Supplementary material

Data source

All analyses were performed on previously published data. 16S rRNA gene abundance data from 2553 samples, comprising 1266 oral, 187 stool, 836 skin and 264 vaginal, originating from the Human Microbiome Project (HMP)[1] were downloaded from their web resource (http://www.hmpdacc.org/). Of these, 172 had less than 1000 reads and were discarded from future analysis. An additional 139 stool metagenomic shotgun sequenced samples were downloaded from the same location. Additional metagenomic shotgun sequencing data originate from samples (368 Chinese samples and 278 samples from the MetaHIT project) described in Qin et al. (2012) and Le Chatelier et al. (2014), respectively. Three additional samples from the US were used, which are described in Schloissnig et al. (2013).

Data, together with code for generating the main figures can be found at: https://hub.docker.com/r/costeapaul/enterotype_figures/. Instructions for pulling and running the docker can also be found there.

Taxonomic and functional analysis

For 16S rRNA gene-based taxonomic composition analysis, we used operational taxonomic unit (OTU) or genus level relative abundances. Genus level abundance matrices were calculated by adding relative abundances of all taxonomically annotated OTUs. OTUs not annotated at genus level are considered "unclassified" and their relative abundances were agglomerated into that category.

For cross-study analyses, shotgun sequencing reads were mapped to a database of selected single copy phylogenetic marker genes (mOTU.v1.padded)[2] and summarized into species-level (mOTU) and genus-level relative abundances. Functional profiles of clusters of orthologous groups (COGs) and KEGG orthology groups (KO), including both those of eukaryotic and bacterial origin, for MetaHIT, Chinese, and HMP samples were computed using MOCAT [3] by mapping shotgun sequencing reads to an annotated reference gene catalogue as described in Voigt et al. [4]. COG category abundances were calculated by summing the abundance of the respective COGs belonging to each category per sample, excluding NOGs.

Weighted and unweighted UniFrac distances were downloaded from the HMP web resource (*http://www.hmpdacc.org/*). Jensen-Shannon distances were computed on genus level relative abundance matrices, as described in the enterotyping tutorial (*http://enterotype.embl.de/enterotypes.html*).

For determining the optimal number of clusters on all data matrices, three different measures were used from the R fpc package[5] (version 2.1.9). The Calinski-Harabasz index[6] and the silhouette index[7] of a given distance matrix and a set number of clusters were computed using the function *pamk* with default parameters. The prediction strength[8] was calculated with a modified version of the *prediction.strength* function, allowing a distance matrix as a parameter, with the dataset being randomized 50 times. For all

the different measures, we varied the number of clusters between two and ten and considered the cluster number with the highest value for each measure to be the optimal one.

Ordination

Visualization of distance matrices was performed using unsupervised ordination methods. Principal coordinate analysis was performed with the R ade4 package (version 1.6.2), using the *dudi.pco* function.

Parallel to each pair of enterotypes, we plotted the abundances of selected genera and functional categories (Fig. 3). For each combination, we performed principal component analysis on the 2-dimensional PCoA coordinates, to identify the axis that explains the greatest variation between enterotypes (i.e. the first eigenvector). This component forms the x-axis for the distributional plots. Subsequently, we log transformed abundance of genera/ functions, then scaled and centered them. The plotted line is a smoothed spline fitted to these transformed abundances, using the R base function *smooth.spline.* To test for significant bimodal distribution of feature abundance, Hartigan's dip test statistic from R package *diptest* was used.

Feature significance testing

Univariate testing for differential abundances of taxonomic and functional features between two or more groups was tested using a Wilcoxon-Rank Sum test or Kruskal-Wallis test (p-value), respectively, corrected for multiple testing using the Benjamini-Hochberg false discovery rate (q- value). COGs and KOs occurring in less than four samples or with an average normalized count abundance < 30 across were excluded from the univariate analysis.

Diversity analysis

Richness and Shannon diversity index for taxonomic and functional features, mOTU and OTU were calculated after rarefaction of matrices to 3,000 and 5,000, units per sample, respectively, using the *vegan* R package. Rarefactions of COG, KO and gene matrices were done using rtk[9] with a per sample rarefaction depth of 6,900,000. In total we performed 30 repetitions, in each of which we measured the richness and Shannon diversity metrics within a rarefaction. The median value of these was taken as the respective richness/ diversity measurement for each sample. These thresholds were chosen to include most samples. Diversity differences were plotted as boxplots (Suppl. Fig. 7 and 8), with boxes defined as 25^{th} and 75^{th} quantiles, and whiskers by 1.5 * interquartile range.

Parameter-dependence of clustering

Clustering algorithms have been developed and employed for nearly 100 years (e.g. Driver and Kroeber 1932) and have more recently been applied to analyze microbial compositions, especially those of the human gut. However, it is challenging to determine whether there are actual clusters present, and if so, how many? A number of clustering optimality measures, as well as distance measures were employed for determining the number of microbial clusters that may be present in the human gut microbiota. To describe inter-sample differences, most studies use a combination of weighted and unweighted UniFrac distances or Jensen-Shannon distance (JSD) (Suppl. Table 1). To determine the optimal number of clusters in the space described by the distance measure, the CH-index[6], silhouette index[7] and prediction strength[8] are commonly used.

Different distance metrics will give a different weight to community features. For example, the UniFrac distance[11], be it weighted (taking abundance into account) or unweighted (presence-absence only), is based on the importance of phylogenetic distance between the components of the community. In contrast, the JSD distance does not take phylogenetic information into account, and measures the mutual information shared between two samples. For both weighted UniFrac and Jensen-Shannon, the underlying hypothesis is that variation in highly abundant members is the most relevant feature for describing similarities. The hypothesis of unweighted analysis is that community membership is the most important feature. While these distances are conceptually very distinct, they may result in the same outcome, though they exhibit quantitatively and qualitatively different properties. Another property worth considering is the absolute numbers of microbes in any given sample; it may be that observed fluctuations in composition poorly reflect the actual cell counts of the members, as the total amount varies considerably. This may constitute a further confounder when trying to disentangle compositional properties.

Within the Human Microbiome Project (HMP) dataset, 2910 samples are available from a range of human body sites[1]. We used this dataset to benchmark the aforementioned distances and optimality criteria, based on the assumption that human body-sites are inhabited by different microbial populations and that their separation should be clear (Fig. 1). The PCoA projections of the distance space into two dimensions shows that the largest part of the variation does not separate the body sites properly, except in the case of the JSD distance on genus level (Suppl. Fig. 4A). This metric and weighted UniFrac both recovered the four expected clusters in conjunction with PS or Silhouette index (Suppl. Fig. 4B). However, even when recovered, the separation appears not to be very strong, with silhouette values being low (0.4 at best). Since often three or less clusters are chosen to be the optimal cluster number, we conclude that the clustering approach may underpowered.

Cross-study enterotype comparison

Comparability of the structure across multiple datasets is a necessary characteristic of the enterotype concept. However, although similar genera were reported as being most abundant in gut stratifications (Suppl. Table 1), this does not automatically imply similar communities or structure. To test the assumption of comparability, we used three unrelated large datasets, with different sampling procedures (US HMP[1], Chinese diabetes type 2 study[12], European MetaHIT consortium[13]) and clustered these with the PAM clustering algorithm on a JSD distance at genus level[14]. The obtained clusters had an overrepresentation of *Prevotella*, *Bacteroides* or Firmicutes (the latter represented by *Ruminococcus*, *Eubacterium* and *Subdilogranulum* respectively), as expected.

Although we do not exclude alternative scenarios (see Fig. 2), we first trained a LASSO logistic regression classifier[15] to recover the three enterotypes within the MetaHIT samples. This was then used to classify samples from the other two studies. The respective ROC-AUC was high (Suppl. Fig. 11), meaning that the classifier and unsupervised clustering mostly assigned the same cluster memberships to samples. One difference is that the classifier can be used on any arbitrarily small dataset. This approach could also be expanded to classify single samples based on other machine learning techniques, e.g. a trained DMM model.

Furthermore, if enterotypes reflect community compositions and not just differences in the driver species, i.e. are reflecting different ecological networks, we expect the classification to remain pertinent after removing *Bacteroides* and *Prevotella* from the data. Indeed, although with lower accuracy than when including these two taxa, the classification still captures the initial enterotypes in all datasets.

Further, using the abundance of gene families within each respective enterotype, the prediction of enterotype state is even stronger in cross validation than when using taxa abundances (Suppl. Fig 11). For this analysis, we used only commonly represented functional categories (i.e., COG's that have representative genes in at least five of the 50 most abundant genera), ensuring that the classifier does not exploit functional categories which are restricted to taxonomic subgroups.

Determining if samples are within enterotype space

Using the HMP dataset[1], we compute the distance between all stool samples using a genus summarized OTU table. This allows us to define the expected distance distribution of stool samples. For any novel sample, we compute the distance to all stool samples in the HMP data and consider it to be in the enterotyping space if its average distance is within one standard deviation of the stool distance distribution. Using this approach, we correctly identify western-like stool sample and reject all other body-site samples as not being in the enterotyping space. Furthermore, we also correctly classify infant samples as being outside the enterotyping space (data not shown).

131

132 **Supplementary Table 1:** Microbial community studies researching the presence of enterotypes (ET).
133 Abbreviations: F=ET F (Firmicutes enriched), B=ET B (Bacteroides enriched), P=ET P (Prevotella enriched),
134 CH= Calinski−Harabasz pseudo F−statistic, SIL= Silhouette internal cluster optimality criterion.

Study	Year	Technology	ET reported	Optimal Cluster number	Notes
[14]	2011	454 rRNA, illumina WGS, Sanger WGS	B, F, P	CH (3)	First study to show ET's
[16]	2011	Sanger rRNA	B, F, P	visual	clone library
[17]	2011	454 rRNA	(F+B), P	CH (3), SIL (2)	Diet relation to ET's
[18]	2012	454 rRNA	B, F, P	[19] (3), SIL (2)	Species network based ET identification
[20]	2012	454 rRNA	P, (F+B), *Bifidobacteria*	SIL (2)	Includes children that form a separate cluster
[21]	2012	454 rRNA	gradient	visual	Analysis not based on clustering, HMP
[22]	2013	454 rRNA, illumina WGS	various	SIL(2), CH(3) (rDNA); 2 (WGS)	Extensive testing of methodology, HMP
[23,24]	2012		P, B, *Ruminococcus* , *Oscillibacter, Alistipes, Odoribacter*	CAGs (6), SIL(2), CH(2)	co-abundance groups
[12]	2012	illumina WGS	B, F, P	SIL (3)	
[25]	2012	illumina sg	B, F, P	CH (3), SIL (2)	Atherosclerosis associated to ET F
[26]	2012	Sanger rRNA	Similar to F, P, B, F2	Dirichlet Multinomial Mixtures (4)	
[27]	2012	illumina rRNA	B, F, P	CH (3), SIL(2/3)	Chimpanzee
[28]	2013	454 rRNA	B, F, P	[26] (3)	Association of ET P to CD risk allele
[29]	2013	454 rRNA	B, F, P	Based on composition	obesity and NASH in adolescents
[30]	2013	illumina rRNA	F, B	CH (2), SIL (2)	Mouse; ET B shows links to inflammation
[31]	2013	454 rRNA	B, F, P	Based on composition	Time series

[32]	2014	454 rRNA	Similar to B, F, P, F2	Dirichlet Multinomial Mixtures (4)	HMP reanalysis
[33]	2014	454 rRNA	B, F, P	Complete linkage, Bray-Curtis clustering, SIL(3)	HMP reanalysis
[34]	2014	qPCR	B, P	Prevotella to Bacteroides ratio	Time series on Food trials
[35]	2014	454 rRNA	B, P	Weighted Unifrac SIL(2)	
[36]	2014	454 rRNA	B, Robinsella (Firmicutes dominated)	CH (2)	Wild mice; predictable ET switch after capture
[37]	2014	illumina WGS	B, F, P	CH(3), SIL(3)	4 datasets combined [1,12,38,39]
[40]	2015	illumina rRNA	Similar to F, P	CH(2)	Gorilla, no association to SIV infection
[41]	2015	454 rRNA	F, P	CH(2), SIL(2)	Swine, juvenile development into adult enterotypes
[42]	2016	illumina rRNA	F,B,P	JSD clustering (2,3), DMM(4)	3984 Samples from US and Europe

135

136

Supplementary Table 2: Percentage of CAZY enzymes annotated within 8 substrate categories on a selected subset of gut specific bacterial genomes as published in[43]. Bacteroides contains 15 genomes of genus Bacteroides, Firmicutes are 104 genomes of phylum Firmicutes and Prevotella contains 3 genomes of genus Prevotella. Note that due to multiple substrate specificities, percentage do not add up to 100%.

CAZY category	Bacteroidetes	Firmicutes	Prevotella
Plant.Cell.Wall.Carbohydrates	50%	35%	42%
Chitin	0%	0%	0%
Alpha.glucans	5%	20%	10%
Animal.Carbohydrates	50%	28%	35%
Bacterial.Cell.Wall.Carbohydrates	4%	23%	12%
Fructans	1%	4%	1%
Fungal.Carbohydrates	11%	7%	9%
Dextran	0%	0%	0%

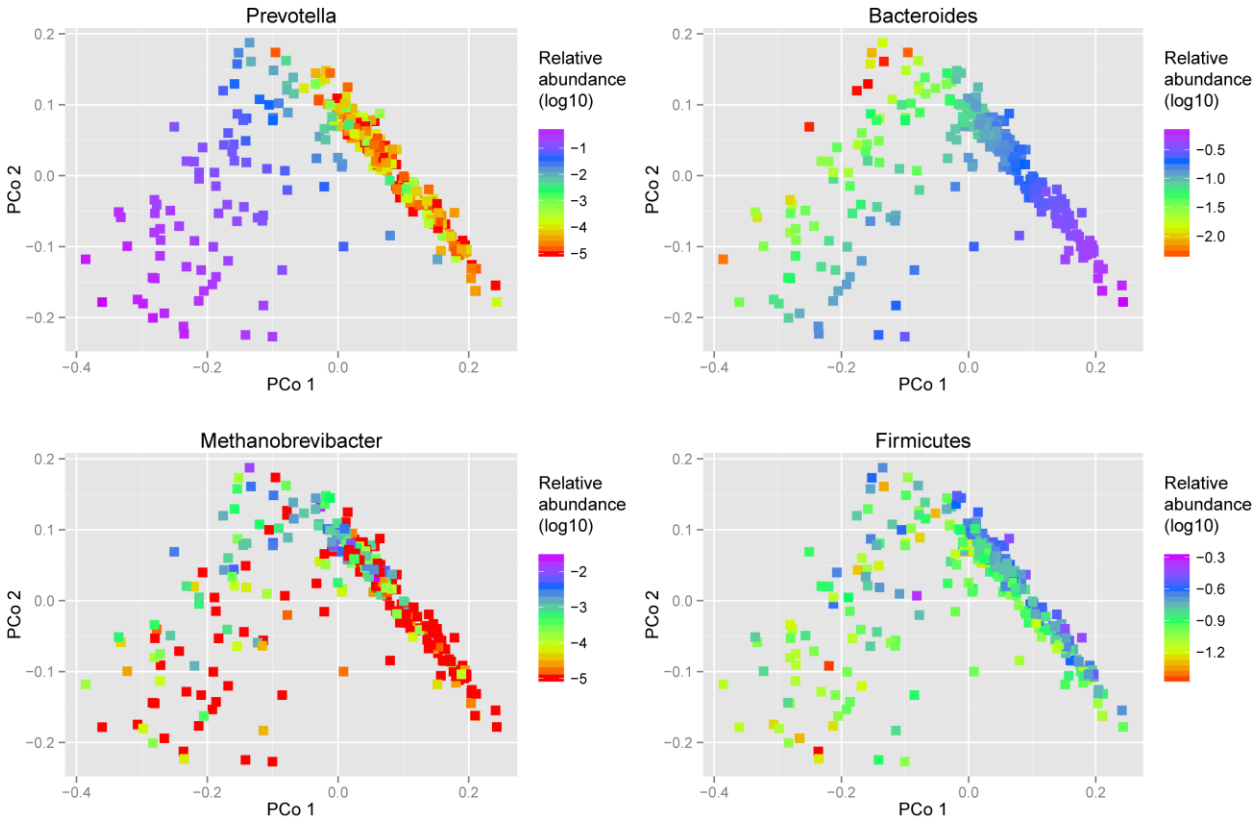
Supplementary Table 3: Functional differences between 3 different enterotype models: 2 Types represents a model comparing ET P against a combined ET F+ ET B, 3 Types compares the three first reported enterotypes (ET B, ET F, ET P) and 4 Types are enterotypes as determined by DMM modelling. Used gene families are derived from COG[44] and KEGG[45] annotations.

Supplementary Table 4: Associations between obesity related parameters reported in[13] and ET state, split for the 2, 3 and 4 clusters.

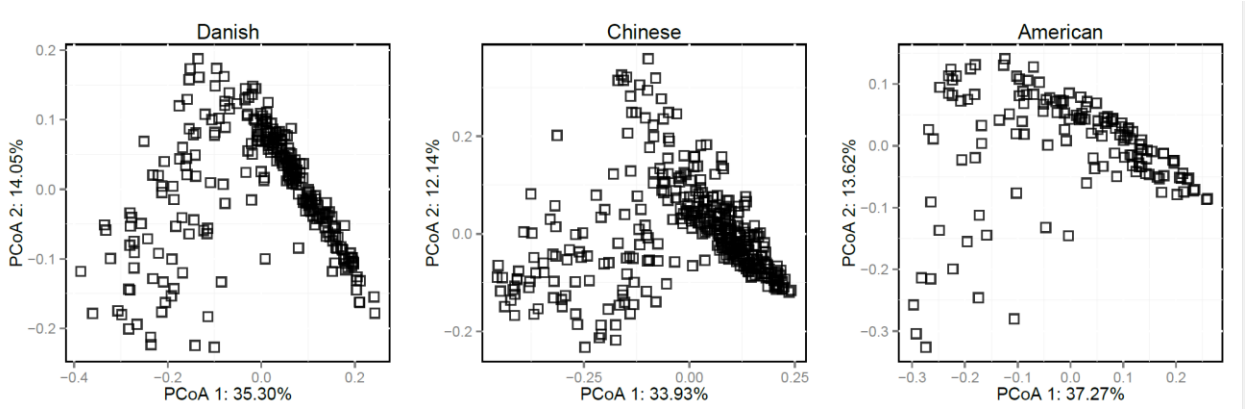
Supplementary Table 5: Studies reporting associations between enterotype drivers and host states

Study	Enriched driver	Phenotype
[25]	Bacteroides	Atherosclerosis
[46,47]	Bacteroides	High-fat diet
[28]	Prevotella	CD risk allele
[48,49]	Prevotella	Colitis susceptible mice
[17,20,50–52]	Prevotella	Fiber-rich diet
[29]	Bacteroides	NASH and ROS
[17,52]	Bacteroides	Protein & animal fat
[53]	Bacteroides	Fibers & fructans
[13,23,30,54,55]	Bacteroides	Low-grade inflammation, CRP and insulin resistance

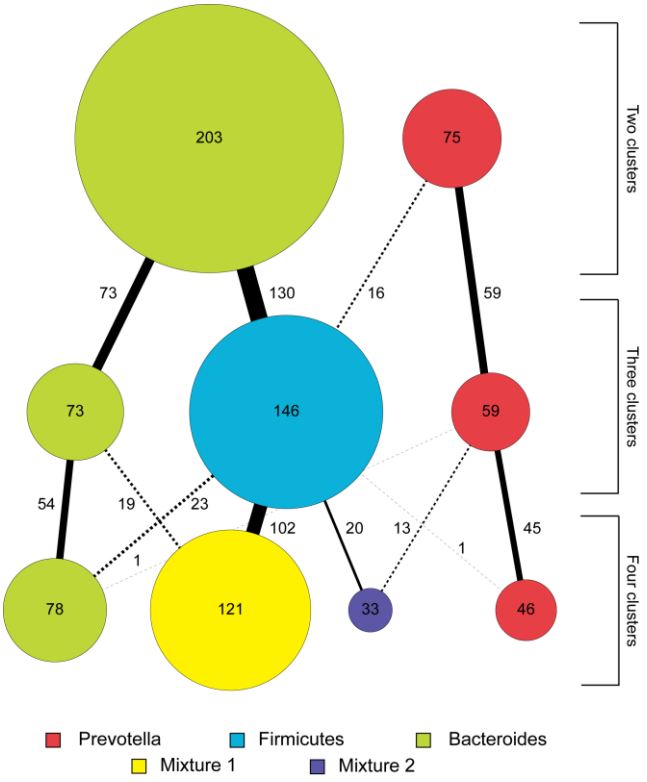
Supplementary Figures



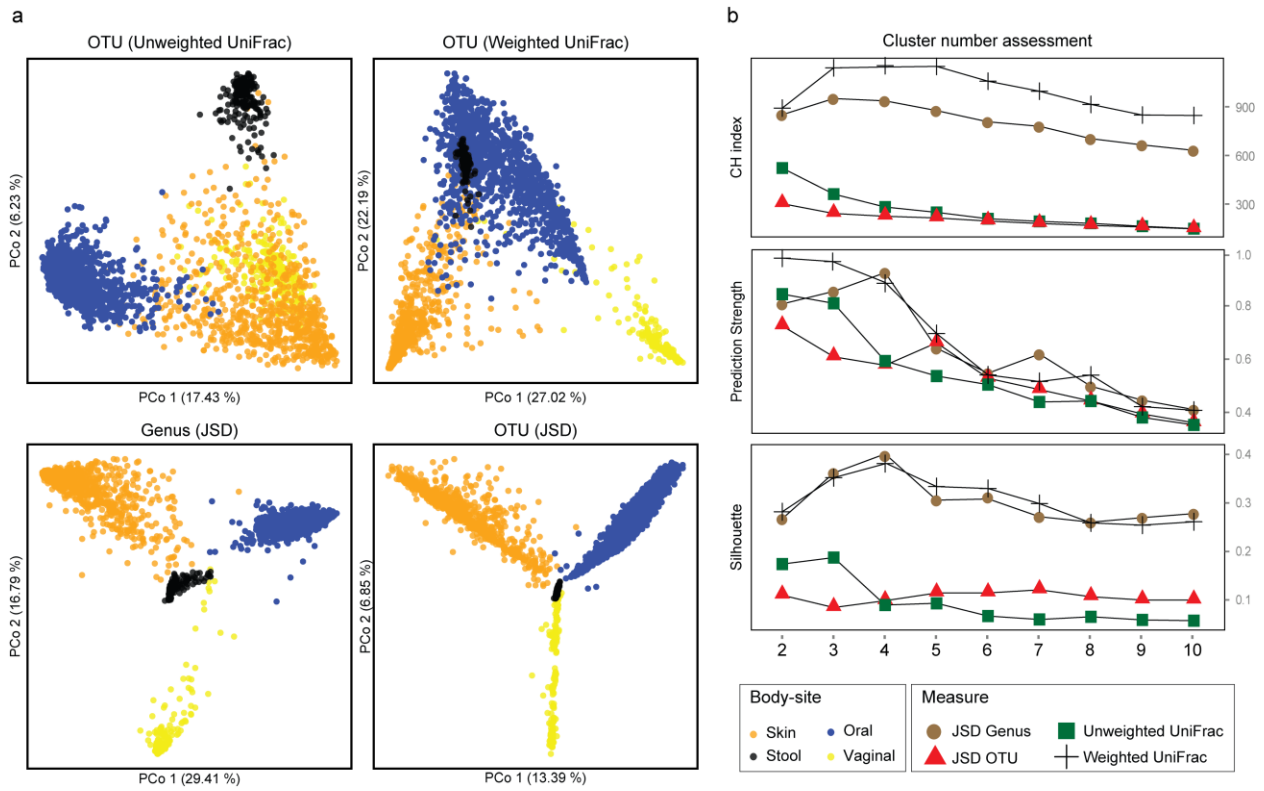
Supplementary Figure 1: Log10 transformed relative abundance of relevant genera, superimposed onto the PCoA ordination of the MetaHIT dataset[13], consisting of 278 samples, showing the bimodal distributions of *Prevotella* and *Methanobrevibacter* and the unimodal distribution of *Bacteroides* and *Firmicutes*, similar to Fig. 2A and 3A where the modality can be more clearly seen based on density distributions.



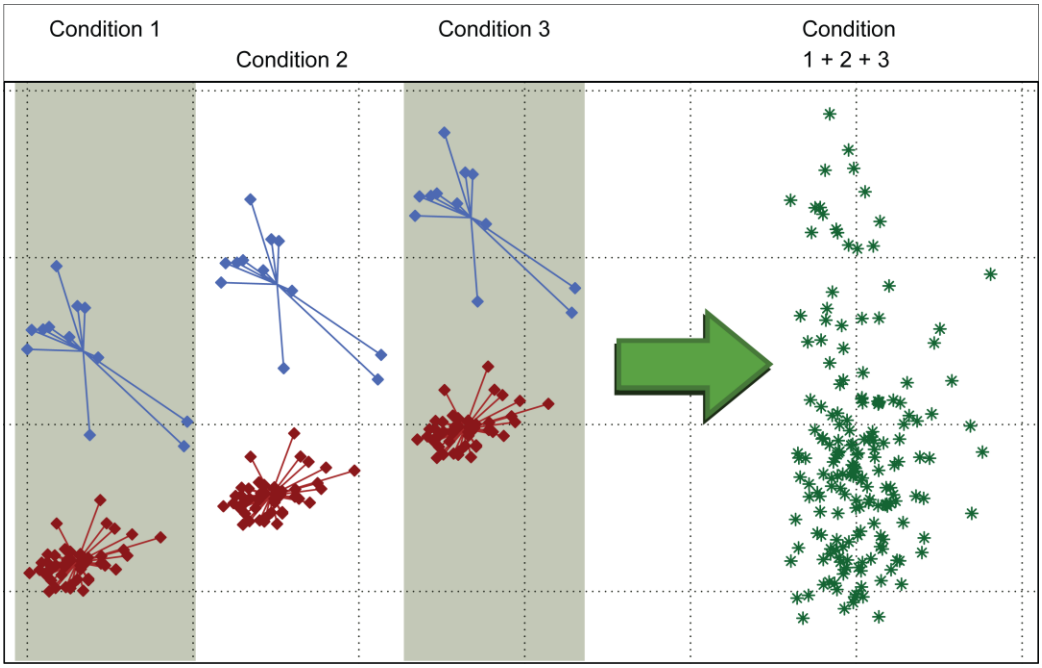
Supplementary Figure 2: PCoA of three datasets (Danish represented by MetaHit[13] (278 samples), Chinese[12] (368 samples) and American by HMP[1] (142 samples)), on a Jensen-Shannon distance computed on the genus abundance profiles of the samples. In all cases, the samples are not randomly distributed throughout the space, with similar higher density regions.



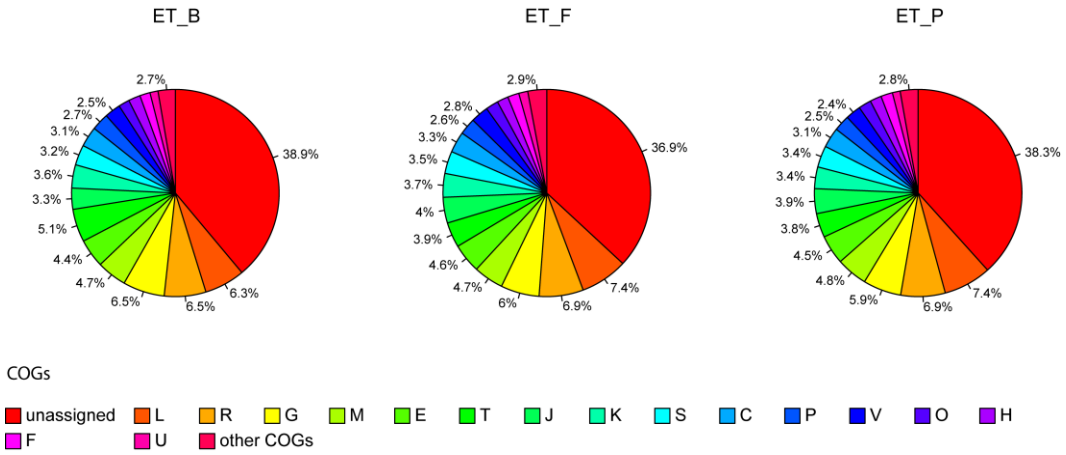
Supplementary Figure 3: Hierarchical structure of different clusterings using 278 MetaHIT[13] samples. Each circle represents a cluster, as obtained by PAMk for 2 and 3 clusters and DMM for 4 clusters. The connecting lines show the number of samples that overlap between the cluster definitions. Overall, the different clusterings are highly associated, forming a hierarchical structure.



Supplementary Figure 4: Clustering of human body sites, based on the genus level abundance of 2381 HMP samples. Body-site separation as in Fig. 1, using frequently used enterotype clustering methods. (A) Ordination of the HMP 16S rRNA (v35) dataset using four common inter-sample distance measures. (B) The optimal cluster number calculated within each distance measure using common clustering optimality measures. Body site separation was recovered by Jensen Shannon divergence (JSD) distance and weighted UniFrac.



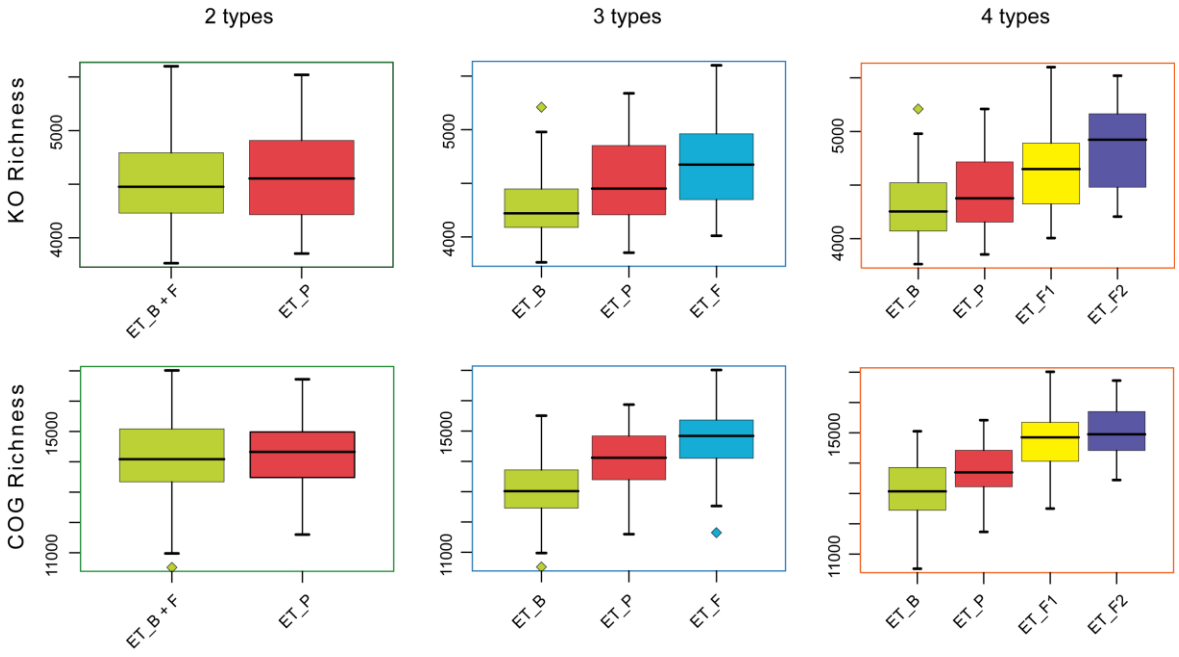
Supplementary Figure 5: Systematic shifts in the microbiota composition under different conditions can obscure clusters in the data. In this theoretical example, the discrete mouse enterotypes F and B, represented as red and blue data points, respectively (60 samples; original data from publication[30]). These were linearly shifted on the y-axis, representing three hypothetical confounders acting on the microbiome. When combining these three conditions (green points), the discrete clustering is no longer observable. Such a systematic confounder that shifts all samples within each of the three exemplarily conditions, could for example be the immune system or different diets, as well as technical biases, such as different DNA extraction methods. For example, in the mouse microbiota[30] a discrete separation between ET F and ET B is possible. Due to the design of this study, more factors were controlled for than possible in human studies, like diet and environment. Thus, this figure illustrates how confounders may impact the observed composition space and the clustering of samples therein.



197

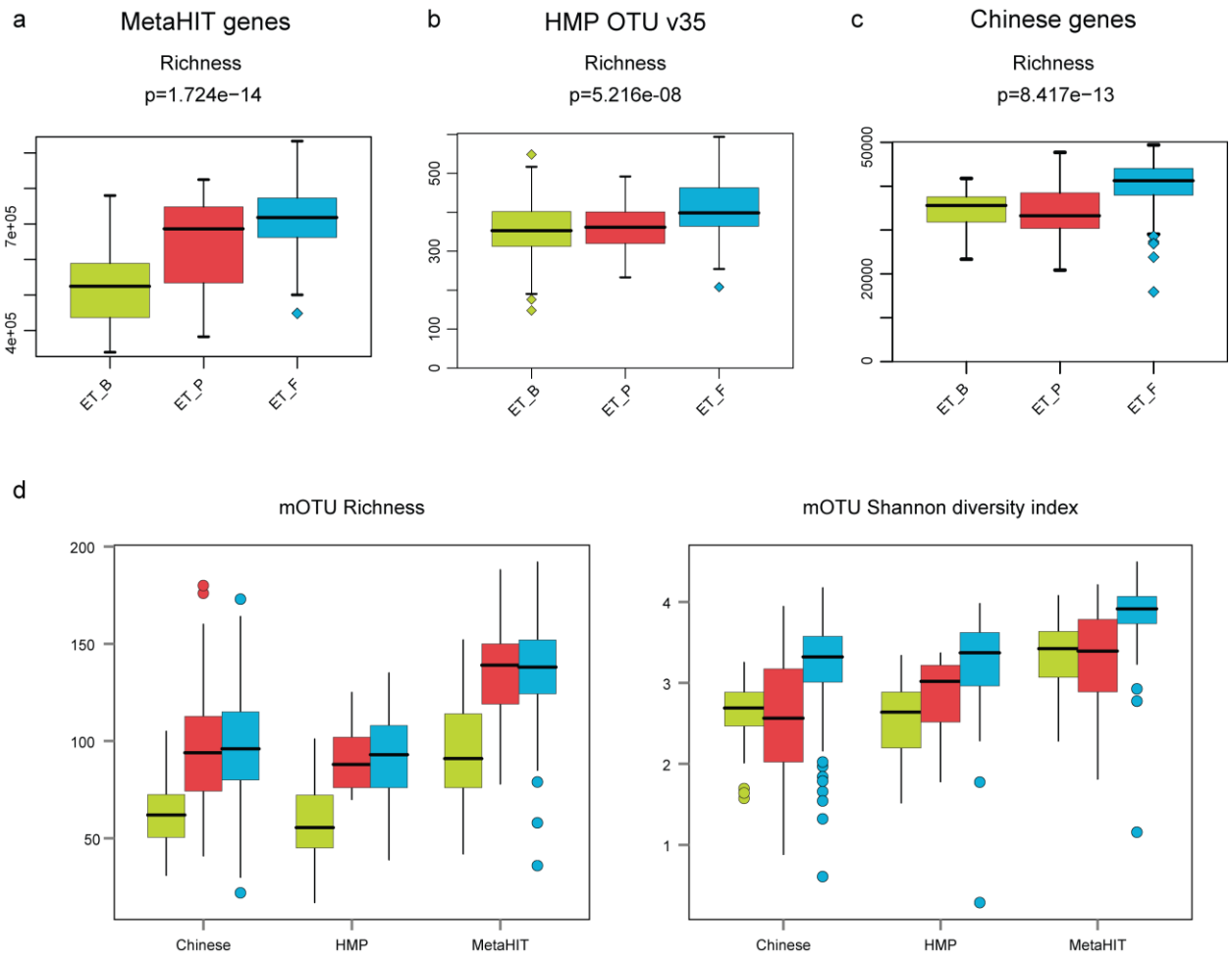
198 **Supplementary Figure 6:** Although 23/25 eggNOG categories are significantly different between
199 three enterotypes defined on the 278 MetaHIT[13] samples (table S3), the overall composition
200 remains relatively stable between them, due to consistently small effect size differences
201 between enterotypes in their functional composition, which is more stable than the taxonomic
202 composition.

203

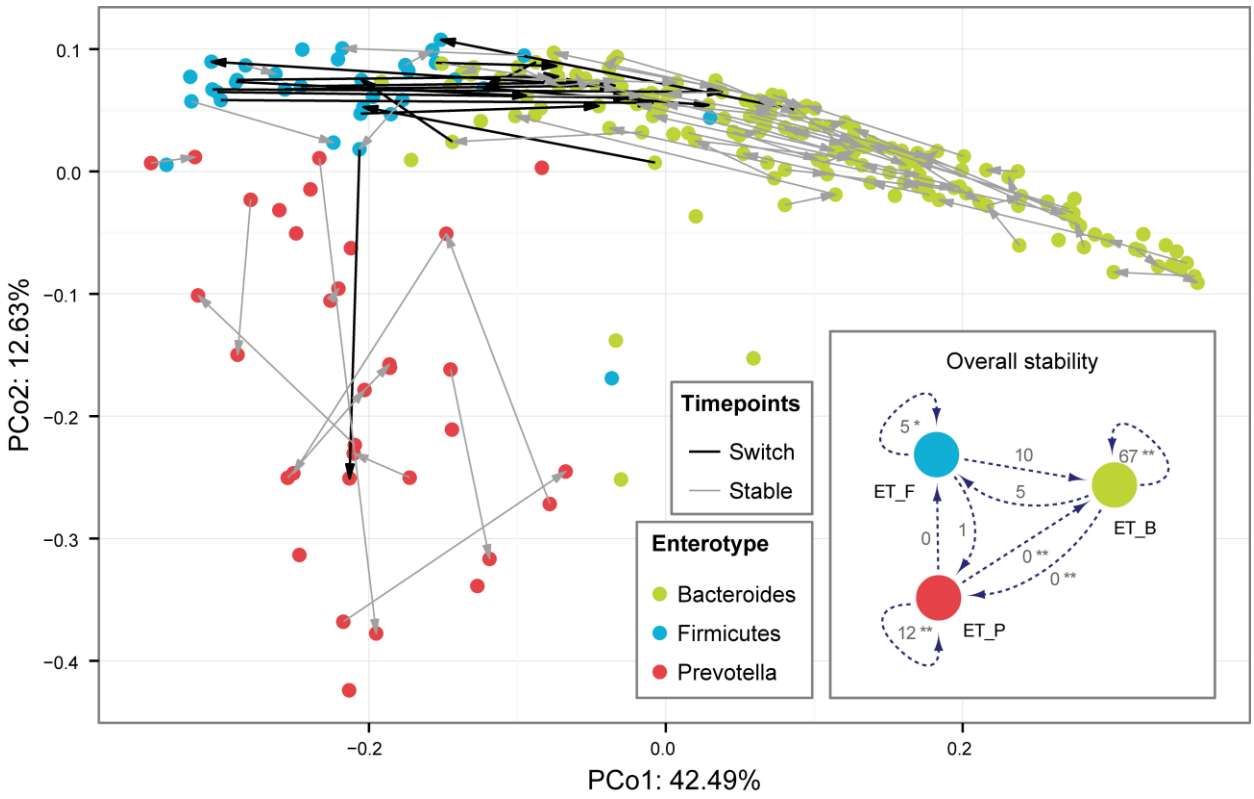


204

205 **Supplementary Figure 7:** Functional richness differs substantially between enterotypes defined
206 on the MetaHIT dataset (278 samples), as measured on COG[44] and KO[45] level data. Boxes
207 represent the 25th to 75th quantiles, and whiskers extend to 1.5x the interquartile range.



Supplementary Figure 8: ET richness differences calculated on (a) 278 MetaHIT samples[13], (b) 142 HMP samples[1] and (b) 368 Chinese samples[12]. Boxes represent the 25th to 75th quantiles, and whiskers extend to 1.5x the interquartile range. High consistency in diversity distribution between enterotypes is visible, despite the underlying data being different community profiling techniques: MetaHIT and Chinese samples are gene richness estimates derived from a gene catalog that encompasses functional as well as taxonomic diversity, while HMP richness was estimated based on v35 16S rRNA gene OTUs, thus only representing taxonomic diversity. Further, using the marker genes based mOTU approach[2] to calculate richness and diversity, similar trends could be derived (d).

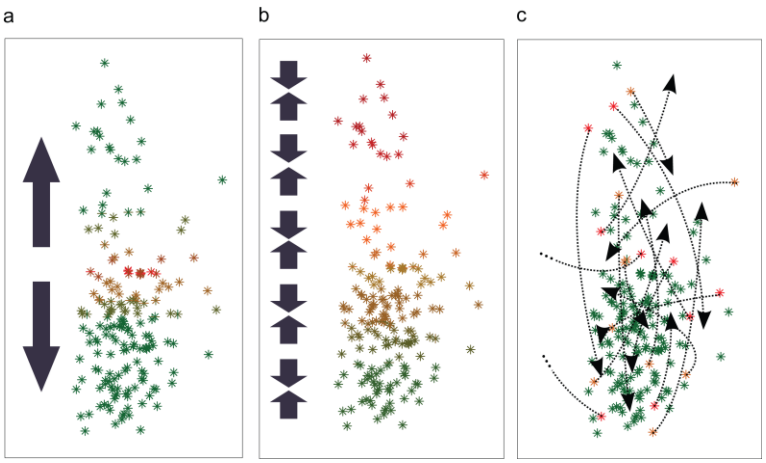


218

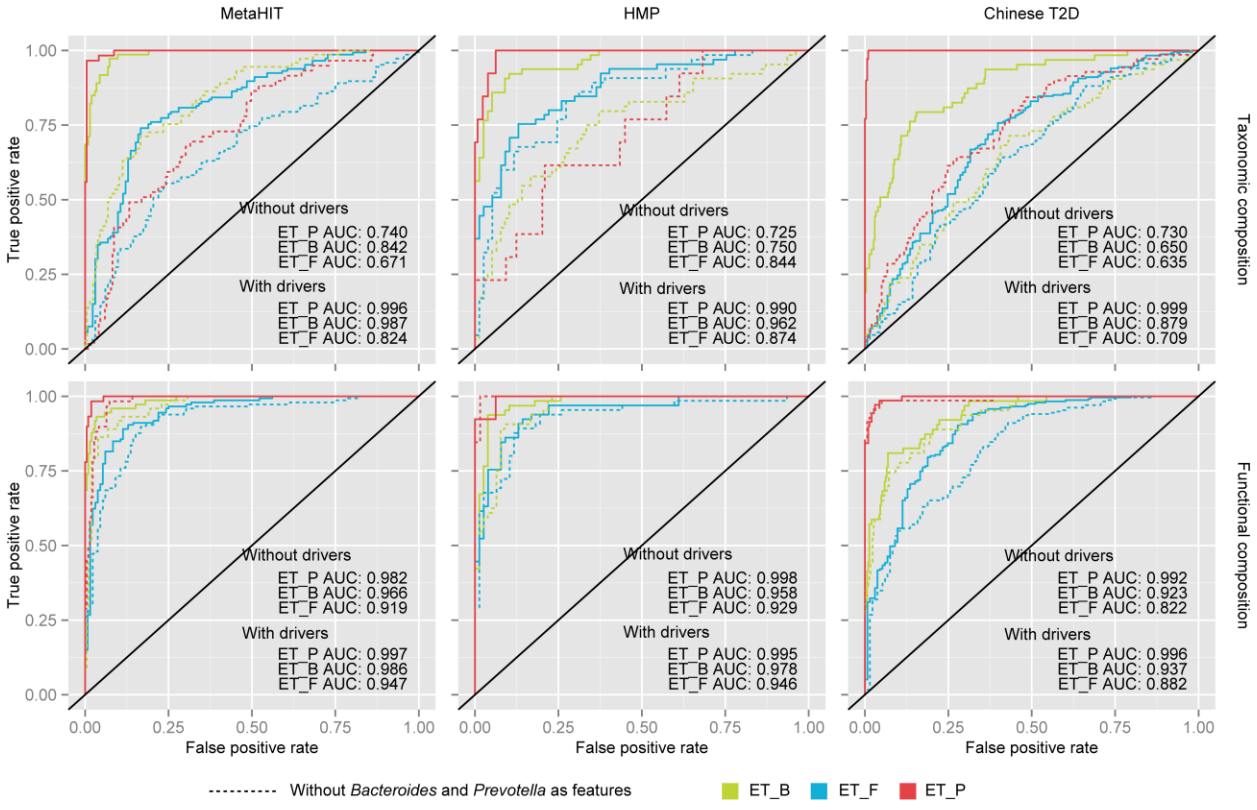
219 **Supplementary Figure 9**: Stability of enterotypes: Considering 100 HMP individuals sampled
220 with metabarcoding (16S) at two different time points, roughly 200 days apart, we illustrate
221 their enterotype stability in PCoA projection. Samples from the same individual are connected
222 through a line, with the arrow pointing to the later sample. Significance of stability is assessed
223 through a permutation test and illustrated in the bottom right panel (* $p < 0.05$, ** $p < 0.01$).
224 The insert "Overall stability" describes the number of samples that were switching between
225 time points or remained within the same enterotype.

226

227



Supplementary Figure 10: Three hypotheses that could account for the observed enterotype gradient in temporal data: (A) the Bacteroides/Firmicutes gradient could be driven by 2 global optimal states or (B) temporal samples are auto correlating to an optimum specified by each microbiota individually, but not driven by global attractors. Last, (C) the Bacteroides/Firmicutes gradient does not reflect an ecological pattern and is subject to strong temporal changes, but this is unlikely to apply to most samples.



Supplementary Figure 11: Robust classification of enterotypes across studies. A three enterotype model classifier was trained on log-ratio transformed genus level abundances of the 278 MetaHIT[13] metagenomic samples in order to address possible compositional effects. This model is shown to also recover enterotypes in 142 HMP samples as well as 368 Chinese samples. The receiver operating characteristic area under the curve (ROC-AUC) for classifier performance on the MetaHIT (internal cross-validation), Chinese and HMP datasets are shown, with the clustering ground truth being estimated using unsupervised clustering of samples in the respective dataset. Although there are known batch effects between these datasets[38], the properties of the enterotypes are comparable and recoverable. Furthermore, the classification is possible even when removing the genera *Bacteroides* and *Prevotella* from the feature set (labelled "Without drivers"). The classification of enterotypes on functional (COG) abundances in almost all cases outperforms the taxonomic classification across all three datasets. In the functional context, "Without drivers" represents a dataset where COGs that contain a gene from either the *Bacteroides* or *Prevotella* genus were removed prior to training and subsequent classification.

Supplementary References

1. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).

2. Sunagawa, S. *et al.* Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* **10,** 1196–1199 (2013).

3. Kultima, J. R. *et al.* MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One* **7,** e47656 (2012).

4. Voigt, A. Y. *et al.* Temporal and technical variability of human gut metagenomes. *Genome Biol.* **16,** 73 (2015).

5. Hennig, C. fpc: Flexible procedures for clustering. (2014).

6. Calinski, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat. - Theory Methods* **3,** 1–27 (1974).

7. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20,** 53–65 (1987).

8. Tibshirani, R. & Walther, G. Cluster Validation by Prediction Strength. *J. Comput. Graph. Stat.* **14,** 511–528 (2005).

9. Saary, P., Forslund, K., Bork, P. & Hildebrand, F. RTK: efficient rarefaction analysis of large datasets. *Bioinformatics* 1–2 (2017). doi:10.1093/bioinformatics/btx206

10. Driver, H. E. & Kroeber, A. L. Quantitative expression of cultural relationships. *Univ. Caltfornia Publ. Am. Archeol. Ethnol.* **31,** 211–256 (1932).

11. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71,** 8228–8235 (2005).

12. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).

13. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500,** 541–546 (2013).

14. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–80 (2011).

15. Tibshirani, R. Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B* **58,** 267–288 (1994).

16. Morotomi, N. *et al.* Evaluation of Intestinal Microbiotas of Healthy Japanese Adults and Effect of Antibiotics Using the 16S Ribosomal RNA Gene Based Clone Library Method. *Biol. Pharm. Bull.* **34,** 1011–20 (2011).

17. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–8 (2011).

18. Zupancic, M. L. *et al.* Analysis of the Gut Microbiota in the Old Order Amish and Its Relation to

292 the Metabolic Syndrome. *PLoS One* **7,** e43052 (2012).

293 19. Zhou, J. *et al.* Functional molecular ecological networks. *MBio* **1,** (2010).

294 20. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–
295 7 (2012).

296 21. Huse, S. M., Ye, Y., Zhou, Y. & Fodor, A. a. A core human microbiome as viewed through 16S rRNA
297 sequence clusters. *PLoS One* **7,** e34242 (2012).

298 22. Koren, O. *et al.* A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial
299 Community Structures in Human Microbiome Datasets. *PLoS Comput. Biol.* **9,** e1002863 (2013).

300 23. Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly.
301 *Nature* **488,** 178–84 (2012).

302 24. Jeffery, I. B., Claesson, M. J., O'Toole, P. W. & Shanahan, F. Categorization of the gut microbiota:
303 enterotypes or gradients? *Nat. Rev. Microbiol.* **10,** 591–592 (2012).

304 25. Karlsson, F. H. *et al.* Symptomatic atherosclerosis is associated with an altered gut metagenome.
305 *Nat. Commun.* **3,** 1245 (2012).

306 26. Holmes, I., Harris, K. & Quince, C. Dirichlet Multinomial Mixtures: Generative Models for
307 Microbial Metagenomics. *PLoS One* **7,** e30126 (2012).

308 27. Moeller, A. H. *et al.* Chimpanzees and humans harbour compositionally similar gut enterotypes.
309 *Nat. Commun.* **3,** 1179 (2012).

310 28. Quince, C. *et al.* The impact of Crohn's disease genes on healthy human gut microbiota: a pilot
311 study. *Gut* **0,** 2012–2014 (2013).

312 29. Zhu, L. *et al.* Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH)
313 patients: A connection between endogenous alcohol and NASH. *Hepatology* **57,** 601–9 (2013).

314 30. Hildebrand, F. *et al.* Inflammation-associated enterotypes, host genotype, cage and inter-
315 individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol.* **14,** R4
316 (2013).

317 31. Mardanov, A. V *et al.* Metagenomic Analysis of the Dynamic Changes in the Gut Microbiome of
318 the Participants of the MARS-500 Experiment, Simulating Long Term Space Flight. *Acta Naturae*
319 **5,** 116–125 (2013).

320 32. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the
321 human body. *Nature* (2014). doi:10.1038/nature13178

322 33. Zhou, Y. *et al.* Exploration of bacterial community classes in major human habitats. *Genome Biol.*
323 **15,** R66 (2014).

324 34. Roager, H. M., Licht, T. R., Poulsen, S. K., Larsen, T. M. & Bahl, M. I. Microbial enterotypes,
325 inferred by the prevotella-to-bacteroides ratio, remained stable during a 6-month randomized
326 controlled diet intervention with the new nordic diet. *Appl. Environ. Microbiol.* **80,** 1142–9
327 (2014).

35. Zhang, J. *et al.* Mongolians core gut microbiota and its correlation with seasonal dietary changes. *Sci. Rep.* **4,** 5001 (2014).

36. Wang, J. *et al.* Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E2703-10 (2014).

37. Karlsson, F. H., Nookaew, I. & Nielsen, J. Metagenomic Data Utilization and Analysis (MEDUSA) and Construction of a Global Gut Microbial Gene Catalogue. *PLoS Comput. Biol.* **10,** e1003706 (2014).

38. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498,** 99–103 (2013).

39. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).

40. Moeller, A. H. *et al.* Stability of the gorilla microbiome despite simian immunodeficiency virus infection. *Mol. Ecol.* **24,** 690–697 (2015).

41. Mach, N. *et al.* Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ. Microbiol. Rep.* **7,** 554–569 (2015).

42. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science (80-.).* **352,** 560–564 (2016).

43. Kaoutari, A. El, Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).

44. Tatusov, R. L., Koonin, E. V & Lipman, D. J. A genomic perspective on protein families. *Science* **278,** 631–7 (1997).

45. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

46. Hildebrandt, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137,** 1716-24–2 (2009).

47. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–4 (2009).

48. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145,** 745–57 (2011).

49. Brinkman, B. M. *et al.* Caspase deficiency alters the murine gut microbiome. *Cell Death Dis.* **2,** e220 (2011).

50. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14691–6 (2010).

51. Ou, J. *et al.* Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *Am. J. Clin. Nutr.* **98,** 111–20 (2013).

52. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* (2013). doi:10.1038/nature12820

53. Sonnenburg, E. D. *et al.* Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141,** 1241–52 (2010).

54. Rath, H. & Herfarth, H. Normal luminal bacteria, especially Bacteroides species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/human beta2 microglobulin transgenic rats. *J. Clin. …* 945–953 (1996).

55. Bloom, S. M. *et al.* Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9,** 390–403 (2011).