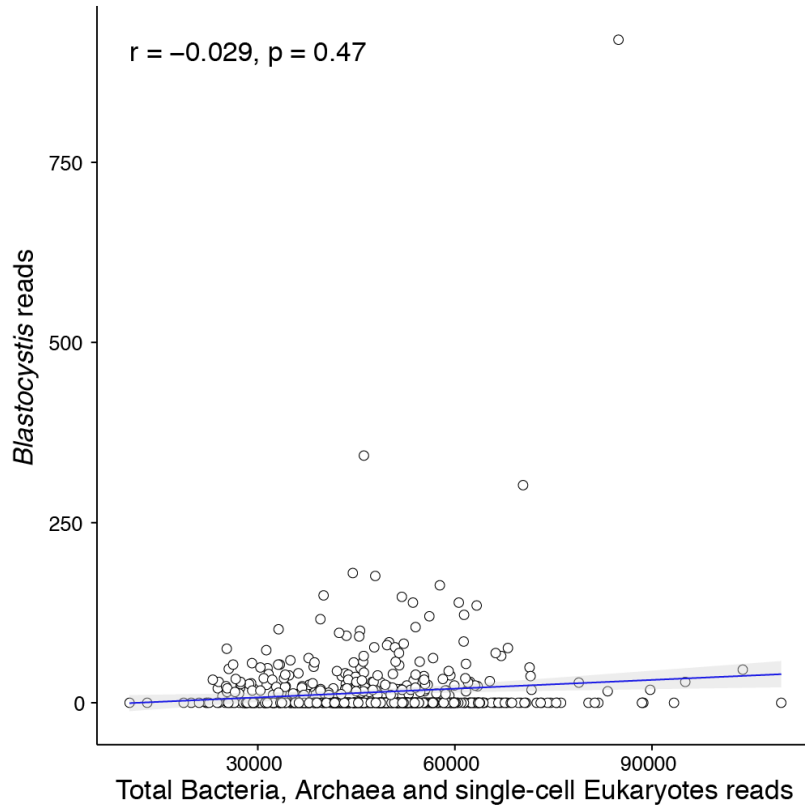


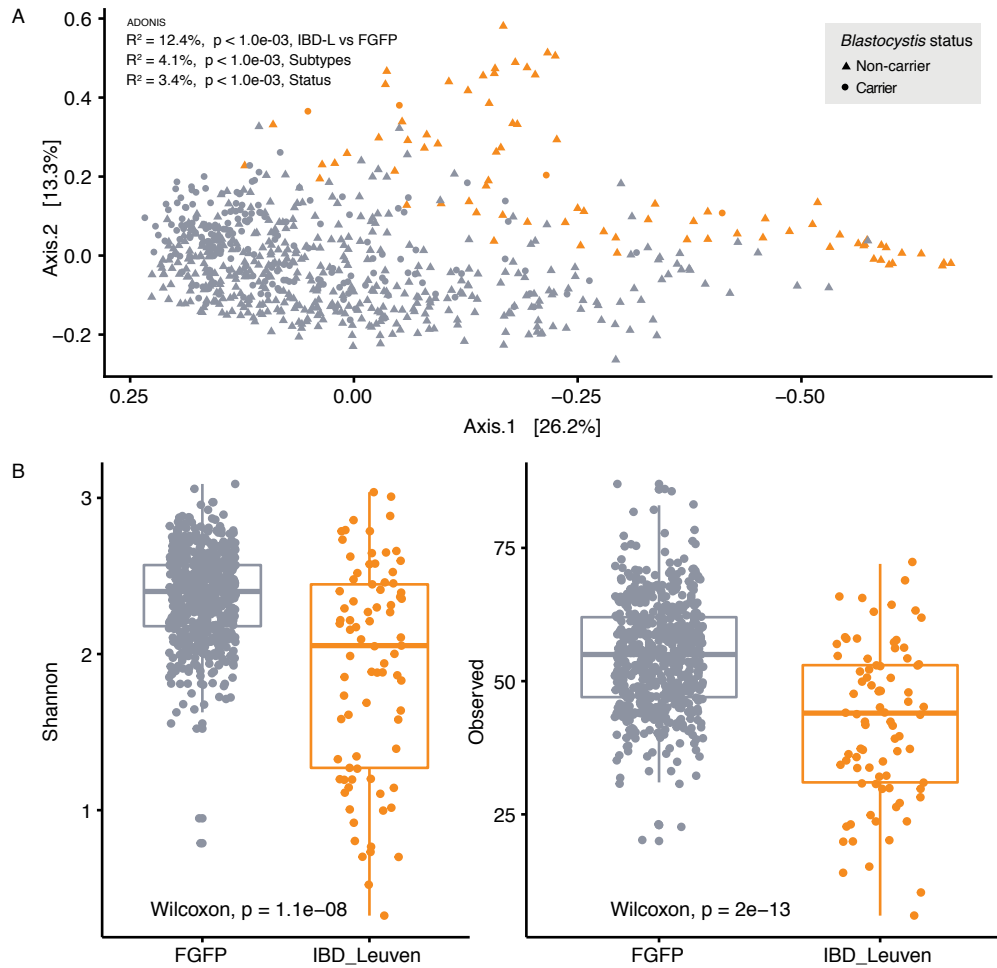
# **Population-level analysis of *Blastocystis* subtype prevalence and variation in the human gut microbiota**

Raul Y. Tito<sup>1,2,3</sup>, Samuel Chaffron<sup>4</sup>, Clara Caenepeel<sup>5</sup>, Gipsi Lima-Mendez<sup>1,2</sup>, Jun Wang<sup>1,2</sup>, Sara Vieira-Silva<sup>1,2</sup>, Gwen Falony<sup>1,2</sup>, Falk Hildebrand<sup>6</sup>, Youssef Darzi<sup>1,2</sup>, Leen Rymenans<sup>1,2</sup>, Chloë Verspecht<sup>1,2</sup>, Peer Bork<sup>6,7,8</sup>, Severine Vermeire<sup>5</sup>, Marie Joossens<sup>1,2</sup>, Jeroen Raes<sup>1,2‡</sup>

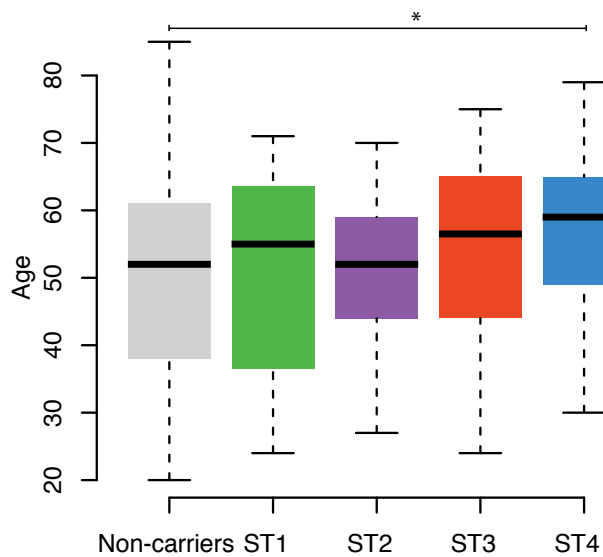
<sup>1</sup>Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven, Herestraat 49, B-3000 Leuven, Belgium; <sup>2</sup>VIB, Center for Microbiology, Herestraat 49, B-3000 Leuven, Belgium; <sup>3</sup>Research Group of Microbiology, Department of Bioengineering Sciences, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. <sup>4</sup>Laboratoire des Sciences du Numérique de Nantes (LS2N), CNRS UMR 6004 – Université de Nantes, École Centrale de Nantes, IMT Atlantique, 2 rue de la Houssinière, 44322 Nantes, France. <sup>5</sup>Translational Research Center for Gastrointestinal Disorders (TARGID), KU Leuven, B-3000 Leuven, Belgium. <sup>6</sup>European Molecular Biology Laboratory, 69117 Heidelberg, Germany. <sup>7</sup>Max Delbrück Centre for Molecular Medicine, 13125 Berlin, Germany. <sup>8</sup>Department of Bioinformatics, University of Würzburg, 97074 Würzburg, Germany.



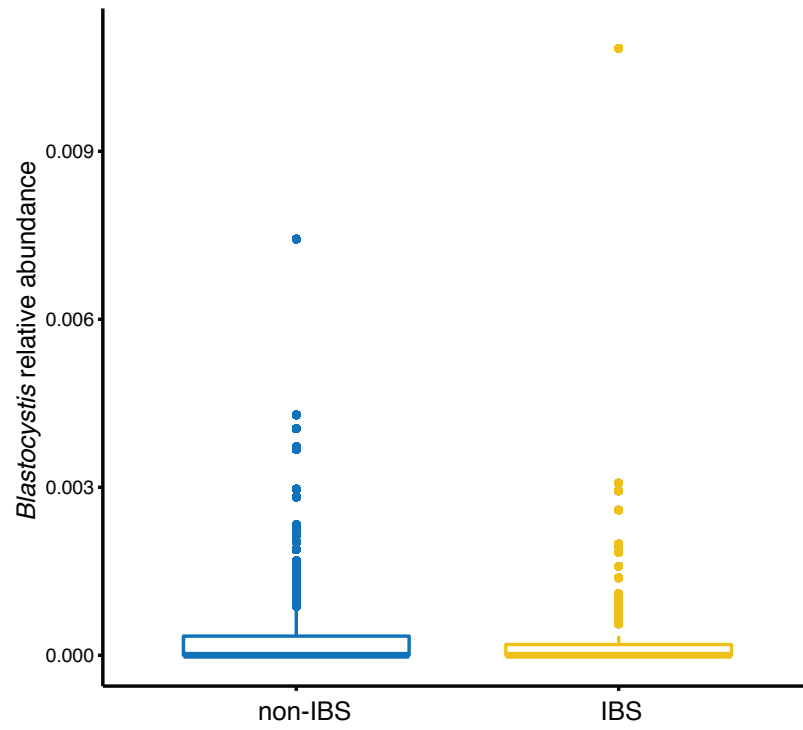
**Supplementary Figure S1.** Correlation of *Blastocystis* reads and total number of reads per sample. There is no correlation between number of reads assigned to *Blastocystis* and the total number of reads generated using a set of primers for 16S rRNA and 18S rRNA genes.



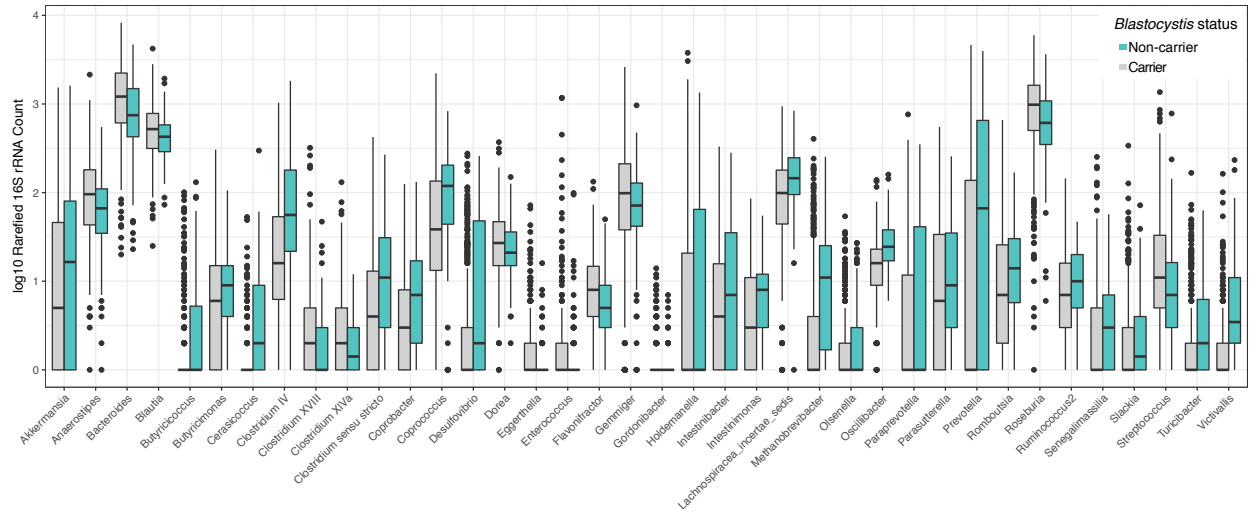
**Supplementary Figure S2.** Microbial community variation in the IBD-L (IBD Leuven cohort). A) Principal coordinate analysis (PCoA) of genus-level Bray-Curtis dissimilarity of the bacterial and archaeal fraction of IBD-L and FGFP. The percentage of variation explained by the two first PCoA dimensions are reported on the axes. B) Shannon diversity index (SDI) and Observed genus richness across IBD-L and FGFP. IBD-L patients have lower SDI and richness compared to FGFP. The body of the box plot represents the first and third quartiles of the distribution, and the median line. The whiskers extend from the quartiles to the last data point within  $1.5 \times \text{IQR}$ , with outliers beyond.



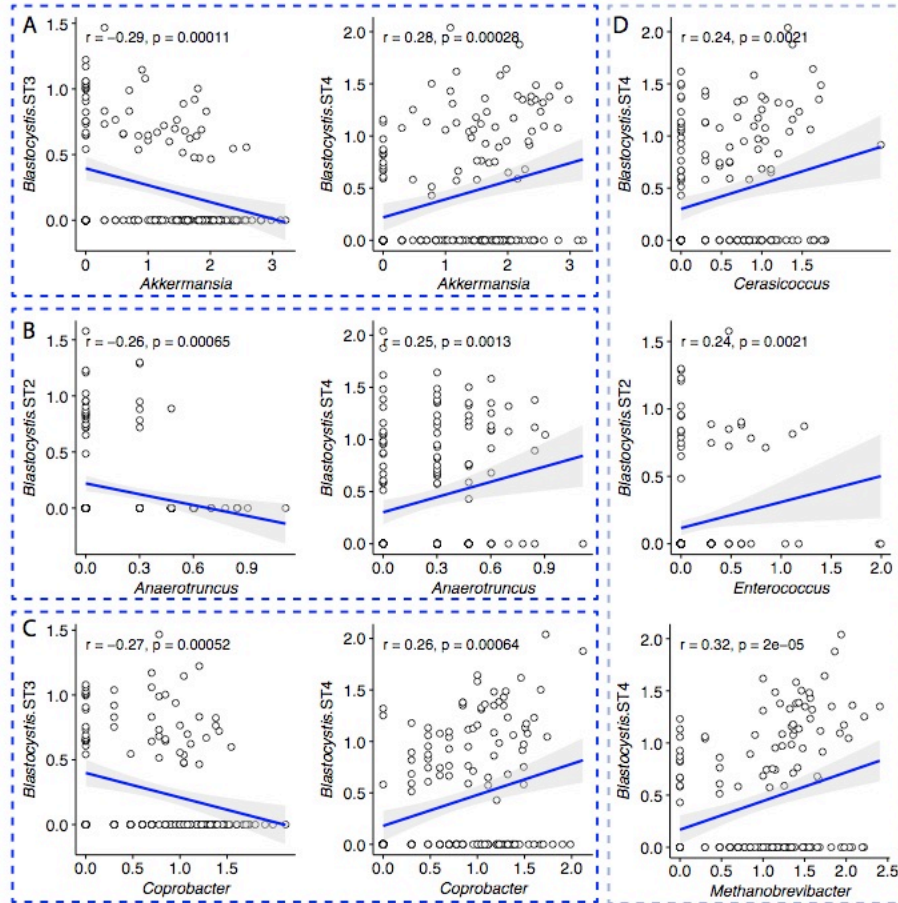
**Supplementary Figure S3.** *Blastocystis* Subtype 4 carriers are significantly older than *Blastocystis* non-carriers. Wilcoxon test, (FDR) (\*) FDR = 7.1e-05.



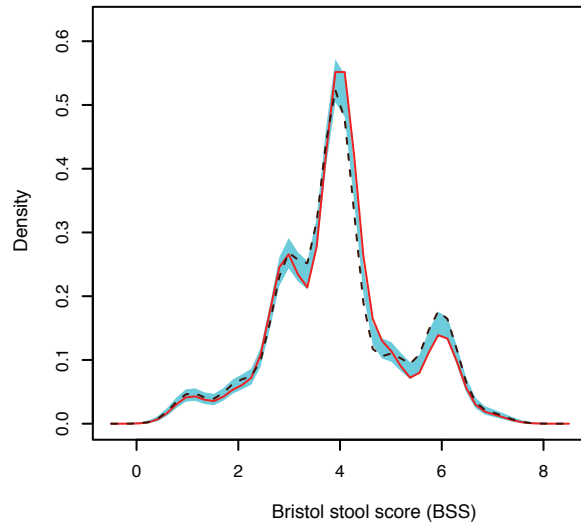
**Supplementary Figure S4.** *Blastocystis* relative abundances are not significantly different between IBS and non-IBS individuals (Wilcoxon test,  $W = 32518$ ,  $p$ -value = 0.50).



**Supplementary Figure S5.** Relative abundances of 38 genera exhibiting significant differential abundance between *Blastocystis* carriers and non-carriers (Kruskal–Wallis test and FDR on genera after filtering by prevalence of 20 % and removing unclassified (unknown) genera (FDR < 0.05)).

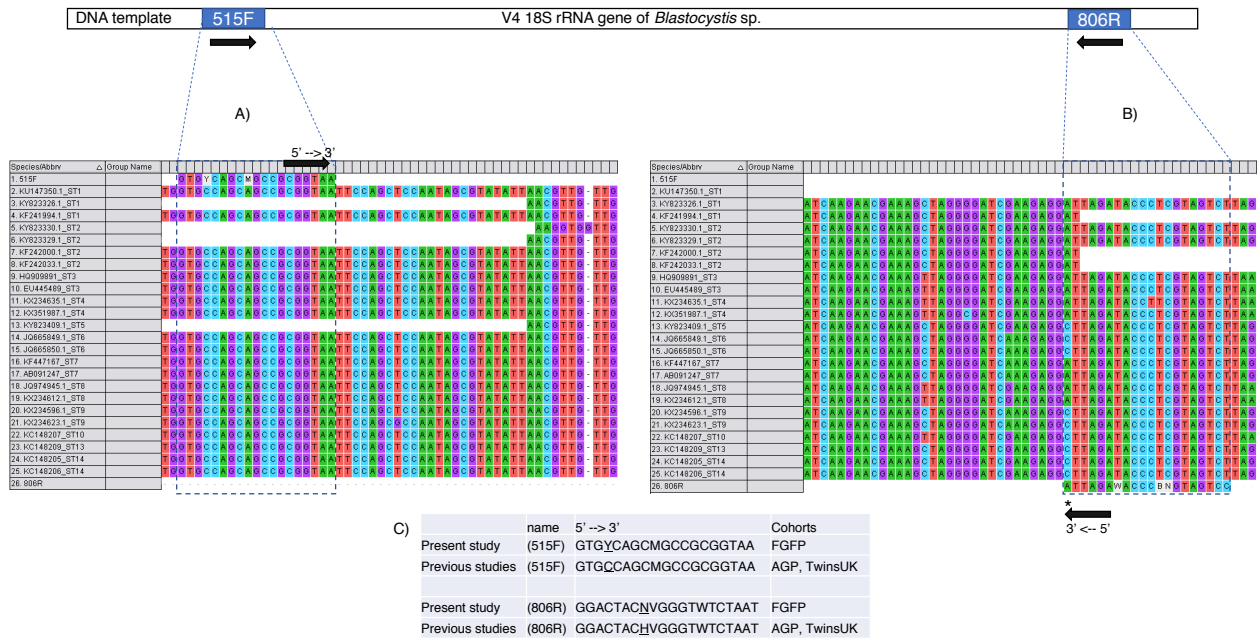


**Supplementary Figure S6.** Association between *Blastocystis* subtypes and Bacteria and Archaea relative abundance. Antagonistic correlations between *Blastocystis* subtypes are presented in A, B, and C. Additional correlations are presented in D. (Spearman correlation, FDR < 0.1).



**Supplementary Figure S7.** Density distributions of BSS in the full FGFP cohort (n=1106, black dashed line) and the analysed subset cohort (n=616, red line).





**Supplementary Figure S8.** Universal primers 515F and 806R Bacteria/Archaea aligned with the different *Blastocystis* subtypes. A) The 515F primer exhibits a perfect alignment with several *Blastocystis* subtypes sequences, including the nine subtypes identified in humans and three other subtypes reported in other vertebrates. B) The 806R primer shows a good alignment with several *Blastocystis* subtypes sequences, including the nine subtypes identified in humans and three other subtypes reported in other vertebrates. Contrary to what is predicted using in-silico tools, the mismatching A in the 3' end of 806R, observed in some *Blastocystis* subtypes (ST5, ST6 and ST9), does not prevent alignment and amplification, as demonstrated by the identification of two of these subtypes (ST6 and ST9) using the same primers in the AGP data (Figure 1, Supplementary Table S4). All *Blastocystis* sequences were retrieved from the NCBI Nucleotide database and aligned using the MEGA7 software<sup>54</sup>. C) The primers used for our study presented three differences with those used for the AGP and TwinUK studies. In the 515F, a C (cytosine) was replaced by a Y (cytosine or

thymine), while in the 806R, a degenerated H (adenine, cytosine and thymine) was substituted by a degenerated N (adenine, cytosine, guanine and thymine).

## LotuS scripts and settings.

#####Bacteria, Archaea and Eukaryotes LotuS run (genus)

```
{PATH}/lotus.pl -i {Fastq_files}/ -m {PATH}/map_file.txt -o out_Directory -s  
{PATH}/sdm_miSeqFR.txt -c {PATH}/IOTUs.cfg -p miSeq  
{PATH}/lotus.pl -i {Fastq_files}/ -m {PATH}/map_file.txt -o out_Directory -s  
{PATH}/sdm_miSeqFR.txt -c {PATH}/IOTUs.cfg -p miSeq -refDB PR2 -simBasedTaxo 1 -  
redoTaxOnly 1 -threads 8
```

##### sdm\_miSeqFR.txt

```
#sdm options file to control sequence quality filtering,  
demultiplexing and preparation (can also be used without  
demultiplexing)  
#* indicates alternative quality filtering options, saved in  
*.add.fna etc. files separately from initial quality filtered  
dataset  
#sequence length refers to sequence length AFTER removal of  
Primers, Barcodes and trimming. this ensures that downstream  
analysis tools will have appropriate sequence information  
#options with a star in front are lenient parameters for mid  
qual sequences (only used for estimating OTU abundance, not for  
OTU building itself).  
minSeqLength 170  
maxSeqLength 1000  
minAvgQuality 27  
*minSeqLength 170  
*minAvgQuality 20  
#truncate total Sequence length to X (length after Barcode,  
Adapter and Primer removals, set to -1 to deactivate)  
TruncateSequenceLength 170  
  
#Ambiguous bases in Sequence  
maxAmbiguousNT 0  
*maxAmbiguousNT 1  
  
#sequence is discarded if a homonucleotide run in sequence is  
longer  
maxHomonucleotide 8  
  
#Filter whole sequence if one window of quality scores is below  
average  
QualWindowWidth 50  
QualWindowThreshold 25
```

```
#Trim the end of a sequence if a window falls below quality
threshold. Useful for removing low quality trailing ends of
sequence
TrimWindowWidth      20
TrimWindowThreshhold 25

#Probabilistic max number of accumulated sequencing errors.
After this length, the rest of the sequence will be deleted.
Complimentary to TrimWindowThreshhold. (-1) deactivates this
option.
maxAccumulatedError 0.75
*maxAccumulatedError -1
#Binomial error model of expected errors per sequence (see
https://github.com/fpusan/moira), to deactivate, set
BinErrorModelAlpha to -1
BinErrorModelMaxExpError 2.5
BinErrorModelAlpha 0.005

#Max Barcode Errors
maxBarcodeErrs 0
maxPrimerErrs 0

#keep Barcode / Primer Sequence in the output fasta file - in a
normal 16S analysis this should be deactivated (0) for Barcode
and de-activated (0) for primer
keepBarcodeSeq 0
keepPrimerSeq 0

#set fastqVersion to 1 if you use Sanger, Illumina 1.8+ or NCBI
SRA files. Set fastqVersion to 2, if you use Illumina 1.3+ -
1.7+ or Solexa fastq files. "auto" will look for typical
characteristics of either of these and choose the quality offset
score automatically.
fastqVersion auto

#if one or more files have a technical adapter still included
(e.g. TCAG 454) this can be removed by setting this option
TechnicalAdapter

#delete X NTs (e.g. if the first 5 bases are known to have
strange biases)
TrimStartNTs 0
```

```
#correct PE header format (1/2) this is to accommodate the
illumina miSeq paired end annotations 2="@XXX 1:0:4" instead of
1="@XXX/1". Note that the format will be automatically detected
PEheaderPairFmt      1
```

```
#sets if sequences without match to reverse primer
(ReversePrimer) will be accepted (T=reject ; F=accept all);
default=F
```

```
RejectSeqWithoutRevPrim  T
```

```
.*RejectSeqWithoutRevPrim      F
```

```
#sets if sequences without a forward (LinkerPrimerSequence)
primer will be accepted (T=reject ; F=accept all); default=F
```

```
RejectSeqWithoutFwdPrim  T
```

```
.*RejectSeqWithoutFwdPrim      F
```

```
#this option should be "T" if your amplicons are possibly
shorter than a single read in a paired end sequencing run (e.g.
if the 16S amplicon length is 200bp in a 250x2 miSeq run, set
this to "T"). This option increases runtime by 10%, if in doubt
just set to "T". *Requires* LinkerPrimerSequence and
ReversePrimer to be defined in mapping file.
```

```
AmpliconShortPE      F
```

```
#options for difficulties during sequencing library construction
#checks if pair1 and pair2 were switched (ignore if single read
data)
```

```
CheckForMixedPairs  F
```

```
#checks if whole amplicon was reverse-transcribed sequenced (not
switched, just reverse translated)
```

```
CheckForReversedSeqs      F
```

```
##### Eukaryotes LotuS run (OTUs)
/{PATH}/lotus.pl -i /{Fastq_files}/ -m /{PATH}/map_file.txt -o out_Directory -s
/{PATH}/sdm_miSeqFR.130.txt -c /{PATH}/LOTUs.cfg -p miSeq -refDB PR2 -simBasedTaxo
1 -threads 8
```

```
##### sdm_miSeqFR.130.txt
```

```
#sdm options file to control sequence quality filtering,
demultiplexing and preparation (can also be used without
demultiplexing)
#* indicates alternative quality filtering options, saved in
*.add.fna etc. files separately from initial quality filtered
dataset
#sequence length refers to sequence length AFTER removal of
Primers, Barcodes and trimming. this ensures that downstream
analysis tools will have appropriate sequence information
#options with a star in front are lenient parameters for mid
qual sequences (only used for estimating OTU abundance, not for
OTU building itself).
minSeqLength 130
maxSeqLength 1000
minAvgQuality 27
*minSeqLength 130
*minAvgQuality 20
#truncate total Sequence length to X (length after Barcode,
Adapter and Primer removals, set to -1 to deactivate)
TruncateSequenceLength 130

#Ambiguous bases in Sequence
maxAmbiguousNT 0
*maxAmbiguousNT 1

#sequence is discarded if a homonucleotide run in sequence is
longer
maxHomonucleotide 8

#Filter whole sequence if one window of quality scores is below
average
QualWindowWidth 50
QualWindowThreshold 25

#Trim the end of a sequence if a window falls below quality
threshold. Useful for removing low quality trailing ends of
sequence
TrimWindowWidth 20
TrimWindowThreshold 25
```

```
#Probabilistic max number of accumulated sequencing errors.
After this length, the rest of the sequence will be deleted.
Complimentary to TrimWindowThreshold. (-1) deactivates this
option.
maxAccumulatedError 0.75
*maxAccumulatedError -1
#Binomial error model of expected errors per sequence (see
https://github.com/fpusan/moira), to deactivate, set
BinErrorModelAlpha to -1
BinErrorModelMaxExpError 2.5
BinErrorModelAlpha 0.005

#Max Barcode Errors
maxBarcodeErrs 0
maxPrimerErrs 0

#keep Barcode / Primer Sequence in the output fasta file - in a
normal 16S analysis this should be deactivated (0) for Barcode
and de-activated (0) for primer
keepBarcodeSeq 0
keepPrimerSeq 0

#set fastqVersion to 1 if you use Sanger, Illumina 1.8+ or NCBI
SRA files. Set fastqVersion to 2, if you use Illumina 1.3+ -
1.7+ or Solexa fastq files. "auto" will look for typical
characteristics of either of these and choose the quality offset
score automatically.
fastqVersion auto

#if one or more files have a technical adapter still included
(e.g. TCAG 454) this can be removed by setting this option
TechnicalAdapter

#delete X NTs (e.g. if the first 5 bases are known to have
strange biases)
TrimStartNTs 0

#correct PE header format (1/2) this is to accommodate the
illumina miSeq paired end annotations 2="@XXX 1:0:4" instead of
1="@XXX/1". Note that the format will be automatically detected
PEheaderPairFmt 1

#sets if sequences without match to reverse primer
(ReversePrimer) will be accepted (T=reject ; F=accept all);
default=F
```

```
RejectSeqWithoutRevPrim  T
#*RejectSeqWithoutRevPrim      F
#sets if sequences without a forward (LinkerPrimerSequence)
primer will be accepted (T=reject ; F=accept all); default=F
RejectSeqWithoutFwdPrim  T
#*RejectSeqWithoutFwdPrim      F

#this option should be "T" if your amplicons are possibly
shorter than a single read in a paired end sequencing run (e.g.
if the 16S amplicon length is 200bp in a 250x2 miSeq run, set
this to "T"). This option increases runtime by 10%, if in doubt
just set to "T". *Requires* LinkerPrimerSequence and
ReversePrimer to be defined in mapping file.
AmpliconShortPE          F

#options for difficulties during sequencing library construction
#checks if pair1 and pair2 were switched (ignore if single read
data)
CheckForMixedPairs      F
#checks if whole amplicon was reverse-transcribed sequenced (not
switched, just reverse translated)
CheckForReversedSeqs    F
```



**Additional file 1**

**Table S1 -- Sequencing data and sample information. FGFP and IBD-L cohorts.**

**Table S2 -- Taxonomic annotation of eukaryotic OTUs using LotuS and BLAST.**

**Table S3 -- Abundance table of Blastocystis subtype OTUs.**

**Table S4 -- UK Twin micro-eukaryotic data (Samples positives for Blastocystis 102, with at least 2 reads).**

**Table S5 -- UK Twin micro-eukaryotic data (Samples positives for Blastocystis 138).**

**Table S6 -- Blastocystis status (Carriers and non-carriers) and phenotypic microbiome covariates (Wilcoxon test).**

**Table S7 -- Blastocystis Subtypes (Non-carriers and different subtypes) and age (Pairwise comparisons using Wilcoxon test).**

**Table S8 -- Bacterial and archaeal genera with distinct relative abundance distributions (Wilcoxon test).**

**Table S9 -- Genera differentially associated to Blastocystis subtypes (Kruskal-Wallis and Wilcoxon tests).**

**Table S10 -- Cooccurrence analysis network.**

**Table S11 -- Distance based redundancy analysis (dbRDA) results.**

**Column Description****Table S1**

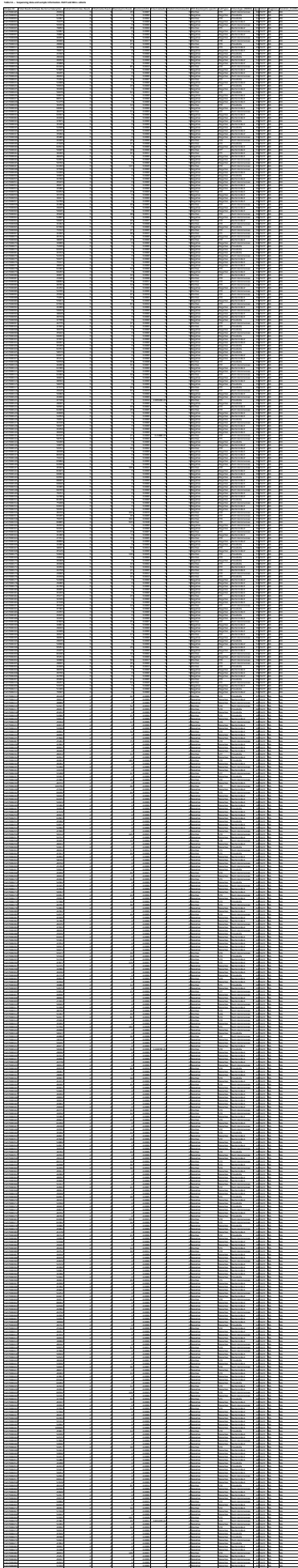
Unique.ID	Sample ID
Total.Reads.Archaea.Bacteria.Eukaryotes	Total Reads
Pentatriconomas.Reads	Reads assigned to Pentatriconomas
Entamoeba.Reads	Reads assigned to Entamoeba
Blastocystis.Reads	Reads assigned to Blastocystis
BlastocystisRA	Blastocystis Relative Abundance (RA)
EntamoebaRA	Entamoeba Relative Abundance
PentatriconomasRA	Pentatriconomas Relative Abundance
PCR.Blastocystis.specific	Presence (1) or Absence (0) of Blastocystis as investi
SubTypes	Blastocystis subtype
Enterotype (DMM4)	Enterotype as defined in Falony et al., Science 2016
Age	Age

**Table S2**

OTU_PR2	OTUs affiliated to Eukarya
Sequence	Sequence
Len	Length
Kingdom	Taxon
Phylum	Taxon
Class	Taxon
Order	Taxon
Family	Taxon
Genus	Taxon
Species	Taxon
Blast_best_match	Accession Number of best match
Max score	Maximum Score
Total score	Total Score
Query cover	Sequence coverage
E value	Expected value
Ident	Identity
name	Description of best match
Subtype assignation	Subtype

**Table S11**

CAP Variable	Type of variable
stepRDA	stepRDA
Cumulative R2	Cumulative R2
group	Type of variable
CAP r2	CAP r2
FDR	q-value (Benjamini Hochberg FDR)
exclusion from plot	exclusion from Figure 4



This image shows a vertical sheet of graph paper. The grid consists of 20 columns and 100 rows of small squares. The lines are thin and black, forming a consistent pattern across the entire page.



Table S2 – Taxonomic annotation of eukaryotic OTUs using LotuS and BLAST

OTU	Sequnce	Len	Kingdom	Phylum	Class	Order	Family	Genus	Species	Blast best #	Max score	Total score	Query cover	E value	Ident	name	Subtype assignment
OTU 306	TTCCAGCTCGAATAGGTTATA	205	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242081.1	372	372	100%	2.00E-99	100%	Blastocystis sp. subtype 4 strain KK126 small subunit ribosomal RNA gene, partial sequence	ST4
OTU 533	TTCCAGCTCGAATAGGTTATA	413	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242073.1	771	771	100%	0	100%	Blastocystis sp. subtype 3 strain KK126 small subunit ribosomal RNA gene, partial sequence	ST3
OTU 702	TTCCAGCTCGAATAGGTTATA	413	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.241954.1	763	763	100%	0	100%	Blastocystis sp. subtype 1 strain KK079 small subunit ribosomal RNA gene, partial sequence	ST1
OTU 867	TTCCAGCTCGAATAGGTTATA	412	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242060.1	761	761	100%	0	100%	Blastocystis sp. subtype 2 strain KK161 small subunit ribosomal RNA gene, partial sequence	ST2
OTU 1083	TTCCAGCTCGAATAGGTTATA	412	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242062.1	761	761	100%	0	100%	Blastocystis sp. subtype 2 strain KK046 small subunit ribosomal RNA gene, partial sequence	ST2
OTU 1644	TTCCAGCTCGAATAGGTTATA	205	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242033.1	363	363	100%	1.00E-96	99%	Blastocystis sp. subtype 2 strain KK119 small subunit ribosomal RNA gene, partial sequence	ST2
OTU 1688	TTCCAGCTCGAATAGGTTATA	205	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.241720.1	359	359	100%	2.00E-95	99%	Blastocystis sp. subtype 1 strain BK118, 18S small subunit ribosomal RNA gene, partial sequence	ST1
OTU 1773	TTCCAGCTCGAATAGGTTATA	226	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242040.1	360	360	100%	1.00E-97	97%	Blastocystis sp. clone 2 (Opalimon) Opalimon (Opalimon) 18S ribosomal RNA gene, partial sequence	Unk
OTU 1809	TTCCAGCTCGAATAGGTTATA	421	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.241715.1	758	758	100%	0	99%	Blastocystis 177 strain 8, clone v19, 178S ribosomal RNA gene, partial sequence	ST7
OTU 1831	TTCCAGCTCGAATAGGTTATA	415	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242485.1	756	756	100%	0	99%	Blastocystis sp. subtype 3 strain KK148, 18S ribosomal RNA gene, complete sequence	ST8
OTU 2093	TTCCAGCTCGAATAGGTTATA	411	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242024.1	754	754	100%	0	99%	Blastocystis sp. subtype 3 strain KK126 small subunit ribosomal RNA gene, partial sequence	ST3
OTU 2380	TTCCAGCTCGAATAGGTTATA	426	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242024.1	760	760	100%	0	95%	Blastocystis sp. subtype 3 strain KK126 small subunit ribosomal RNA gene, partial sequence	Unk
OTU 4258	TTCCAGCTCGAATAGGTTATA	426	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242024.1	660	660	100%	0	95%	Blastocystis sp. subtype 3 strain KK126 small subunit ribosomal RNA gene, partial sequence	Unk
OTU 4846	TTCCAGCTCGAATAGGTTATA	412	Eukaryota	Stramenopila	Opalinata	X	Blastocystis	Blastocystis	Blastocystis	42.242000.1	739	739	100%	0	99%	Blastocystis sp. subtype 2 strain KK300 small subunit ribosomal RNA gene, partial sequence	ST2
OTU 4950	TTCCAGCTCGAATAGGTTATA	220	Eukaryota	Amoebozoa	Archamoeba	Archamoeba	Entamoeba	Entamoeba	Entamoeba	42.242424.1	401	401	100%	3.00E-108	95%	Entamoeba coli strain 18S rRNA gene, full-length	Entamoeba
OTU 2855	TTCCAGCTCGGAGTTTGT	315	Eukaryota	Excavata	Parabasala	Trichomonas	Trichomonas	Pentasticho	Pentasticho	42.242424.1	579	579	100%	1.00E-161	100%	Pentastichomonas hominis 18S ribosomal RNA gene, partial sequence (internal transcribed spacer 1)	Pentastichomonas















Table S4 - AGP micro-eukaryotic data (Samples positives for Blastocystis 445).

ID	Entamoeba	Acanthamoeba	Naegleria	Trichomonas	Dientamoeba	Allovalhampflia	Trichomonas	Vahkampflia	Haemomonchus	ST1	ST2	ST3	ST4	ST6	ST7	ST8	ST9	STUNK	Total	Number of no Blastocystis reads	Blastocystis reads	multiple single-cell Eukaryotes	multiple Blastocystis	SubType
AG_000031524.6983	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000031567.6642	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3	0	0	0	0	ST2
AG_000031799.8967	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST2
AG_000031792.7036	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	11	0	0	0	0	ST2
AG_000032795.7037	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST2
AG_000032814.7038	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3	0	0	0	0	ST2
AG_000032887.7046	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4	0	0	0	0	ST2
AG_000032888.7047	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	12	0	0	0	0	ST2
AG_000032994.6668	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	8	0	0	0	0	ST2
AG_000032958.9014	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	5	0	0	0	0	ST2
AG_000032968.9017	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	31	0	0	0	0	ST2
AG_000033138.8599	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST2
AG_000033438.9050	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST2
AG_000033439.9051	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	11	0	0	0	0	ST2
AG_000033440.9052	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST2
AG_000033441.9053	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST2
AG_000033473.8597	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	20	0	0	0	0	ST2
AG_000034125.7563	0	0	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	38	0	0	0	0	ST2
AG_000035420.7137	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST2
AG_000037982.9079	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	15	0	0	0	0	ST2
AG_000038061.8662	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	8	0	0	0	0	ST2
AG_000038078.8096	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST2
AG_000038187.9118	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	10	0	0	0	0	ST2
AG_000038189.9119	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	5	0	0	0	0	ST2
AG_000038360.9138	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4	0	0	0	0	ST2
AG_000041256.8411	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	10	0	0	0	0	ST2
AG_000041976.8461	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST2
AG_000038098.7619	0	0	0	0	0	0	0	0	0	0	10	0	3	0	0	0	0	0	13	0	0	13	yes	ST1ST3
AG_000041343.7242	0	0	0	0	0	0	0	0	0	0	3027	0	0	0	0	0	0	0	3027	0	0	3027	yes	ST1Entamoeba
AG_000001099.25	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4	0	0	0	0	ST1
AG_000002073.509	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3	0	0	0	0	ST1
AG_000001670.908	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3	0	0	0	0	ST1
AG_000002928.4685	0	0	0	0	0	0	0	0	0	0	44	0	0	0	0	0	0	0	44	0	0	0	0	ST1
AG_000006635.1063	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000007643.1460	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	6	0	0	0	0	ST1
AG_000007173.1465	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4	0	0	0	0	ST1
AG_000009101.1506	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000009218.1152	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	7	0	0	0	0	ST1
AG_000007771.1792	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4	0	0	0	0	ST1
AG_000011350.1956	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000011341.1965	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST1
AG_000011092.1992	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	12	0	0	0	0	ST1
AG_000012486.3424	0	0	0	0	0	0	0	0	0	0	102	0	0	0	0	0	0	0	102	0	0	0	0	ST1
AG_000012334.5223	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	5	0	0	0	0	ST1
AG_000013255.7271	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	7	0	0	0	0	ST1
AG_000013440.2563	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	6	0	0	0	0	ST1
AG_000014111.2684	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	6	0	0	0	0	ST1
AG_000014132.3441	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000014971.2064	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000015329.2951	0	0	0	0	0	0	0	0	0	0	51	0	0	0	0	0	0	0	51	0	0	0	0	ST1
AG_000016333.3498	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000017490.6186	0	0	0	0	0	0	0	0	0	0	73	0	0	0	0	0	0	0	73	0	0	0	0	ST1
AG_000002937.7281	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	11	0	0	0	0	ST1
AG_000021945.4528	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000025252.7319	0	0	0	0	0	0	0	0	0	0	88	0	0	0	0	0	0	0	88	0	0	0	0	ST1
AG_000002027.6195	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000026426.5772	0	0	0	0	0	0	0	0	0	0	144	0	0	0	0	0	0	0	144	0	0	0	0	ST1
AG_000024688.6816	0	0	0	0	0	0	0	0	0	0	144	0	0	0	0	0	0	0	144	0	0	0	0	ST1
AG_000027245.5804	0	0	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	38	0	0	0	0	ST1
AG_000027318.5817	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	5	0	0	0	0	ST1
AG_000027898.6235	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000028105.7391	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000028154.5901	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	9	0	0	0	0	ST1
AG_000028445.5926	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	19	0	0	0	0	ST1
AG_000028711.5953	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST1
AG_000029571.8943	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	ST1
AG_000029430.6040	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST1
AG_000029483.6364	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	ST1
AG_000029596.6369	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	4	0	0	0	0	ST1
AG_000030124.6446	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	47	0	0	0	0	ST1
AG_0000																								

Table S5 -- UK Twin micro-eukaryotic data (Samples positives for Blastocystis 138)

Twins UK micro-eukaryotic data	Dientamoeba	Entamoeba	ST1	ST2	ST3	ST4	ST7	Total of micro-eukaryotes reads	No Blastocystis reads	Blastocystis reads	Multiple single-cell Eu	Multiple Blastocystis subtypes	SubType
ERR561184	0	0	0	0	0	94	0	94	0	94	no		ST4
ERR561054	0	3	0	0	65	0	68	68	3	65	yes	no	ST4Entamoeba
ERR561159	0	0	0	0	62	0	62	0	0	62	no	no	ST4
ERR560996	0	0	0	0	52	0	52	0	0	52	no	no	ST4
ERR561305	2	0	0	0	46	0	48	48	2	46	yes	no	ST4Dientamoeba
ERR561547	0	0	0	0	45	0	45	0	0	45	no	no	ST4
ERR561509	0	0	0	0	42	0	42	0	0	42	no	no	ST4
ERR560767	0	0	0	0	35	0	35	0	0	35	no	no	ST4
ERR561391	0	0	0	0	32	0	32	0	0	32	no	no	ST1
ERR561605	0	0	0	0	30	0	30	0	0	30	no	no	ST4
ERR560994	0	0	0	0	29	0	29	0	0	29	no	no	ST3
ERR560671	0	0	0	0	27	0	27	0	0	27	no	no	ST4
ERR561617	0	0	26	0	0	0	26	0	0	26	no	no	ST1
ERR561638	0	0	0	0	24	0	24	0	0	24	no	no	ST4
ERR561649	0	0	0	0	24	0	24	0	0	24	no	no	ST4
ERR560636	0	0	0	0	20	0	20	0	0	20	no	no	ST4
ERR561561	0	0	0	0	20	0	20	0	0	20	no	no	ST4
ERR561404	0	0	0	19	0	0	19	0	0	19	no	no	ST1
ERR560941	0	0	0	19	0	0	19	0	0	19	no	no	ST3
ERR561617	0	0	0	19	0	0	19	0	0	19	no	no	ST3
ERR561618	0	0	0	0	18	0	18	0	0	18	no	no	ST4
ERR560881	1	0	0	0	17	0	18	1	17	17	yes	no	ST4Dientamoeba
ERR560776	0	0	0	16	0	0	16	0	0	16	no	no	ST1
ERR560713	0	0	13	0	0	0	13	0	0	13	no	no	ST1
ERR561283	0	0	15	0	0	0	15	0	0	15	no	no	ST1
ERR560588	0	0	14	0	0	0	14	0	0	14	no	no	ST1
ERR561436	0	0	0	14	0	0	14	0	0	14	no	no	ST3
ERR560788	0	0	0	13	0	0	13	0	0	13	no	no	ST4
ERR561055	0	0	0	0	13	0	13	0	0	13	no	no	ST4
ERR561393	0	0	12	0	0	0	12	0	0	12	no	no	ST1
ERR561452	0	0	0	12	0	0	12	0	0	12	no	no	ST1
ERR561392	0	0	0	12	0	0	12	0	0	12	no	no	ST4
ERR561352	0	0	0	0	12	0	12	0	0	12	no	no	ST4
ERR560658	0	0	0	0	12	0	12	0	0	12	no	no	ST7
ERR560736	0	0	0	11	0	0	11	0	0	11	no	no	ST1
ERR561565	0	0	0	11	0	0	11	0	0	11	no	no	ST3
ERR560640	0	0	0	0	11	0	11	0	0	11	no	no	ST4
ERR561245	0	0	0	0	11	0	11	0	0	11	no	no	ST4
ERR561418	0	0	0	0	11	0	11	0	0	11	no	no	ST4
ERR561422	0	11	0	0	11	0	11	0	0	11	no	no	ST4
ERR561624	0	0	0	0	11	0	11	0	0	11	no	no	ST4
ERR561175	0	0	0	0	10	0	10	0	0	10	no	no	ST4
ERR561349	0	0	0	9	0	0	9	0	0	9	no	no	ST2
ERR561647	0	0	0	9	0	0	9	0	0	9	no	no	ST4
ERR561622	0	0	0	0	8	0	8	0	0	8	no	no	ST3
ERR561369	0	0	0	0	8	0	8	0	0	8	no	no	ST4
ERR561537	0	0	0	0	8	0	8	0	0	8	no	no	ST4
ERR561610	0	0	0	0	8	0	8	0	0	8	no	no	ST4
ERR561288	0	0	0	7	0	0	7	0	0	7	no	no	ST2
ERR561613	0	0	0	7	0	0	7	0	0	7	no	no	ST2
ERR560881	0	0	0	7	0	0	7	0	0	7	no	no	ST4
ERR560986	0	0	6	0	0	0	6	0	0	6	no	no	ST1
ERR561060	0	0	6	0	0	0	6	0	0	6	no	no	ST1
ERR561487	0	0	6	0	0	0	6	0	0	6	no	no	ST2
ERR561343	0	0	0	6	0	0	6	0	0	6	no	no	ST3
ERR560750	0	0	0	6	0	0	6	0	0	6	no	no	ST4
ERR560838	0	0	0	6	0	0	6	0	0	6	no	no	ST4
ERR561496	0	0	5	0	0	0	5	0	0	5	no	no	ST1
ERR560725	0	0	0	5	0	0	5	0	0	5	no	no	ST2
ERR561553	0	0	0	5	0	0	5	0	0	5	no	no	ST2
ERR561155	0	0	0	5	0	0	5	0	0	5	no	no	ST3
ERR561198	0	0	0	5	0	0	5	0	0	5	no	no	ST3
ERR561329	0	0	0	5	0	0	5	0	0	5	no	no	ST3
ERR560620	0	0	0	0	5	0	5	0	0	5	no	no	ST4
ERR560800	0	0	0	0	5	0	5	0	0	5	no	no	ST4
ERR561009	0	0	0	0	5	0	5	0	0	5	no	no	ST4
ERR561465	0	0	0	0	5	0	5	0	0	5	no	no	ST4
ERR561545	0	0	0	0	5	0	5	0	0	5	no	no	ST4
ERR560612	0	0	0	4	0	0	4	0	0	4	no	no	ST3
ERR560845	0	0	0	4	0	0	4	0	0	4	no	no	ST3
ERR561317	0	0	0	4	0	0	4	0	0	4	no	no	ST3
ERR561494	0	0	0	4	0	0	4	0	0	4	no	no	ST3
ERR560968	0	0	0	0	4	0	4	0	0	4	no	no	ST4
ERR561464	0	0	0	0	4	0	4	0	0	4	no	no	ST4
ERR561669	0	0	0	0	4	0	4	0	0	4	no	no	ST4
ERR561650	1	0	0	0	3	0	4	1	3	3	yes	no	ST4Dientamoeba
ERR560815	0	0	3	0	0	0	3	0	0	3	no	no	ST1
ERR560867	0	0	3	0	0	0	3	0	0	3	no	no	ST1
ERR560788	0	0	0	3	0	0	3	0	0	3	no	no	ST2
ERR561520	0	0	0	3	0	0	3	0	0	3	no	no	ST2
ERR561035	0	0	0	3	0	0	3	0	0	3	no	no	ST3
ERR561253	0	0	0	3	0	0	3	0	0	3	no	no	ST3
ERR560688	0	0	0	3	0	0	3	0	0	3	no	no	ST4
ERR560683	0	0	0	0	3	0	3	0	0	3	no	no	ST4
ERR560885	0	0	0	0	3	0	3	0	0	3	no	no	ST4
ERR560955	0	0	0	0	3	0	3	0	0	3	no	no	ST4
ERR561356	0	0	0	0	3	0	3	0	0	3	no	no	ST4
ERR561409	0	0	0	0	3	0	3	0	0	3	no	no	ST4
ERR561454	0	0	0	0	3	0	3	0	0	3	no	no	ST4
ERR561325	0	0	0	2	0	0	2	0	0	2	no	no	ST2
ERR561414	0	0	0	2	0	0	2	0	0	2	no	no	ST2
ERR561498	0	0	0	2	0	0	2	0	0	2	no	no	ST3
ERR560602	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR560644	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR560921	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR561018	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR561162	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR561326	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR561349	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR561508	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR561588	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR561625	0	0	0	0	2	0	2	0	0	2	no	no	ST4
ERR560921	0	1	0	0	1	0	2	1	1	1	yes	no	ST4Entamoeba
ERR561246	0	0	1	0	0	0	1	0	0	1	no	no	ST1
ERR560675	0	0	0	1	0	0	1	0	0	1	no	no	ST2
ERR560728	0	0	0	1	0	0	1	0	0	1	no	no	ST2
ERR560854	0	0	0	1	0	0	1	0	0	1	no	no	ST2
ERR561093	0	0	0	1	0	0	1	0	0	1	no	no	ST2
ERR561376	0	0	0	1	0	0	1	0	0	1	no	no	ST2
ERR561465	0	0	0	1	0	0	1	0	0	1	no	no	ST2
ERR561484	0	0	0	1	0	0	1	0	0	1	no	no	ST2
ERR560813	0	0	0	1	0	0	1	0	0	1	no	no	ST3
ERR561033	0	0	0	1	0	0	1	0	0	1	no	no	ST3
ERR561092	0	0	0	1	0	0	1	0	0	1	no	no	ST3
ERR561285	0	0	0	1	0	0	1	0	0	1	no	no	ST3
ERR561362	0	0	0	1	0	0	1	0	0	1	no	no	ST3
ERR561602	0	0	0	1	0	0	1	0	0	1	no	no	ST3
ERR560575	0	0	0	0	1	0	1	0	0	1	no	no	ST4
ERR560617	0	0	0	0	1	0	1	0	0	1	no	no	ST4
ERR560670	0	0	0	0	1	0	1	0	0	1	no	no	ST4
ERR560743	0	0	0	0	1	0	1	0	0	1	no	no	ST4
ERR560843	0	0	0	0	1	0	1	0	0	1	no	no	ST4
ERR560851	0	0	0	0	1	0	1	0	0	1	no	no	ST4
ERR560886	0	0	0	0	1	0	1	0	0	1	no	no	ST4
ERR560920	0	0	0	0	1	0	1	0	0	1	no	no	

Table S6 -- Blastocystis status (Carriers and non-carriers) and phenotypic microbiome covariates (Wilcoxon test)

Wilcoxon test	W	p-value	q-values (FDR)
Age	42951	0.00032037	0.022
HOMA_IR	15697.5	0.004345176	0.143
sugary_soda_consumption_frequency_last_week	31526	0.006221943	0.143
ANTIBIOTICS_J01CA04	33082	0.008481428	0.146
GFR_CDK.EPI	31488	0.014717572	0.203
BetaGlobulines.g.L	31750	0.021310298	0.239
chocolate_type	39847.5	0.027570115	0.239
Ureum.mg.dL	40300.5	0.027698895	0.239
Alfa2Globulines.g.L	32390.5	0.049375497	0.364
fruits_consumption_frequency_last_week	39619	0.058016837	0.364
RBC.milj.mm3	32470.5	0.054613054	0.364
coffee_last_consumption_before_sampling	32898.5	0.073870642	0.368
GPT.U.L	32738	0.074617914	0.368
meat_average_consumption_last_week	32885	0.071346977	0.368
Gastroenteritis_other_treatment	35001	0.083026001	0.382
age_at_first_work	38818.5	0.09099964	0.384
soy_products_average_consumption_last_week	38476.5	0.100250418	0.384
Triglyceriden.mg.dL	32955	0.095551938	0.384
MEDICATION_A06AD15_65	35616	0.124164655	0.428
MEDICATION_L04AX01	35616	0.124164655	0.428
ANTIBIOTICS_J01CR02	34779	0.133467846	0.439
bloating_last_week	34432.5	0.18655845	0.483
BMI	33693	0.201220561	0.483
do_you_want_to_do_something_about_your_weight	38324.5	0.186892217	0.483
FOODSUPPLEMENT_Mg	35579	0.199584985	0.483
how_do_you_find_yourself	38459	0.191580058	0.483
MEDICATION_G03AA12	35784	0.210107769	0.483
MEDICATION_L04AB	35784	0.210107769	0.483
most_eaten_bread_type	37774	0.158929629	0.483
workQuestions.workStyle.option.13	35300	0.204004576	0.483
MEDICATION_N06AX16	36513	0.230565324	0.513
CK.U.L	38267.5	0.258077677	0.556
alcohol_average_consumption_last_week	34341	0.331095038	0.598
ANTIBIOTICS_J01XE01	35626	0.303045178	0.598
beer_consumption	34489.5	0.355389648	0.598
constipated_last_week	34608	0.321268735	0.598
frequency_of_sick_leaves_last_year	34468.5	0.354735283	0.598
MEDICATION_G03CA04_CC06	36382	0.328444635	0.598
PARTICIPANT_Inflammatory_Bowel_Disease_Colitis_ulcerosa	35626	0.303045178	0.598
sleeping_hours_per_day	37881	0.324034248	0.598
Urinezuur.mg.dL	34335	0.347037972	0.598
fruits_last_consumption_before_sampling	34573	0.37832461	0.607
Gamma.GT.U.L	34435	0.374498341	0.607
number_of_work_daysaweek	34613	0.39674129	0.622
MEDICATION_G03DA04	36251	0.490281737	0.752
MCHC.g.dL	37278.5	0.525674971	0.789
PARTICIPANTS_ALLERGY_food	35656	0.542385031	0.796
averageCigarettesConsumptionPerDayLast3Months	35584.5	0.556904851	0.797
last_week_break_fast_count	36851.5	0.565649504	0.797
Hemoglobine.g.dL	35081	0.584136983	0.799
PARTICIPANT_Spastische_darm	35376	0.590398953	0.799
average_week_work_hours	35400	0.692249172	0.821
doYouInhale	35447.5	0.689925903	0.821
FOODSUPPLEMENT_Q10	35999	0.686199073	0.821
HDL.chol..mg.dL	36808	0.717068516	0.821
MEDICATION_A07EC02	35999	0.686199073	0.821
MEDICATION_R06AX28	35999	0.686199073	0.821
PARTICIPANT_CVS	36469	0.651248689	0.821
Smoker	35770	0.710429979	0.821
spirits_consumption	35308.5	0.634345954	0.821
stoole_score	35497.5	0.726212656	0.821
Gender	36637	0.751769403	0.837
previous_relief	36515.5	0.831166528	0.91
MEDICATION_N03AE01	36046	0.844739773	0.911
Creatinine.mg.dL	35845	0.884828606	0.925
height	36420.5	0.874183217	0.925
householdCompositionQuestions.petCohabitationLast3Months.option.2	36229	0.941006704	0.969
Heupontrek	33138	0.967732776	0.982
residency_desc_younger_than_5	36097.5	0.989917711	0.99

**Table S7 -- Blastocystis Subtypes (Non-carriers and different subtypes) and age (Pairwise comparisons using Wilcoxon test)**  
P value adjustment method: fdr

Pairwise comparisons using Wilcoxon test	Non-carriers	ST01	ST02	ST03
ST01	0.72591	-	-	-
ST02	0.72591	0.984	-	-
ST03	0.15295	0.69522	0.55738	-
ST04	0.00071	0.23687	0.15102	0.44908

**Table S8 -- Bacterial and archaeal genera with distinct relative abundance distributions (Wilcoxon test)**

	W	p-value	q-values
Akkermansia	30275	1.6E-03	3.4E-03
Anaerostipes	44538	9.3E-06	2.7E-05
Bacteroides	45487	8.1E-07	3.2E-06
Blautia	42300	1.1E-03	2.5E-03
Butyrivicoccus	31098	6.5E-04	1.6E-03
Butyrivicimonas	30339	2.2E-03	4.4E-03
Cerasicoccus	20888	1.3E-25	3.9E-24
Clostridium IV	20530	2.2E-16	2.6E-15
Clostridium sensu stricto	25968	6.3E-08	3.4E-07
Clostridium XIVa	44882	2.0E-06	6.9E-06
Clostridium XVIII	45080	5.8E-07	2.5E-06
Coprobacter	25256	5.4E-09	3.6E-08
Coprococcus	24122	2.6E-10	2.6E-09
Desulfovibrio	26353	1.4E-09	1.2E-08
Dorea	41248	6.9E-03	1.3E-02
Eggerthella	43965	2.0E-07	1.0E-06
Enterococcus	40194	7.5E-03	1.3E-02
Flavonifractor	45409	9.5E-07	3.6E-06
Gemmiger	41212	7.3E-03	1.3E-02
Gordonibacter	40235	2.3E-03	4.4E-03
Holdemanella	30743	1.1E-03	2.5E-03
Intestinibacter	31238	9.1E-03	1.5E-02
Intestinimonas	28259	2.3E-05	5.9E-05
Lachnospiracea_incertae_sedis	24858	3.0E-09	2.3E-08
Methanobrevibacter	20190	1.1E-20	2.3E-19
Olsenella	31510	4.2E-03	7.9E-03
Oscillibacter	20049	2.5E-17	3.8E-16
Paraprevotella	28800	7.4E-06	2.3E-05
Parasutterella	31869	2.4E-02	3.8E-02
Prevotella	26042	3.0E-08	1.8E-07
Romboutsia	28548	6.3E-05	1.6E-04
Roseburia	45698	4.6E-07	2.1E-06
Ruminococcus2	30244	1.9E-03	4.0E-03
Senegalimassilia	28444	8.0E-06	2.4E-05
Slackia	32278	2.2E-02	3.5E-02
Streptococcus	44438	1.2E-05	3.2E-05
Turicibacter	28028	2.1E-06	6.9E-06
Victivallis	17056	1.0E-29	6.0E-28

Table S9 -- Genera differentially associated to Blastocystis subtypes (Kruskal-Wallis and Wilcoxon tests).

Kruskal-Wallis	Chi2	p-value	q-values (FDR)	Wilcoxon (effect size)	ST1_ST2_r	ST1_ST3_r	ST1_ST4_r	ST2_ST3_r	ST2_ST4_r	ST3_ST4_r
Anaerotruncus	16.133	0.001065	0.06391265	Anaerotruncus	0.32			0.27	0.41	0.19
Coprobacter	13.386	0.003872	0.08760931	Coprobacter				0.26		0.33
Enterococcus	13.122	0.00438	0.08760931	Enterococcus				0.35	0.30	
Akkermansia	12.488	0.005885	0.08827822	Akkermansia						0.32
Cerasioccus	9.7093	0.02121	0.21206325	Wilcoxon FDR	ST1_ST2_FRD	ST1_ST3_FRD	ST1_ST4_FRD	ST2_ST3_FRD	ST2_ST4_FRD	ST3_ST4_FRD
Methanobrevibacter	9.9096	0.01935	0.21206325	Anaerotruncus	0.046	0.675	0.378	0.046	0.000	0.064
Turicibacter	7.9359	0.04736	0.40590052	Coprobacter	0.381	0.479	0.218	0.070	0.874	0.002
Alistipes	6.1041	0.1067	0.41185265	Enterococcus	0.324	0.113	0.212	0.011	0.011	0.489
Coproccoccus	6.2937	0.09816	0.41185265	Akkermansia	0.498	0.330	0.330	0.108	0.630	0.003
Flavonifractor	6.3657	0.09511	0.41185265	Wilcoxon p-value	ST1_ST2_p-value	ST1_ST3_p-value	ST1_ST4_p-value	ST2_ST3_p-value	ST2_ST4_p-value	ST3_ST4_p-value
Intestinibacter	6.3455	0.09596	0.41185265	Anaerotruncus	0.023	0.675	0.315	0.016	0.000	0.043
Intestinimonas	5.9867	0.1123	0.41185265	Coprobacter	0.254	0.399	0.109	0.023	0.874	0.000
Lachnospiraceae_incertae_sedis	6.4906	0.09003	0.41185265	Enterococcus	0.270	0.056	0.141	0.002	0.004	0.489
Oscillibacter	6.6756	0.08299	0.41185265	Akkermansia	0.415	0.220	0.183	0.036	0.630	0.001
Prevotella	5.8978	0.1167	0.41185265							
Veillonella	6.3553	0.09555	0.41185265							
Victivallis	6.3163	0.0972	0.41185265							
Clostridium_IV	5.2986	0.1512	0.50397086							
Romboutsia	5.1385	0.1619	0.51135872							
Butyrivibrio	4.8212	0.1854	0.556104							
Clostridium_XIVb	4.3795	0.2233	0.63797444							
Bifidobacterium	3.63	0.3043	0.77794008							
Clostridium_sensu_stricto	3.5749	0.3112	0.77794008							
Slackia	3.6759	0.2987	0.77794008							
Bilophila	3.2824	0.3501	0.77800065							
Paraprevotella	3.3714	0.3378	0.77800065							
Parasutterella	3.3508	0.3406	0.77800065							
Faecalicoccus	3.1427	0.3701	0.79316966							
Dorea	3.0351	0.3862	0.79911351							
Fusicatenibacter	2.9072	0.4062	0.8070867							
Parabacteroides	2.7581	0.4304	0.8070867							
Phascolarctobacterium	2.7779	0.4272	0.8070867							
Actinomyces	2.0854	0.5549	0.81065646							
Bacteroides	1.9885	0.5748	0.81065646							
Blautia	2.4777	0.4793	0.81065646							
Butyrivibrio	2.0518	0.5617	0.81065646							
Desulfovibrio	2.2451	0.5231	0.81065646							
Dialister	2.2836	0.5157	0.81065646							
Gemmiger	2.4881	0.4774	0.81065646							
Gordonibacter	1.895	0.5945	0.81065646							
Odoribacter	2.0744	0.5571	0.81065646							
Roseburia	2.3987	0.4939	0.81065646							
Senegalimassilia	2.6365	0.4511	0.81065646							
Streptococcus	1.9562	0.5815	0.81065646							
Anaerostipes	1.6709	0.6434	0.81247994							
Clostridium_XVIII	1.8068	0.6135	0.81247994							
Lactobacillus	1.7452	0.6269	0.81247994							
Ruminococcus	1.6416	0.65	0.81247994							
Barnesiella	1.4647	0.6904	0.84542768							
Clostridium_XIVa	1.232	0.7453	0.84827889							
Eggerthella	1.3205	0.7243	0.84827889							
Faecalibacterium	1.1714	0.7599	0.84827889							
Holdemania	1.1565	0.7635	0.84827889							
Lactococcus	1.2952	0.7303	0.84827889							
Collinsella	0.84468	0.8388	0.89275976							
Olsenella	0.85694	0.8358	0.89275976							
Ruminococcus2	0.80562	0.8481	0.89275976							
Holdemania	0.31254	0.9577	0.97388413							
Sutterella	0.35669	0.949	0.97388413							
Sporobacter	0.1711	0.9821	0.98211356							



**Table S10 -- Co-occurrence analysis network**

Associated genera	Edge	Cooccurrence_method	interactionType	pval	qval
Cloacibacillus	OTU-ST04->OTU-304	correl_spearman;dist_kullbackleibler	copresence	1.8E-11	2.4E-09
Akkermansia	OTU-23->OTU-ST04	correl_spearman;dist_kullbackleibler	copresence	4.2E-05	1.4E-03
Coprobacter	OTU-ST04->OTU-161	correl_spearman;dist_kullbackleibler	copresence	2.1E-03	4.7E-02
Akkermansia	OTU-23->OTU-ST03	correl_spearman;dist_kullbackleibler	mutualExclusion	2.9E-07	1.9E-05
Methanobrevibacter	OTU-ST04->OTU-78	correl_spearman;dist_kullbackleibler	copresence	3.9E-06	1.8E-04
Anaerotruncus	OTU-ST04->OTU-290	correl_spearman;dist_kullbackleibler	copresence	1.8E-03	4.7E-02

Table S11 -- Distance based redundancy analysis (dbRDA) results

CAP Variable	stepRDA	Cumulative R2	group	CAP r2	FDR	exclusion from plot
Status: ST1, ST2, ST3, ST4 and non-carriers	1	0.031346	Blastocystis	0.03542608	0.00475	included
Bristol stool score	2	0.040972	Bowel	0.008065	0.00475	included
A06AD15_65 (osmotic laxatives)	3	0.050676	Medication	0.00757658	0.00475	included
Age	4	0.057972	Anthropometrics	0.00797695	0.00475	included
Relative abundance of Blastocystis	5	0.064948	Blastocystis	0.02241826	0.00475	included
J01CA04 (amoxicillin)	6	0.071039	Medication	0.00851593	0.00475	included
Bread type preference	7	0.076857	Dietary Information	0.00923985	0.00475	included
Magnesium supplement	8	0.082257	Dietary Information	0.00341009	0.02955556	included
HDL cholesterol	9	0.087075	Blood Parameters	0.00369955	0.006608696	included
Time since previous relief	10	0.091107	Bowel	0.00621267	0.00475	included
R06AX28 (rupatadine)	11	0.094557	Medication	0.00210583	0.057	included
Fruits (freq. past week)	12	0.097408	Dietary Information	0.00396126	0.006608696	included
Shift work	13	0.099614	Lifestyle	0.00225304	0.038	included
Gastroenteritis (other treatment)	14	0.101915	Health	0.0030723	0.018424242	included
Height	15	0.104136	Anthropometrics	0.00407335	0.006608696	included
Red blood cell counts	16	0.106342	Blood Parameters	0.00850329	0.00475	included
Beer consumption	17	0.108651	Dietary Information	0.00386302	0.010857143	included
Carriers or non-carriers		0.108651	Blastocystis	0.03458764	0.00475	included
Presence of ST4		0.108651	Blastocystis	0.01228901	0.00475	included
Glomerular filtration rate		0.108651	Health	0.0076032	0.00475	included
Presence of ST3		0.108651	Blastocystis	0.00705173	0.00475	included
Hemoglobin		0.108651	Blood Parameters	0.00546469	0.00475	included
Gender		0.108651	Anthropometrics	0.0047907	0.006608696	included
Presence of ST1		0.108651	Blastocystis	0.00430263	0.00475	included
Presence of ST2		0.108651	Blastocystis	0.00422679	0.006608696	included
L04AX01 (azathioprine)		0.108651	Medication	0.00406195	0.010857143	included
J01CR02 (amoxicillin and enzyme inhibitor)		0.108651	Medication	0.00405839	0.00475	included
L04AB (TNF-alpha inhibitors)		0.108651	Medication	0.0040441	0.010857143	included
Triglycerides		0.108651	Blood Parameters	0.00364601	0.006608696	included
Coffee (days since consumed)		0.108651	Dietary Information	0.00350997	0.0095	included
Sleep (hours/day)		0.108651	Lifestyle	0.00346966	0.006608696	included
BetaGlobulines.g.L		0.108651	Blood Parameters	0.00333459	0.015724138	included
Chocolate type preference		0.108651	Dietary Information	0.00318647	0.010857143	excluded
Ulcerative colitis		0.108651	Health	0.00306962	0.018424242	excluded
Number of breakfasts past week		0.108651	Dietary Information	0.00301261	0.01716129	excluded
Uric acid		0.108651	Blood Parameters	0.00287959	0.020117647	excluded
BMI		0.108651	Anthropometrics	0.00287249	0.01716129	excluded
N03AE01 (clonazepam)		0.108651	Medication	0.00259617	0.023885714	excluded
HOMA-IR		0.108651	Blood Parameters	0.00246837	0.057422222	excluded
Fruits (days since consumed)		0.108651	Dietary Information	0.00224397	0.034918919	excluded
Gamma-glutamyltransferase		0.108651	Blood Parameters	0.00200026	0.057	excluded
Smoker		0.108651	Lifestyle	0.00198467	0.048195122	excluded
A07EC02 (mesalazine)		0.108651	Medication	0.00187954	0.064434783	excluded
Spirits consumption		0.108651	Dietary Information	0.00183624	0.0456	excluded
Cigarettes (average/day past 2 months)		0.108651	Lifestyle	0.00179798	0.054285714	excluded
Smoking with inhalation		0.108651	Lifestyle	0.00173022	0.042871795	excluded
Q10 supplement		0.108651	Dietary Information	0.00172219	0.077490196	excluded
Creatinine		0.108651	Blood Parameters	0.00166504	0.069531915	excluded
Alcohol (average past week)		0.108651	Dietary Information	0.00160862	0.06992	excluded
Irritable bowel syndrome		0.108651	Health	0.00156206	0.069666667	excluded
Alfa-2-globuline		0.108651	Blood Parameters	0.00150732	0.084603774	excluded
G03AA12 (drospirenone and ethinylestradiol)		0.108651	Medication	0.00146153	0.095345455	excluded
Work days per week		0.108651	Lifestyle	0.00143158	0.06992	excluded
Blood urea		0.108651	Blood Parameters	0.00141385	0.078923077	excluded
Bloating (days past week)		0.108651	Bowel	0.00127067	0.094296296	excluded
Alanine Aminotransferase			Blood Parameters	0.0012117	0.114	excluded
N06AX16 (venlafaxine)			Medication	0.00114837	0.114	excluded
Chronic fatigue syndrome			Health	0.00120843	0.114	excluded
Working hours (hours/week)			Lifestyle	0.00109864	0.124949153	excluded
Food allergy			Health	0.00098884	0.130466667	excluded
Constipation (days past week)			Bowel	0.00093254	0.163213115	excluded
Hip circumference			Anthropometrics	0.00080475	0.165483871	excluded
Meat (average past week)			Dietary Information	0.00078915	0.173714286	excluded
Soda (freq. past week)			Dietary Information	0.00073084	0.1983125	excluded
G03DA04 (progesterone)			Medication	0.00063511	0.220984615	excluded
J01XE01 (nitrofurantoin)			Medication	0.00049392	0.228	excluded
Weight perception			Anthropometrics	0.00045762	0.253333333	excluded
Plan to change weight			Anthropometrics	0.00047551	0.253333333	excluded
Soy products (average past week)			Dietary Information	0.00054507	0.253333333	excluded
Sick leave (freq. past week)			Lifestyle	0.00036343	0.287714286	excluded
G03CA04_CC06 (estrogens)			Medication	0.0002256	0.317915493	excluded
Mean corpuscular hemoglobin concentration			Blood Parameters	0.00022409	0.329333333	excluded
Childhood residence type			Lifestyle	9.90E-05	0.386246575	excluded
Creatine Kinase			Blood Parameters	-5.94E-05	0.461135135	excluded
Pets (past 3 months)			Lifestyle	-0.0004384	0.664746667	excluded
Working start age			Lifestyle	-0.0012125	0.995	excluded