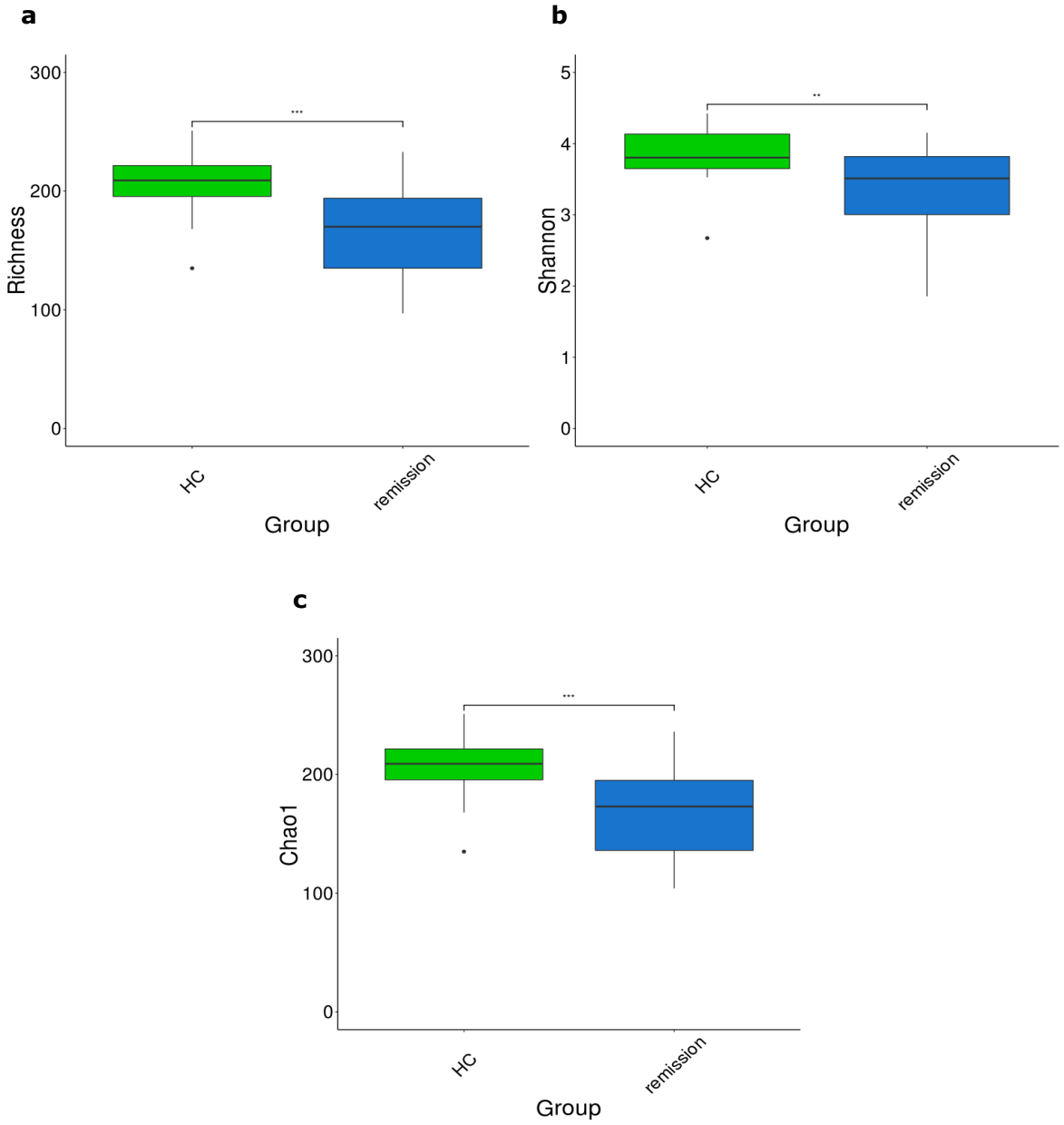
Supplementary Table 7 F statistics and p-value resulting from PERMANOVA on microbiota composition using the Montreal classification factors, medication usage, smoking,surgery, and sample depth.

	F	Pr(>F)
Disease localization	0.82734	0.655
Age at diagnosis	0.843010	0.558
Phenotype	2.084877	0.016
Thiopurines	0.408099	0.971
Mesalazines	0.655223	0.773
Biological	0.581659	0.828
Corticosteroids	1.044867	0.293
Proton Pump Inhibitors	0.781195	0.61
Surgery	1.554117	0.123
Smoking	0.879233	0.591
Number of liquid stools/day	1.1946	0.247
Sequencing depth	1.709101	0.089

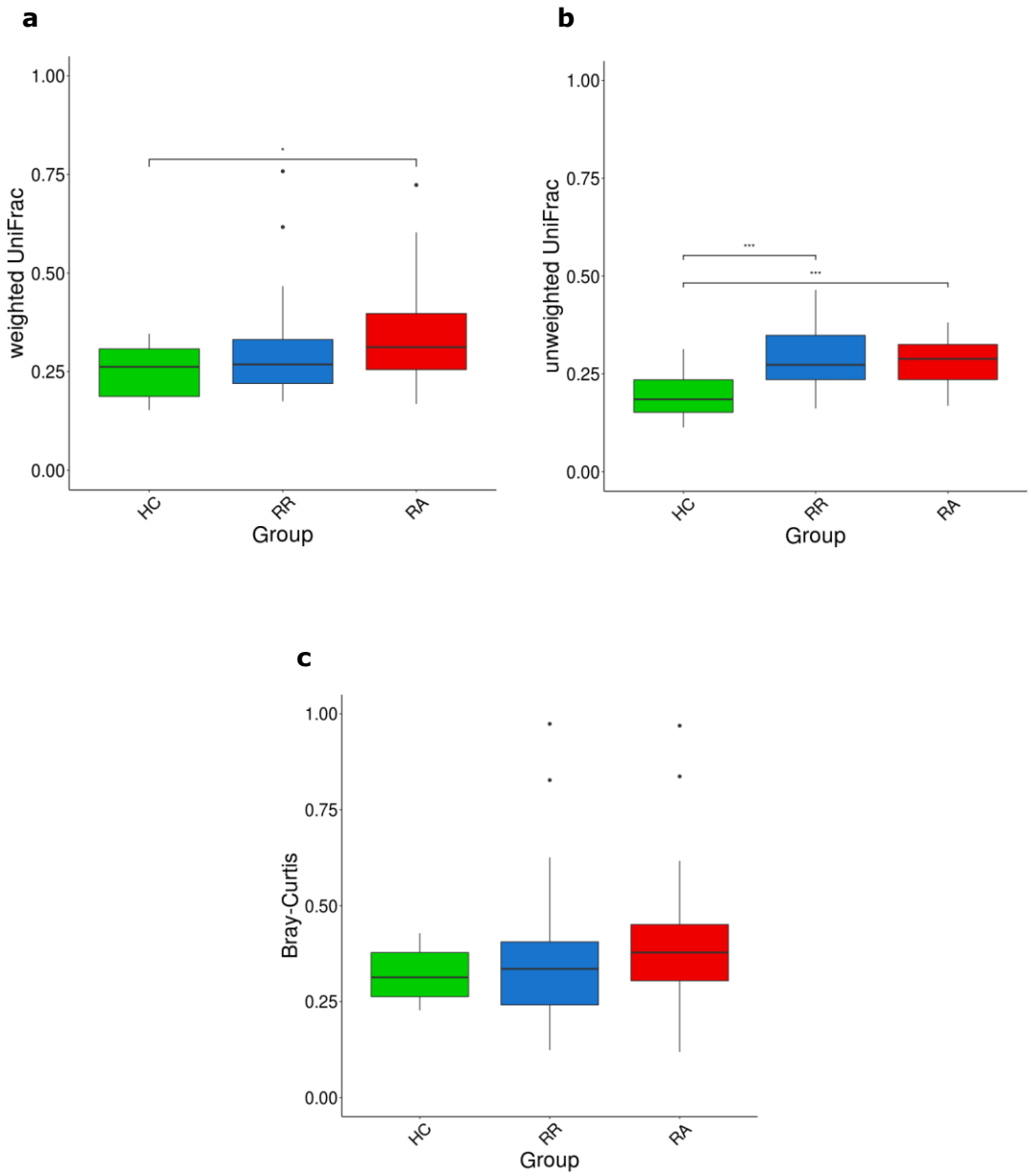
Supplementary Table 8 F statistics and p-value resulting from PERMANOVA on microbiota composition using the disease phenotype: B1(non stricturing, non penetrating); B2(structuring); B3(penetrating).

	F	Pr(>F)
B1 vs. B2	0.703237	0.691
B1 vs. B3	2.839477	0.013
B2 vs. B3	2.082749	0.039

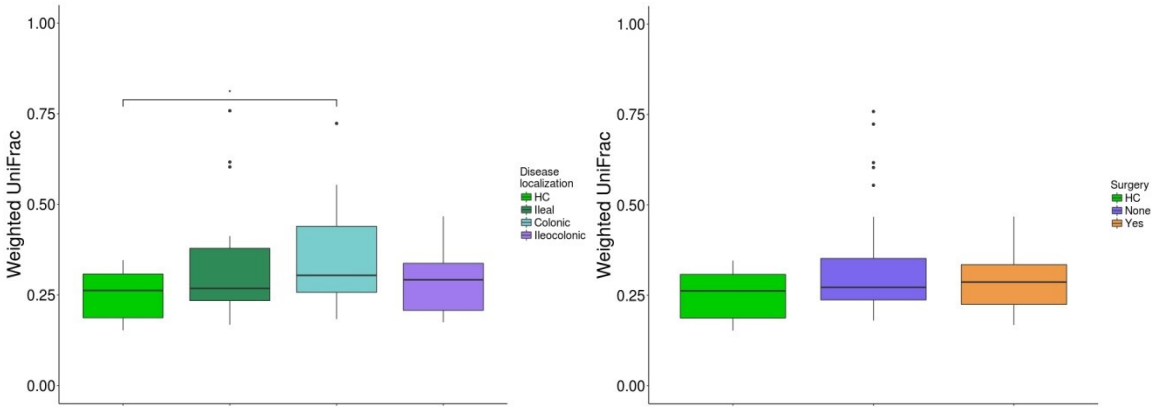
SUPPLEMENTARY FIGURES



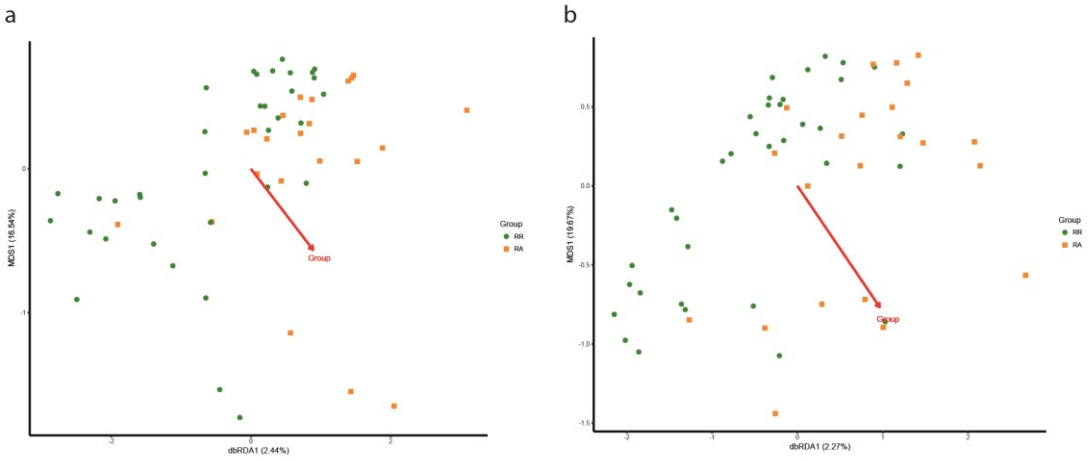
Supplementary Figure 1 Baseline alpha diversity indices: a)Observed species (richness), b)Chao1, and c) Shannon index within healthy controls (HC) and Crohn's disease patients at baseline. All patients were in remission state at baseline. Significance was tested using Wilcoxson Signed-Ranks Test; *** indicates $p<0.01$.



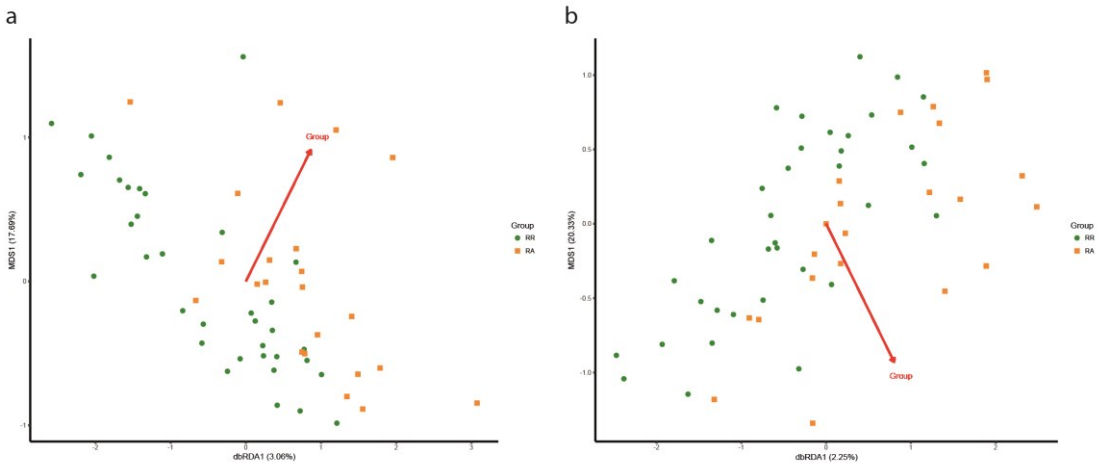
Supplementary Figure 2 Within-subject beta-diversity between two subsequent sampling time-points (T1-T2): a)Weighted UniFrac, b) Unweighted UniFrac and c) Bray-Curtis distance, in healthy individuals (HC), CD patients maintaining in remission (RR) and CD patients in remission and subsequent exacerbation (RA).



Supplementary Figure 3 Within-subject weighted UniFrac distance between two subsequent sampling time-points (T1-T2) among: a) CD patients with different disease localization and healthy subjects, and b) CD patients that did or did not received abdominal surgery .



Supplementary Figure 4 Distance-based redundancy analysis (dbRDA) based on weighted UniFrac distances using patient group as explanatory variable at **a**) baseline (T1) and **b**) at second time point (T2). Significance was tested using PERMANOVA resulting in p=0.12 and p=0.15 for T1 and T2 respectively.



Supplementary Figure 5 Distance-based redundancy analysis (dbRDA) based on weighted UniFrac distances using patient group as explanatory variable and partially filtering out Bacteroides:Prevotella ratio, age, gender, and medications use at **a)** baseline (T1) and **b)** at second time point (T2). Significance was tested using PERMANOVA resulting in p=0.12 and p=0.17 for T1 and T2 respectively.