Supporting Information

Levy and Borenstein 10.1073/pnas.1300926110

SI Text

Human Oral Community Stringent Growth Comparison. We performed a more stringent analysis of the human oral community, repeating the growth rate comparison described in *Methods* but using our own assessment of growth. Specifically, in this more stringent analysis, we used manually curated data gathered from previous studies (1–6) and compared only growth rates measured as part of the same experiment (i.e., the same experimental conditions and in the same reference) and without conflicting evidence (e.g., if two species grow on pegs but not in flow cells). This more stringent analysis provided qualitatively similar results. Species with better impact on the growth of a partner tend to have lower metabolic competition indices and higher metabolic complementarity indices ($P < 0.048$ and $P < 0.024$, respectively; paired one-sided t test).

We additionally aimed to confirm that *Porphyromonas gingivalis*' metabolic competition and complementarity with other community members represents its ability to form mutualistic biofilms with many species (5). Specifically, we wished to demonstrate that *P. gingivalis* is the most complemented by and poses the least competition to all other species. We first noted that for three of the six target species it poses the lowest metabolic competition (Table S1*B*, Pg column) and in the other three the second lowest; in all six cases it is the most complemented species (Table S1*C*, Pg row). Comparing the indices associated with *P. gingivalis* to those associated with other species, we additionally found that the set of scores denoting the metabolic competition posed by *P. gingivalis* are significantly lower than all other corresponding scores ($P < 0.003$, one-tailed Wilcoxon rank sum test). Similarly, the scores denoting the complementation received by *P. gingivalis* are significantly higher than all other complementarity scores ($P < 1.28 \times 10^{-4}$, one-tailed Wilcoxon rank sum test). Finally, examining all pairwise comparisons of the competition scores posed by *P. gingivalis* (or the complementarity scores received by *P. gingivalis*) to that of other species using a one-sided rank sum test, we found that the median competition associated with *P. gingivalis* is lower in all six comparisons, significantly so in four of six comparisons ($P < 0.01$, one-sided Wilcoxon rank sum test); the complementarity scores associated with *P. gingivalis* are significantly lower in all cases ($P < 0.01$, one-sided Wilcoxon rank sum test). These results are in line with observation of the interaction between *P. gingivalis* and other oral species; although it is not necessarily the preferred growth partner of all species, it can form mutualistic biofilms with many species.

Alternative Co-Occurrence Metrics and Sensitivity to Undersampling. For the results reported in the main text we applied the widely used Jaccard similarity index as a measure of co-occurrence (7). However, because a number of co-occurrence metrics have been used in previous ecological studies (e.g., refs. 8 and 9) and no standardized similarity metric has been fully established, we confirmed here that our main findings are not an artifact of the specific metric used and that the observed patterns hold under several alternative measures of co-occurrence. Specifically, we examined the correlation between our interaction indices and several previously introduced ecological co-occurrence metrics, including in addition to the Jaccard similarity index the Bray–Curtis similarity, the Morisita–Horn similarity, and Cosine similarity. We additionally repeated our analysis using a range-normalized transform of the abundance data: For each species, the abundance was scaled such that the lowest observed abundance value was 0 and the highest was 1. In this way, these metrics are not quantifying the similarity of abundance profiles, but rather the similarity in the changes of abundance across samples. For example, without such normalization, species that have very high abundances in all samples would seem to have similar profiles, even when a rise in the abundance of one is associated with a decrease in the abundance of the other. We find that using any of the above metrics or normalization schemes does not qualitatively change the results reported in the main text (Table S3*D*). We additionally examined Pearson and Spearman coefficients of correlation as a measure of co-occurrence. Using these measures, which are known to attribute spurious associations in relative compositional abundance data (10), resulted in generally similar, but weaker, patterns.

Owing to the limited number of individuals sampled by the MetaHIT study, we further sought to determine whether undersampling of individuals might have any detrimental effects on observed co-occurrence values. We repeatedly subsampled at random 62 individuals (50% of total) uniformly, with no regard to nationality, health state, body mass index, or enterotype. Using these samples, we recalculated the co-occurrence of all species pairs using all metrics described above. We found that the Jaccard similarity index is the most robust, with relatively little variation in obtained co-occurrence values (Fig. S1). We consequently used this co-occurrence metric in the main text.

Alternative Reverse-Ecology Interaction index. In addition to the interaction indices discussed in the main text, we also investigated the association of species co-occurrence with a previously described reverse-ecology interaction measure, the Effective Metabolic Overlap (EMO) score (11). Similarly to the metabolic competition index described in the main text, EMO is a network-based heuristic for estimating the competition between two species. The two indices, EMO and metabolic competition index, are significantly correlated ($\rho = 0.312$, $P < 10^{-3}$). It is important, however, to note a fundamental difference between these two measures: Whereas the metabolic competition index directly quantifies the amount of niche overlap between species, EMO aims to quantify the deleterious downstream effects of a competing partner on the growth of a species. Briefly, to determine the EMO of two species, the nutritional profiles of both species are calculated, overlapping metabolites are removed from the nutritional profile of the query species, and the network expansion algorithm (12) is used to determine how many essential metabolites this species is still capable of synthesizing. Using the EMO score to predict competitive interaction between species, we obtained results qualitatively similar to those observed using our metabolic competition index, with EMO being positively correlated with co-occurrence (correlation was weak but significant). Restricting our analysis to coherent EMO scores further improved the correlation ($\rho = 0.140$, $P < 10^{-4}$). In 110 species (71%), partners have greater EMO than excluders.

Analysis of Coherently Predicted Interactions. Because the nutritional profiles of species vary substantially in size, our predicted interaction indices are not necessarily symmetric. Consider, for example, a species A with a nutritional profile containing many compounds, and a second species B with a nutritional profile containing only a few compounds, all of which also appear in the nutritional profile of species A. In this extreme example, the metabolic competition index of species A on species B is 1 whereas the metabolic competition index of species B on species

A is much smaller than 1 (and approaches 0 as the size of A's nutritional profile increases). Because it is hard to interpret the exact effect of niche overlap in such extreme scenarios (see, e.g., ref. 13), we wish here to control for these cases. We therefore repeated our analysis using only coherent interactions: pairs of mutual indices that are within 0.1 of one another. We found that with this control the observed association between predicted interaction indices and co-occurrence increases and is still highly significant ($\rho = 0.249$, $P < 10^{-4}$ and $\rho = -0.204$, $P < 10^{-4}$, metabolic competition index and metabolic complementarity index, respectively, Mantel correlation test). We additionally found that as the definition of coherent interactions is made more stringent, the magnitude of correlation between interaction indices and co-occurrence increases, potentially indicating that strongly reciprocated interactions exert a larger influence on co-occurrence patterns. We similarly repeated these results analyzing only pairs of species with nutritional profiles whose sizes are within 10 compounds. Again, we found a similar association ($\rho = 0.285$, $P < 10^{-4}$ and $\rho = -0.193$, $P < 10^{-4}$, metabolic competition index and metabolic complementarity index, respectively, Mantel correlation test).

Consistency in Definition of Partners and Excluders. We tested the robustness of our results to the definition of partner and excluder species. We define each species' partners as those with which it shares the greatest co-occurrence, and excluders as those with which it shares the lowest. In the main text, we used a threshold of 25% of species for determining high and low co-occurrence. Here, we examined threshold values ranging from 1% (the most extreme cases of partners and excluders) to 50% (77 species). Using all threshold values, we found that in at least 80% of species the mean competition index with partners is greater than with excluders. We also compared the mean metabolic competition index of partners to the mean metabolic competition index of excluders in different phylogenetic distance bins (discussed in the main text), using each threshold definition. We found that at any threshold less extreme than 25%, species have significantly greater metabolic competition with partners than with excluders in any phylogenetic bin ($P < 0.05$ in all bins, one-tailed Mann–Whitney U test). Using thresholds of 15% or 20%, species have significantly greater metabolic competition with partners than with excluders in all bins but that of the lowest phylogenetic distance. For more extreme threshold values, metabolic competition still tends to be greater with partners than with excluders, but because fewer pairs of species are placed in each bin, significance could not be well established.

Comparison of Species' Partners and Excluders to a Null Model. We used the Mantel test to compare the interaction indices of partners and excluders to a null distribution. Each species' partners and excluders were determined as described in the main text. The number of species that have greater metabolic competition with partners than with excluders, and lower metabolic complementarity with partners than with excluders was determined. To determine the significance of these associations, we randomly shuffled species co-occurrence 10,000 times. For each shuffled matrix, we again determined partners and excluders, and the mean metabolic indices. The P value was calculated as the fraction of shuffled matrices in which a higher or equal number of species were observed with greater metabolic competition with partners than with excluders, or with lower metabolic complementarity with partners than with excluders. We found that the separation of partners and excluders by metabolic competition index and metabolic complementarity index was significantly high ($P < 2 \times 10^{-4}$ and $P < 1 \times 10^{-4}$, respectively, Mantel test).

Metabolic Versatility Cannot Fully Account for the Observed Habitat-Filtering Patterns. Species with larger nutritional profiles (i.e., larger seed sets) are potentially more metabolically versatile and may be able to survive in many environments using subsets of their nutritional profiles (14, 15). Such environmental generalists may be therefore able to survive with a wider range of interacting species, mitigating the competitive influence of niche overlap. Here, we therefore aimed to confirm that the differences in metabolic interaction indices between partners and excluders are not in fact the outcome of variation in nutritional profile size. First, we found no significant correlation between nutritional profile size and mean co-occurrence rank ($\rho = 0.049$, $P < 0.270$, Mantel correlation test), suggesting that such species will not necessarily be considered partners of many other species. Furthermore, we calculated the partial correlation between co-occurrence and metabolic interaction indices and found that controlling for nutritional profile size does not lead to a significant reduction in correlation ($\rho = 0.210$, $P < 0.10^{-3}$ and $\rho = -0.199$, $P < 0.10^{-3}$, metabolic competition and complementarity, respectively; Mantel partial correlation test). We also examined whether the nutritional profile size of partners is consistently different from that of excluders at different phylogenetic distances and exhibits a pattern similar to that observed for metabolic competition in Fig. 3*B*. Using the same phylogenetic relatedness bins as those used in Fig. 3*B* we found that the difference in nutritional profile size between partners and excluders is not consistent. In only four of the six bins do partners have larger nutritional profiles than excluders, and in the most populated bin partners, in fact, have significantly smaller nutritional profiles. Finally, determining the average nutritional profile size among the partners and excluders of each species (as was done for metabolic competition in Fig. 2), we found that for only 58% of the species (90 of 154, a nonsignificant enrichment) do partners have larger nutritional profiles than excluders, compared with 82% of the species in which partners have higher metabolic competition than excluders (main text).

Testing Host Nationality and Enterotype. Because our data include samples from two different cohorts, Danish and Spanish, we further examined whether variation between these two cohorts can account for the observed habitat-filtering pattern. Partitioning our samples and repeating the analysis above considering separately samples from each nationality, we again did not find any qualitative change in the trends reported in the main text (Dataset S1 *G* and *H* and Table S3*C*). Furthermore, because it has recently been suggested that variation in the human intestinal microbiota tends to cluster into three discrete states (termed enterotypes, ref. 16), we similarly confirmed that the association between co-occurrence and metabolic interaction indices holds when controlling for the various enterotypes found in our dataset (Dataset S1 *I*–*L* and Table S3*C*).

Metabolic Competition of Consistent and Inconsistent Partners and Excluders. We determined the consistency of co-occurrence patterns across health states and examined whether consistency is associated with predicted metabolic interactions scores. To this end, we partitioned the samples into two groups: healthy individuals and those with inflammatory bowel disease (IBD). We determined each species' partners and excluders in each group separately as before (*Methods*). We found that the vast majority (96%) of co-occurrence partnerships were consistent across health states: 3,800 pairs consistently co-occur and 3,912 pairs consistently exclude across healthy and IBD samples, whereas only 140 pairs co-occur in healthy and exclude in IBD and 201 pairs exclude in healthy and co-occur in IBD. Examining the metabolic competition index for consistent and inconsistent species pairs separately, we again found a clear association between co-occurrence and metabolic interaction: Consistent partners

exhibit significantly higher metabolic competition than consistent excluders, with inconsistent partners/excluders exhibiting intermediate competition levels (Fig. S2).

Consistency of Species' Partners and Excluders Separation Across Species' Ecological Traits. We examined whether the difference observed between a species' partners and excluders is consistent across species and, specifically, whether species associated with a certain ecological attribute escape these assembly rules. To this end, ecological characteristics were collected from the National Center for Biotechnology Information Genome Project's table of prokaryotic genomes. Six species did not have an exact strain match, and an alternative strain was used. Three characteristics relevant to the ecology of species were recorded: oxygen requirement, habitat preference, and pathogenicity. These data were used to label each species with a subset of the following attributes: pathogen, human pathogen, anaerobe, facultative anaerobe, host-associated, and cosmopolitan (Table S2*B*). Species not listed as pathogenic were assumed to be benign. Other omitted annotations were treated as missing information. For each ecological attribute (e.g., anaerobes), the number of species for which partners have higher mean competition index than excluders or for which excluders have higher mean complementarity index than partners was counted, as well as the number of total species for which information about this ecological attribute is available. We used a hypergeometric enrichment test to determine whether any of the ecological attributes tested is enriched among species with properly separated partners and excluders. We found that species labeled with each of these attributes exhibit a pattern similar in terms of their metabolic competition and complementarity with partners and with excluders to those not labeled with the attribute (Table S4).

Comparison of Competition, Complementarity, and Phylogeny in Distinguishing Partners vs. Excluders. We examined the ability of three different indices in distinguishing between each species' partners and excluders: the metabolic competition index, the metabolic complementarity index, and phylogenetic relatedness determined by 16s similarity. We compared these metrics in partners and excluders of each of the 143 species for which estimates of phylogenetic relatedness are available (*Methods*). As before, we classified as partners of a given species the 25% of species with which it has the highest co-occurrence scores and as excluders the 25% of species with which it has the lowest co-occurrence scores. We found that each of the three indices above distinguishes partners and excluders roughly equivalently: 81% of species (116 out of 143) have greater metabolic competition with partners than with excluders, 86% (123 out of 143) have lower metabolic complementarity with partners than with excluders, and 78% (112 out of 143) have greater phylogenetic relatedness with partners than with excluders. We found, however, that the sets of species for which each index correctly distinguishes partners and excluders is not identical (Fig. S3), suggesting that these three criteria, competition, complementarity, and phylogeny, encapsulate distinct information about the co-occurrence of species.

Correlation on Co-Occurrence and Metabolic Interaction Indices in Human Microbiome Project Oral Samples. In the oral community, the observed correlation between co-occurrence and metabolic interaction was found to be generally weaker than the correlation obtained for other body sites and was not statistically significant for the Metabolic Complementarity Index. The lack of a clear habitat-filtering signature within the oral community may be attributed to a number of factors. First, whereas the other body sites generally represent a single specific subsite (e.g., the nares in the airways), the oral community was sampled from several distinct subsites, each of which represents a specific niche (17). Second, although the α-diversity of the oral community is higher than that of other communities (18), the number of organisms surveyed that mapped to sequenced genomes was similar to other sites, potentially underrepresenting the community and making it more susceptible to the influence of noise in the data.

1. Palmer RJ, Jr., Kazmerzak K, Hansen MC, Kolenbrander PE (2001) Mutualism versus independence: Strategies of mixed-species oral biofilms in vitro using saliva as the sole nutrient source. *Infect Immun* 69(9):5794–5804.
2. Chalmers NI, Palmer RJ, Jr., Cisar JO, Kolenbrander PE (2008) Characterization of a Streptococcus sp.-Veillonella sp. community micromanipulated from dental plaque. *J Bacteriol* 190(24):8145–8154.
3. Periasamy S, Chalmers NI, Du-Thumm L, Kolenbrander PE (2009) Fusobacterium nucleatum ATCC 10953 requires Actinomyces naeslundii ATCC 43146 for growth on saliva in a three-species community that includes Streptococcus oralis 34. *Appl Environ Microbiol* 75(10):3250–3257.
4. Periasamy S, Kolenbrander PE (2009) Aggregatibacter actinomycetemcomitans builds mutualistic biofilm communities with Fusobacterium nucleatum and Veillonella species in saliva. *Infect Immun* 77(9):3542–3551.
5. Periasamy S, Kolenbrander PE (2009) Mutualistic biofilm communities develop with Porphyromonas gingivalis and initial, early, and late colonizers of enamel. *J Bacteriol* 191(22):6804–6811.
6. Periasamy S, Kolenbrander PE (2010) Central role of the early colonizer Veillonella sp. in establishing multispecies biofilm communities with initial, middle, and late colonizers of enamel. *J Bacteriol* 192(12):2965–2972.
7. Chao A, Chazdon RL, Colwell RK, Shen T-J (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett* 8:148–159.
8. Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325–349.
9. Horn HS (1966) Measurement of "overlap" in comparative ecological studies. *Am Nat* 100:419–424.
10. Aitchison J (1982) The statistical analysis of compositional data. *J R Stat Soc, B* 44: 139–177.
11. Kreimer A, Doron-Faigenboim A, Borenstein E, Freilich S (2012) NetCmpt: A network-based tool for calculating the metabolic competition between bacterial species. *Bioinformatics* 28(16):2195–2197.
12. Ebenhöh O, Handorf T, Heinrich R (2004) Structural analysis of expanding metabolic networks. *Genome Inform* 15(1):35–45.
13. Mahowald MA, et al. (2009) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc Natl Acad Sci USA* 106(14): 5859–5864.
14. Borenstein E, Kupiec M, Feldman MW, Ruppin E (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci USA* 105(38): 14482–14487.
15. Borenstein E, Feldman MW (2009) Topological signatures of species interactions in metabolic networks. *J Comput Biol* 16(2):191–200.
16. Arumugam M, et al.; MetaHIT Consortium (2011) Enterotypes of the human gut microbiome. *Nature* 473(7346):174–180.
17. Faust K, et al. (2012) Microbial co-occurrence relationships in the human microbiome. *PLOS Comput Biol* 8(7):e1002606.
18. Huttenhower C, et al.; Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207–214.

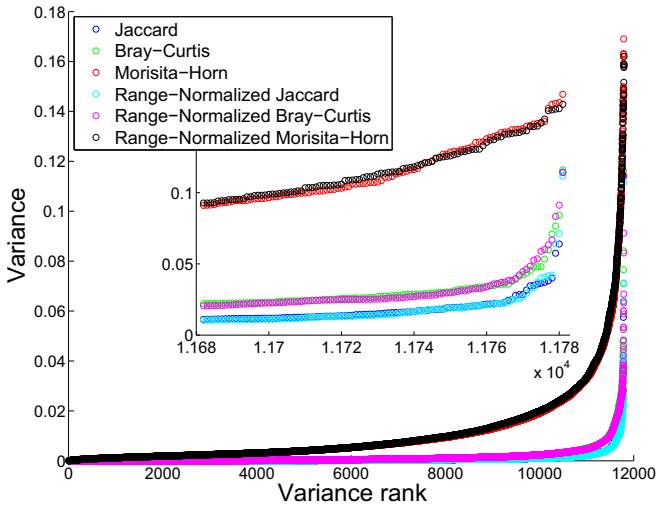

Fig. S1. Robustness of co-occurrence metrics to undersampling. Species abundances were subsampled 1,000 times, from which co-occurrence of pairs was calculated. Variance across each species pair's 1,000 co-occurrence values is plotted, with variance sorted from smallest to largest. (*Inset*) Variance in Jaccard similarity index rises the slowest and has the lowest maximum of all co-occurrence metrics tested.

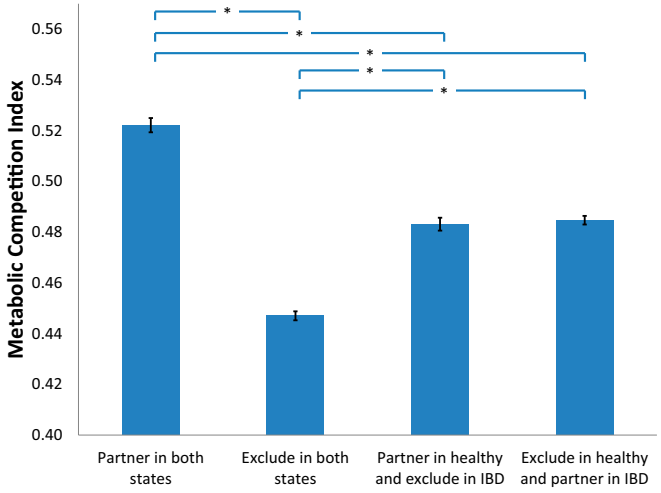

Fig. S2. Metabolic competition index for consistent and inconsistent species co-occurrence. Bars represent the mean metabolic competition and SE for species pairs that are found to be consistent partners or consistent excluders (i.e., co-occur/exclude in both healthy and inflammatory bowel disease samples) and for pairs that exhibit inconsistent co-occurrence patterns. Consistent partners have significantly different metabolic competition index from consistent excluders and from inconsistent pairs ($P < 0.05$; Wilcoxon rank sum test).

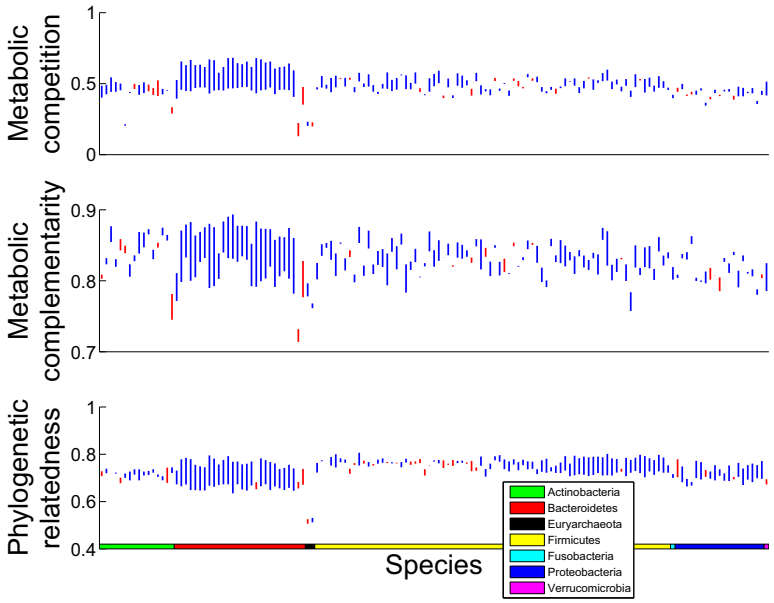

Fig. S3. Comparison of competition, complementarity, and phylogeny in distinguishing partners vs. excluders in the human intestinal community. Bars represent the range from the mean score among partners to the mean score among excluders. Blue bars indicate accurate separation, with partners having higher metabolic competition index, lower metabolic complementarity index, or higher phylogenetic relatedness than excluders.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)
Dataset S1 (XLSX)
Dataset S2 (XLSX)