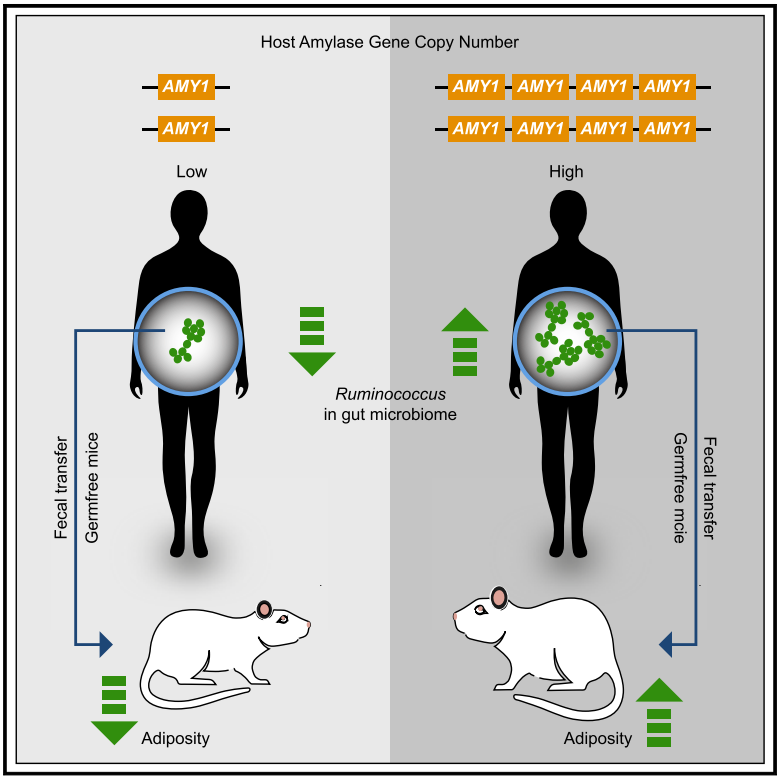
Article

Cell Host & Microbe

Human Salivary Amylase Gene Copy Number Impacts Oral and Gut Microbiomes

Graphical Abstract



Highlights

- High *AMY1* copy number (CN) is associated with higher levels of *Porphyromonas* in saliva
- High *AMY1*-CN stool has more resistant starch degraders; drives more adiposity in GF mice
- Stool short-chain fatty acid levels are predictive of salivary amylase activity
- Upon diet standardization, gut microbiomes converged without eliminating differences

Authors

Angela C. Poole, Julia K. Goodrich, Nicholas D. Youngblut, ..., Daniel H. Huson, James G. Booth, Ruth E. Ley

Correspondence

rley@tuebingen.mpg.de

In Brief

Poole et al. examine how copy-number (CN) variation of the AMY1 gene, encoding salivary amylase, relates to human microbiomes. Individuals with high AMY1-CN have an increased number of oral *Porphyromonas*, which is linked to periodontitis. Their gut microbiotas have a higher abundance of resistant starch-degrading microbes and drive higher adiposity when transferred to germ-free mice.

Poole et al., 2019, Cell Host & Microbe *25*, 553–564
April 10, 2019
https://doi.org/10.1016/j.chom.2019.03.001

CellPress

Cell Host & Microbe
Article

Human Salivary Amylase Gene Copy Number Impacts Oral and Gut Microbiomes

Angela C. Poole,[1,2,7] Julia K. Goodrich,[1] Nicholas D. Youngblut,[1] Guillermo G. Luque,[1] Albane Ruaud,[1] Jessica L. Sutter,[1] Jillian L. Waters,[1] Qiaojuan Shi,[2] Mohamed El-Hadidi,[3,8] Lynn M. Johnson,[4] Haim Y. Bar,[5] Daniel H. Huson,[3] James G. Booth,[6] and Ruth E. Ley[1,2,9,*]

[1]Department of Microbiome Science, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany
[2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[3]Center for Bioinformatics, University of Tübingen, 72076 Tübingen, Germany
[4]Cornell Statistical Consulting Unit, Cornell University, Ithaca, NY 14853, USA
[5]Department of Statistics, University of Connecticut, Storrs, CT 06269, USA
[6]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA
[7]Present address: Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA
[8]Present address: Bioinformatics Group, Center for Informatics Science (CIS), Nile University, Giza, Egypt
[9]Lead Contact
*Correspondence: rley@tuebingen.mpg.de
https://doi.org/10.1016/j.chom.2019.03.001

SUMMARY

Host genetic variation influences microbiome composition. While studies have focused on associations between the gut microbiome and specific alleles, gene copy number (CN) also varies. We relate microbiome diversity to CN variation of the *AMY1* locus, which encodes salivary amylase, facilitating starch digestion. After imputing *AMY1*-CN for ~1,000 subjects, we identified taxa differentiating fecal microbiomes of high and low *AMY1*-CN hosts. In a month-long diet intervention study, we show that diet standardization drove gut microbiome convergence, and *AMY1*-CN correlated with oral and gut microbiome composition and function. The microbiomes of low-*AMY1*-CN subjects had enhanced capacity to break down complex carbohydrates. High-*AMY1*-CN subjects had higher levels of salivary *Porphyromonas*; their gut microbiota had increased abundance of resistant starch-degrading microbes, produced higher levels of short-chain fatty acids, and drove higher adiposity when transferred to germ-free mice. This study establishes *AMY1*-CN as a genetic factor associated with microbiome composition and function.

INTRODUCTION

Host genotype has recently emerged as a significant factor in shaping the relative abundance of specific members of the human gut microbiota (Goodrich et al., 2016; Rothschild et al., 2018). Heritable gut microbes, whose abundances are influenced by host genotype, have been identified, and genome-wide association studies (GWASs) have linked specific gene variants to members or functions of the gut microbiome (Goodrich et al., 2014, 2016, 2017; Lim et al., 2017; Davenport et al., 2016; Turpin et al., 2016; Davenport et al., 2015). In addition to single nucleotide polymorphism (SNP) differences between individuals, another important aspect of human genetic variation is the copy number (CN) of genes. Gene duplications resulting in increased CN provide a rapid means of adaptation to environmental change (Iskow et al., 2012). Copy-number variation (CNV) in genes accounts for more genomic variability than SNPs (Conrad et al., 2010) and significantly influences gene expression (Chiang et al., 2017). This important aspect of genetic variability likely affects microbiome differences between individuals, but links between the CNV of specific human genes and the microbiome remain to be elucidated.

CNV of the *AMY1* gene, encoding the salivary amylase enzyme, is considered one of the strongest signals of recent natural selection on human populations (Perry et al., 2007). Salivary amylase hydrolyzes alpha bonds of starch and glycogen, beginning the process of starch degradation in the mouth. *AMY1*-CN is positively correlated with oral amylase activity (Mandel et al., 2010; Perry et al., 2007). A shift to greater starch consumption during the agronomic transition of the Neolithic period likely selected for the duplications observed within the *AMY1* locus (Kelley and Swanson, 2008; Iskow et al., 2012; Perry et al., 2015). Today, the mean *AMY1*-CN is reported higher in populations with an agrarian background compared to hunter-gatherers (Perry et al., 2007). Across genetic backgrounds, human *AMY1*-CN ranges from 2 to 24 (Usher et al., 2015; Yong et al., 2016; Perry et al., 2007).

Complex carbohydrates, a broad category of polysaccharides that includes starch, first encounter amylase in the mouth, then in the small intestine (SI), where pancreatic amylase is added to the chyme and the liberated sugars are absorbed. SI uptake of sugars liberated by host enzymes yields more energy to the host than uptake in the large intestine (LI) of microbial fermentation products (Walter and Ley, 2011). Host-microbial competition for starch may have driven selection for duplications at the amylase locus. Indeed, amylase supplementation to farm animals enhances starch digestibility and promotes weight gain (Burnett, 1962; Gracia et al., 2003; Jo et al., 2012). Similarly, humans with a high *AMY1*-CN (AMY1H), who produce high levels of salivary amylase, should derive more energy from the same

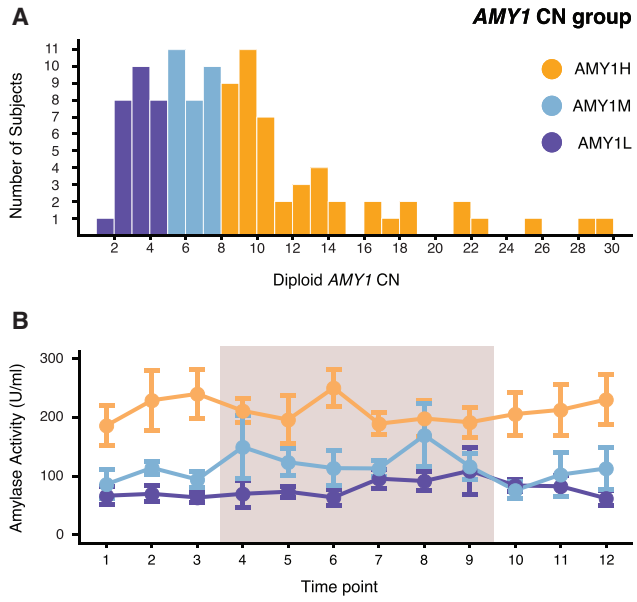


Figure 1. *AMY1*-CN Distribution and Mean Amylase Salivary Activity for the Intervention Group

(A) Diploid *AMY1*-CN distribution for 105 subjects, from which the 25 intervention subjects were selected, was obtained using qPCR with primer sequences previously reported (Perry et al., 2007).

(B) Mean amylase activity per mL of saliva ± SEM at each time point for the 25 individuals in the *AMY1*-CN intervention groups. Measurements were performed in triplicate for both qPCR and salivary amylase activity.

carbohydrate-rich diet than those with a low *AMY1*-CN (AMY1L). Compared to AMY1H, AMY1L individuals might be expected to harbor gut microbiomes with a greater capacity for breakdown of complex carbohydrates, compensating for the lower levels of host amylase.

Because of its link to carbohydrate digestion, *AMY1*-CN has been investigated for associations with BMI and metabolism. Results of these studies vary, with low *AMY1*-CN associated with high BMI in some populations (Viljakainen et al., 2015; Mejía-Benítez et al., 2015; Falchi et al., 2014; Marcovecchio et al., 2016; Bonnefond et al., 2017), but not others (Usher et al., 2015; Yong et al., 2016). Discrepancies may be methodological (Usher et al., 2015) or due to variation in starch intake between individuals (Rukh et al., 2017). Whether the gut microbiome, which is known to impact host metabolism (Sonnenburg and Bäckhed, 2016; Zeevi et al., 2015; Pedersen et al., 2016; Goodrich et al., 2014), responds to host *AMY1*-CN remains to be ascertained.

Here, we address how *AMY1*-CN relates to the diversity and function of the gut microbiomes of healthy individuals with normal BMIs. Using an existing dataset of genotyped individuals with associated fecal microbiome diversity data, we identified taxa that discriminated high and low *AMY1*-CN individuals. We then conducted an intervention study by screening >100 volunteers for *AMY1*-CN and recruiting 25 participants into a 1-month longitudinal study in which diet was standardized for 2 weeks. We used 16S rRNA gene sequence analysis to assess the effects of host *AMY1*-CN and diet intervention on oral and fecal microbiomes. In addition, we obtained a functional characterization of fecal microbiomes through (1) deep metagenomic sequencing, (2) short-chain fatty acid (SCFA) measures, and (3) fecal transfers to germ-free mice. Together, results support an association between host *AMY1*-CN and microbiome diversity and function.

RESULTS

Identification of Gut Microbiota that Discriminate between Hosts with High and Low Predicted *AMY1*-CN Genotypes

We searched for microbial taxa that discriminate fecal microbiomes of subjects with high or low *AMY1*-CN in a genotyped population with available fecal microbiome data (Goodrich et al., 2016; STAR Methods). Available genotype data included 7 of the 10 SNPs that correlate with *AMY1*-CN (Usher et al., 2015). For each of 994 subjects with normal BMIs and available microbiome data, we calculated the sum of the change in *AMY1*-CN values corresponding to each of their 7 alleles. We then selected the top and bottom 5% of the distribution (50 subjects at each extreme of predicted total difference in *AMY1*-CN). Using a bivariate model (hereafter referred to as "Harvest") (Bar et al., 2014), we identified 17 operational taxonomic units (OTUs) whose relative abundances discriminated high and low groups (Table S1). Some of these OTUs were classified as *Ruminococcus*, *Faecalibacterium prausnitzii*, and *Bacteroides*. Members of the Ruminococcaceae family were prominent among the taxa enriched in fecal microbiomes obtained from subjects with predicted high *AMY1*-CN.

Intervention Study and Participants

We collected buccal swabs from 105 volunteers recruited on the Cornell University campus and measured their *AMY1*-CN by qPCR (Figure 1A; Table S2; Data S1), then selected 25 participants across the *AMY1*-CN distribution for further study. We confirmed the *AMY1*-CN of participants using alternate qPCR primers and digital PCR (STAR Methods; Table S3). Based on the results, 11 participants were assigned to a high group (CN > 8, designated AMY1H), 5 to a medium group (5 < CN < 8, designated AMY1M), and 9 to a low group (CN < 5, designated AMY1L). Neither BMI nor body fat percentage differed significantly between groups (Table S3). CN of the gene for pancreatic amylase, *AMY2*, was positively correlated with *AMY1*-CN (Spearman's rho = 0.79, $p = 3 \times 10^{-8}$; Table S3) and had a smaller range.

Dietary Intake Was Similar between *AMY1*-CN Groups throughout the Study

To mitigate the effects of dietary differences between individuals on their microbiomes and to promote frequent starch consumption, during study weeks 2 and 3 we provided all participants with all meals and snacks. Participants ate from the same menu freely, occasionally supplemented it (Table S4), and recorded their dietary intake in food diaries (STAR Methods). Based on dietary records, mean percentages of total carbohydrate, protein, and fat intake did not differ significantly between the AMY1H, AMY1M, and AMY1L groups, regardless of whether meals were consumed during (weeks 2–3) or outside (weeks 1 and 4) of the standardized diet period (Figures S1A–S1C). The intake of all three macronutrients differed between

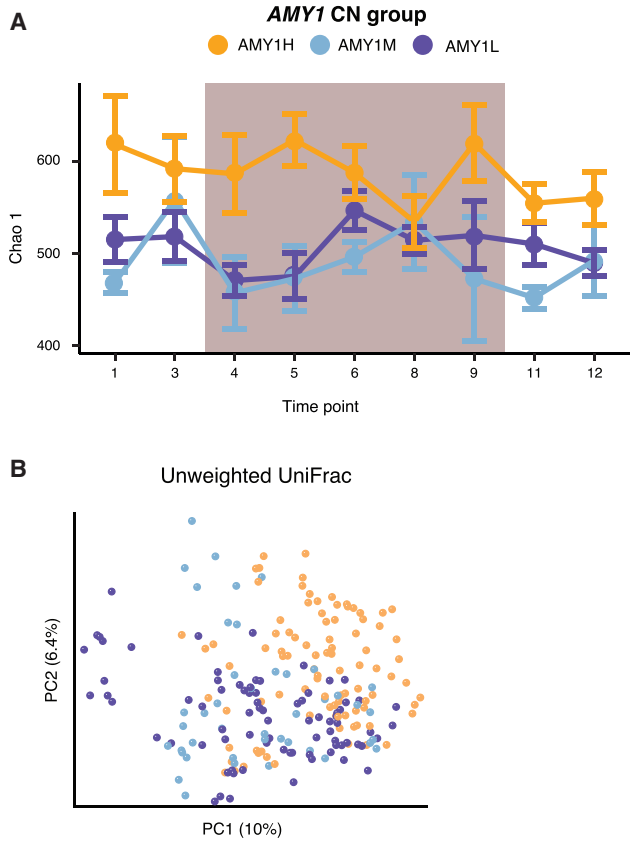


Figure 2. Oral Microbiomes Differ in Diversity between *AMY1*-CN Groups

(A) Alpha diversity assessed with the metric Chao 1 (n = 25).

(B) Principal coordinate analysis (PCoA) of the unweighted UniFrac distances between samples collected from all 25 subjects throughout the study. The first two PCs are plotted. The percent variation explained by each PC is indicated on the axes. Samples are colored according to *AMY1*-CN group. See also Figure S2.

days ($p < 1 \times 10^{-5}$). However, the standardized diet period did not significantly impact mean intake of macronutrients.

AMY1-CN versus Oral and Fecal Amylase Activity

We obtained saliva and stool samples at 12 time points (TPs; 3 per week). Salivary amylase activity (SAA) ranged between 10.2 and 527 units per mL of saliva, similar to previously reported ranges (Mandel et al., 2010). *AMY1*-CN correlated with SAA across all subjects at all TPs (linear mixed model; $p = 2.1 \times 10^{-5}$). SAA levels for the AMY1H were higher than for the AMY1L at all TPs (linear mixed model; $p = 1.9 \times 10^{-4}$; Figure 1B) with AMY1M individuals intermediate (Figure S2A). Fecal amylase activity (FAA) was variable within and between subjects (0.6–1,120 Ug^{-1} stool; Figure S1D). Unlike SAA, FAA did not correlate with *AMY1*-CN. To further characterize FAA, we used an ELISA method (STAR Methods) to measure levels of host pancreatic amylase in stool samples at TPs 6 and 10. Across all 25 subjects, pancreatic amylase levels were highly correlated with FAA (Spearman's rho = 0.80, $p = 3.7 \times 10^{-6}$ for TP 6 and Spearman's rho = 0.78, $p = 6.3 \times 10^{-5}$ for TP 10; Figure S1E), although *AMY2*-CN was not. This finding corroborates previous reports that FAA is largely pancreatic (Macfarlane and Englyst, 1986; Moriyoshi et al., 1991).

Classification of the Oral Microbiota by Host *AMY1*-CN

Saliva samples were profiled for microbial community diversity by Illumina sequencing 16S rRNA gene PCR amplicons (V4 region; Table S2). Sequences were clustered into OTUs using a threshold of 97% pairwise sequence identity (STAR Methods). We observed that oral microbiome richness (alpha diversity) was correlated with *AMY1*-CN (using Chao 1, Observed Species, and Faith's PD metrics $p < 0.01$, but not Shannon's Index, which is also a measure of evenness) and was higher in AMY1H than AMY1L individuals ($p = 0.011$; Figure 2A).

We applied the UniFrac metrics to assess between-subject (beta) diversity. Principal coordinate analysis (PCoA) of unweighted UniFrac distance metrics revealed clustering of saliva microbiomes by subject and a trend for separation by *AMY1*-CN group across all subjects (Figures 2B, S2B, and S2C). Together, these observations indicate that across the *AMY1*-CN gradient, a higher *AMY1*-CN is associated with greater richness of the microbiome without a significant shift in overall diversity.

We searched for OTUs that distinguished AMY1H and AMY1L categories (AMY1M excluded) using a machine learning technique (random forest; STAR Methods). A model trained on 80% of the samples produced an accuracy of 97.67% with a Matthews correlation coefficient (MCC) of 94.92% and an area under the curve (AUC) of 96.66% once it was tested on the remaining 20% of the samples. We then used a feature selection process to identify the relevant features of the model. Among the top OTUs that most discriminated AMY1H and AMY1L groups (AMY1M excluded) were OTUs classified as *Prevotella* and *Porphyromonas* (Figure 3A; Table S5). In order to include all 25 participants, we reclassified the individuals into only two new *AMY1* groups, AMY1H′ (CN > 6) and AMY1L′ (CN < 6), using k-means. Using this assignation, 13 subjects were in the AMY1H′ and 12 were in the AMY1L′. Performing machine learning using the AMY1H′ and AMY1L′ groups yielded similar results (Figure S3).

To gain a time-resolved view into the taxa driving differences for subjects at the extremes of the *AMY1*-CN gradient, we compared AMY1H and AMY1L groups (AMY1M excluded) at each TP using Harvest. This analysis revealed 9 OTUs with significantly different mean relative abundances between AMY1H and AMY1L (Figure 3B); none exhibited different variances. As observed for the machine learning analysis, OTUs that discriminated the AMY1H and AMY1L groups belonged to the genera *Prevotella*, *Haemophilus*, *Neisseria*, and *Porphyromonas* (Figure 3B; Table S5). These patterns highlight that the same OTUs are consistently elevated over time in either AMY1H or AMY1L.

Classification of the Fecal Microbiota by Host *AMY1*-CN

In contrast to oral microbiomes, the alpha diversity of fecal microbiomes was generally similar between *AMY1*-CN groups (using Chao 1, Observed Species, and Faith's PD metrics; Figure 4A). Beta diversity was unrelated to host *AMY1*-CN (Figures 4B and S4A), with some clustering by subject (Figure S4B). Using a random forest analysis, the prediction performed on the 20% of

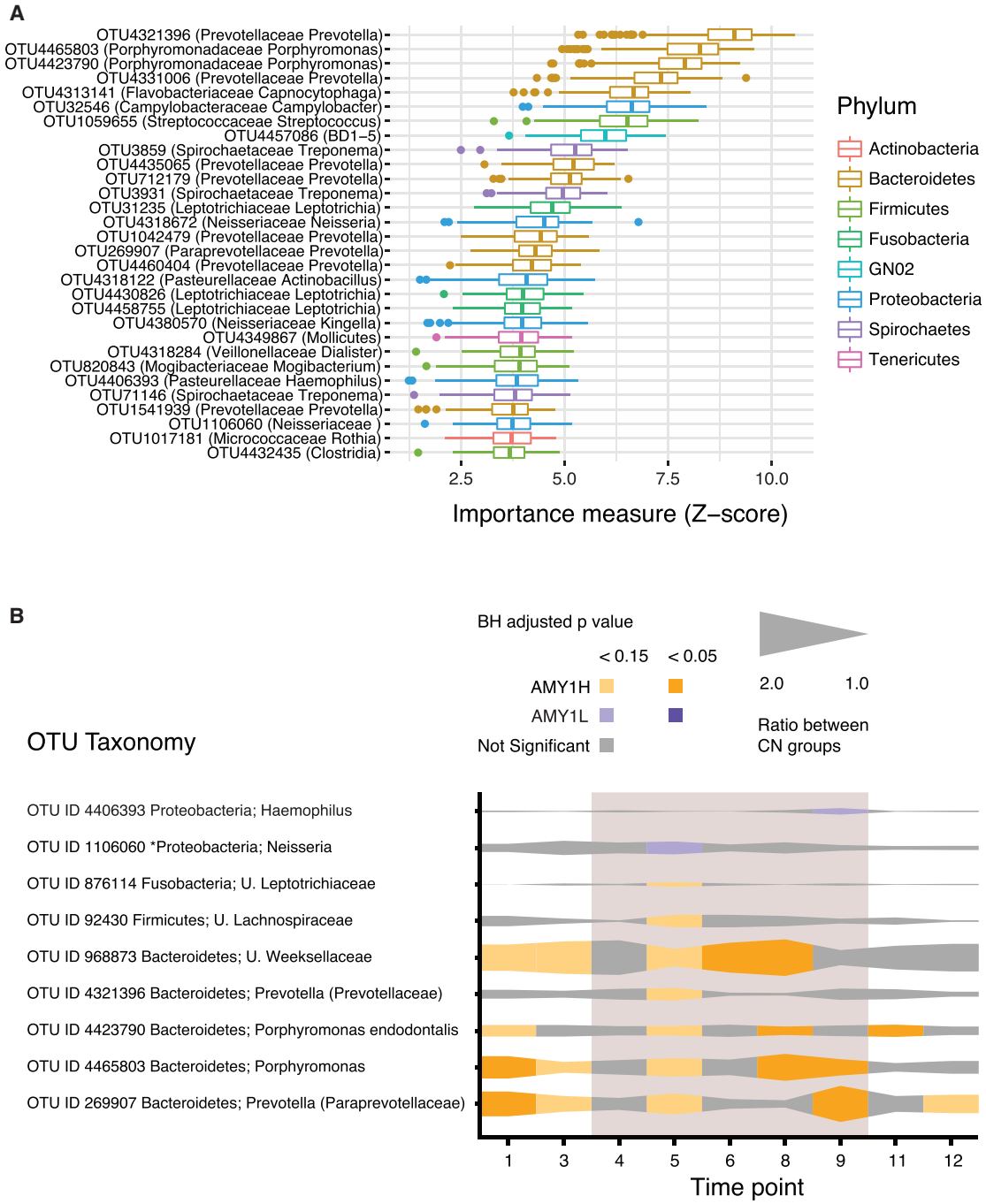


Figure 3. Oral Microbiomes Differ between *AMY1*-CN Groups at the OTU Level

(A) The OTUs shown here were identified using machine learning as distinguishing between AMY1H and AMY1L groups (n = 20). The length of the bar represents the magnitude of the mean decrease in Gini index and the orientation indicates the group in which the OTU is enriched.

(B) The mean relative abundances of the OTUs included in this ribbon plot were significantly different between AMY1H and AMY1L groups at one or more TPs using the statistical model Harvest (n = 20). Each ribbon corresponds to a single OTU with taxonomy indicated to the left, with unclassified abbreviated "U." Taxonomy may be shared by several OTUs. The width of the ribbon at each TP shows the ratio of the mean OTU relative abundances between the *AMY1*-CN groups. If the ribbon is colored orange at a given TP, the AMY1H group has a higher mean relative abundance of the OTU; when purple, the AMY1L has a higher mean relative abundance. When the ribbon is colored gray, the Benjamini-Hochberg (BH) adjusted $p \geq 0.15$. Lighter orange or purple corresponds to a BH adjusted $p < 0.15$; darker colors correspond to BH adjusted $p < 0.05$. The asterisk (*) denotes an OTU that was assigned taxonomy with higher resolution after performing a BLAST search using the representative sequence. See also Figure S3 and Table S5.

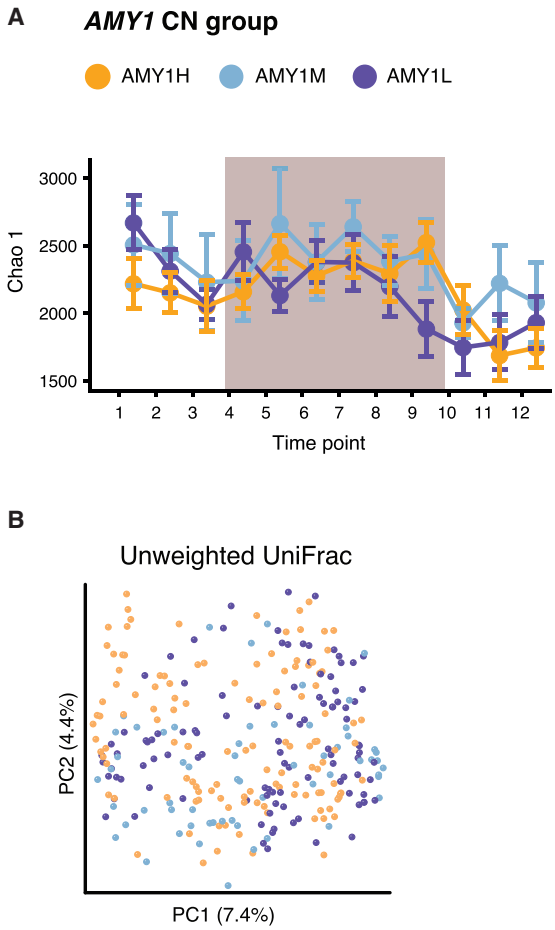


Figure 4. Gut Microbiomes Do Not Differ in Overall Composition between *AMY1*-CN Groups

(A) Alpha diversity (Chao 1) for the gut microbiomes over time.

(B) PCoA of the unweighted UniFrac distances between microbiomes derived from fecal samples.

See legend for Figure 2. See also Figure S4.

the samples reserved as a testing dataset produced an accuracy of 98.21% with an MCC of 96.04% and an AUC of 97.36%. After performing a feature selection process, we identified OTUs discriminating between AMY1H and AMY1L as belonging to Ruminococcaceae (*Ruminococcus* and *Oscillospira*) and Lachnospiraceae (*Blautia*, *Dorea*, and *Roseburia*; Figure 5A; Table S6). Similarly, when all 25 subjects were reclassified into low and high groups based on k-means clustering (AMY1M included, as above), a similar set of discriminatory OTUs was observed (Figure S5).

We also identified taxa that differentiated AMY1H and AMY1L groups at each TP (using Harvest): 7 of the 11 discriminatory OTUs were members of the Ruminococcaceae family, and all but one were elevated in the AMY1H compared to the AMY1L group (Figure 5B; Table S6). As observed in the oral microbiota, the Harvest analysis showed that discriminatory OTUs were consistently enriched in the same *AMY1*-CN category. Members of the Ruminococcaceae have been linked to resistant starch degradation: enrichment of Ruminococcaceae OTUs in the AMY1H is consistent with reduced availability of starches susceptible to host amylase degradation in the distal gut of the AMY1H host. OTUs classified to the *Ruminococcus* genus were also enriched in the 994 subjects (above) with predicted high host *AMY1*-CN.

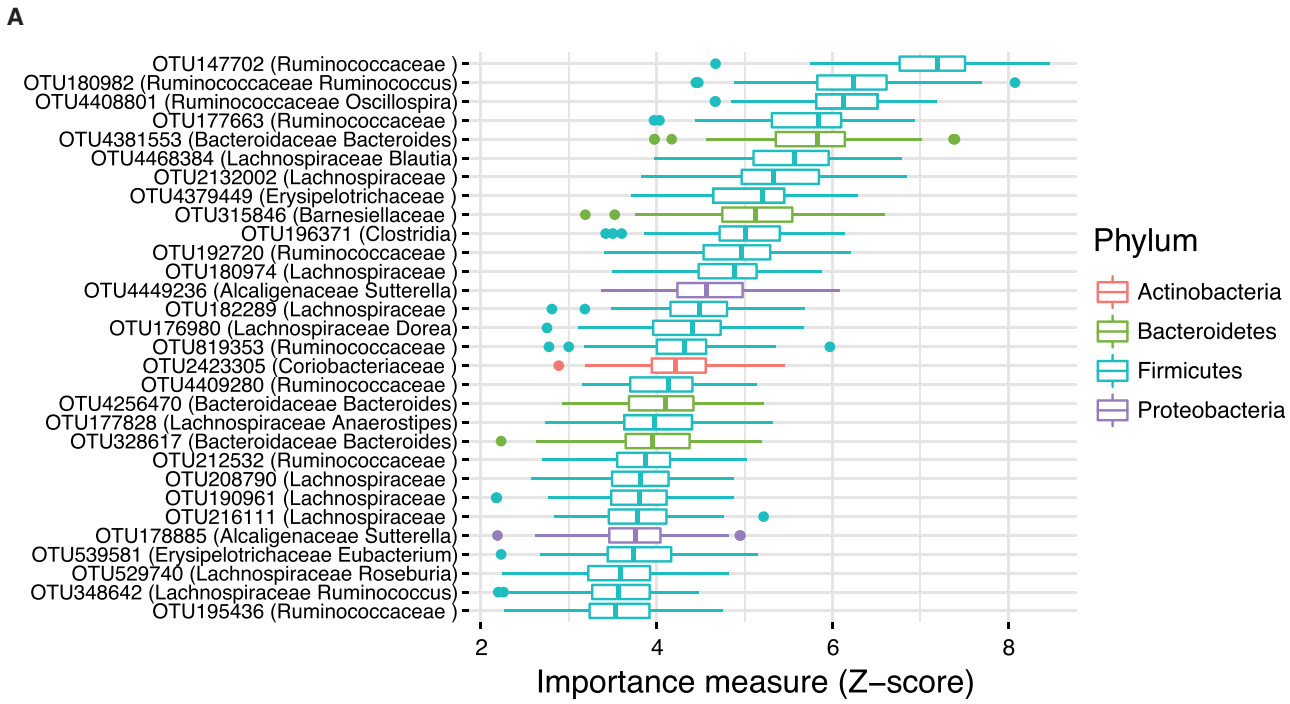
Effect of Diet on the Oral and Fecal Microbiomes

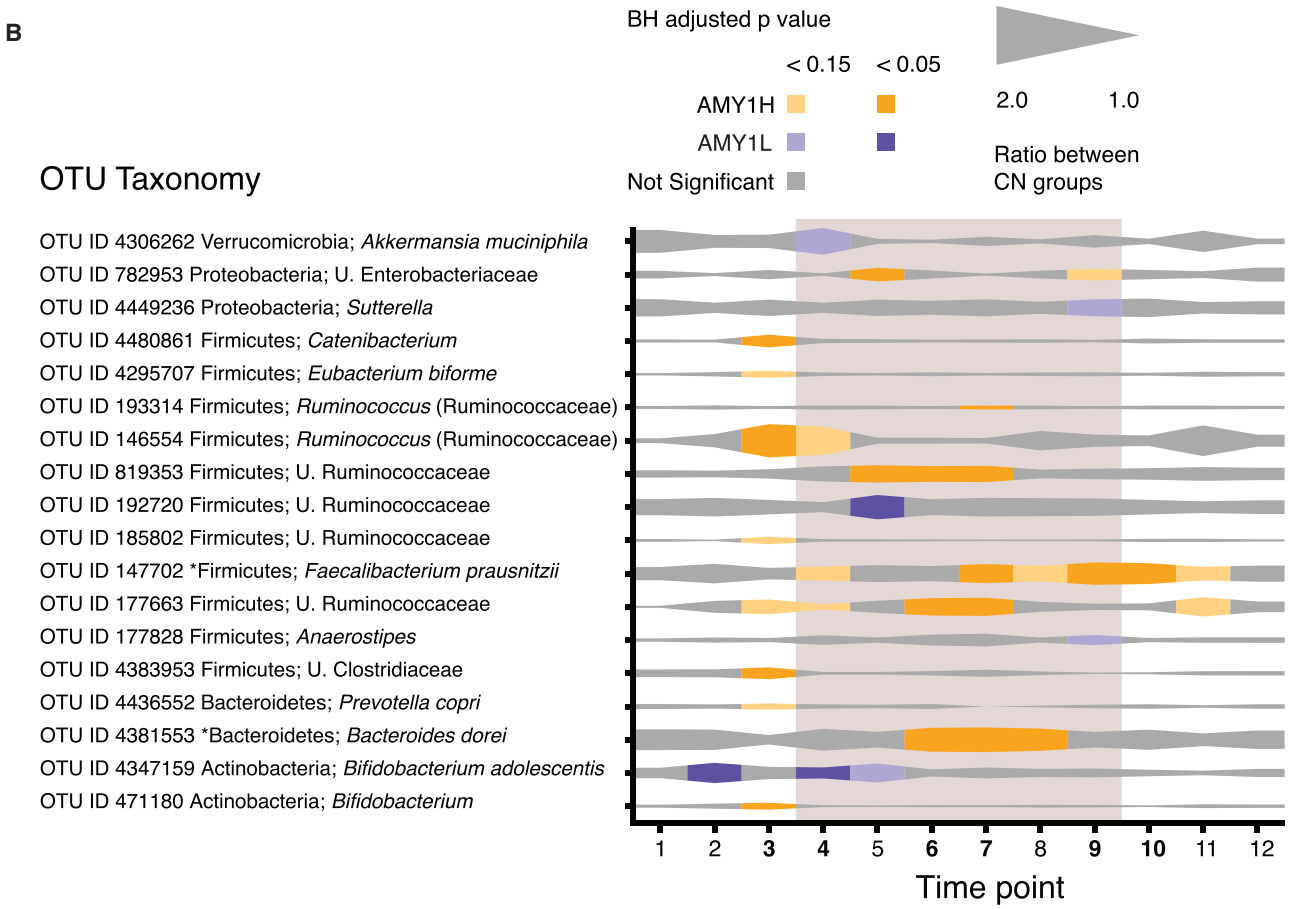
We assessed whether diet standardization resulted in more similar microbiomes (i.e., reduced beta diversity) for AMY1H and AMY1L groups by comparing samples before, during, and after the standardized diet. Non-parametric bootstrap confidence intervals (CIs) for the differences in the weighted and unweighted UniFrac distances between AMY1H and AMY1L groups for each pair of diet periods indicated that diet standardization did not induce convergence of oral microbiomes between AMY1H and AMY1L subjects but did so for the gut microbiomes (Figures S5B and S5C).

Deep Metagenome Sequencing Reveals Differences in Functional Capacity between *AMY1*-CN Gut Microbiomes

We compared the metabolic potentials of the gut microbiomes for AMY1H and AMY1L groups using metagenomes generated for all subjects sampled at 6 TPs (3, 4, 6, 7, 9, and 10). We subsampled 20 million paired-end reads per sample to normalize sequencing depth and used the HMP Unified Metabolic Analysis Network (HUMAnN2) pipeline to classify shotgun metagenomic reads into gene families (Figure 6A). Using non-parametric bootstrapping with 1,000 permutations, we determined that Bray-Curtis distances calculated from gene family counts decreased in mean values between AMY1H and AMY1L individuals during the diet period relative to the pre-diet period (Figure 6B). The number of gene families significantly enriched in the AMY1H group was also lower after TP 4 (Figure 6D), and the number of gene families significantly different between AMY1 groups was lower after TP 4 regardless of taxonomy or function (Figures 6C and 6E). We also observed a spike of differentially abundant genes associated with mobile elements at the start of the diet provision period (TP 3 to TP 4), which is consistent with nutritional stress-induced activation of prophages, lytic bacteriophages, and horizontal gene transfer (Huddleston, 2014; Lerner et al., 2017). Thus, the diet standardization drove the convergence of metabolic functions as well as composition.

Despite the convergence of AMY1H and AMY1L microbiomes over the standardized diet period, the two groups could be differentiated by their functional gene content. We used the statistical software DESeq2 to identify gene families with differential abundances between high and low AMY1 groups at each time point. We identified 481 gene families with significantly different read counts at one or more TPs between AMY1H and AMY1L groups (Figure 6A). Notably, 39% of the 481 gene families were taxonomically assigned to *Bacteroides dorei* and were more abundant in the AMY1H group at multiple TPs, in accordance with our 16S rRNA gene diversity results (Figures 5B and 6A). Eight other species of *Bacteroides* were identified, but only *Bacteroides cellulosilyticus* was enriched in the AMY1L group. In accordance with the results of the 16S rRNA gene diversity analysis, reads for gene families mapping to *Ruminococcus* were enriched in the AMY1H group (Figures 5B and 6A).





(legend on next page)

Sequences mapping to gene families from *Prevotella copri* were also more abundant in AMY1H, but almost exclusively at TP 3, consistent with the 16S rRNA gene diversity data. Together, these data support a relative enrichment in taxa responsible for resistant starch breakdown in the AMY1H microbiomes.

To directly assess functional capacity for carbohydrate degradation, we used hidden Markov models from dbCAN to identify carbohydrate-active enzymes (CAZymes), which include the following enzyme classes: glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), carbohydrate-binding modules (CBM), S-layer homology modules of the cellulosome (SLH), and auxiliary activities (AA) (Lombard et al., 2014). We then used a linear mixed model to assess differences in the abundances of each of the 7 CAZyme classes between AMY1H and AMY1L groups. We observed a higher number of read counts for GH and PL classes in AMY1L individuals (post hoc analysis; G, $p = 0.0054$; PL, $p = 0.030$; Figures 7A and 7B), and a similar trend for AA and CE classes (Figure S6). These enzyme classes are involved in the breakdown of complex carbohydrates; their enrichment in AMY1L individuals is consistent with the notion that a greater proportion of complex carbohydrates reaches the distal gut in AMY1L individuals.

Fecal SFCAs Relate to Salivary Amylase Activity

As an assessment of microbial metabolic output, we measured levels of SCFAs in stool samples collected at all TPs. We used machine learning to assess whether SCFAs levels were predictive of *AMY1*-CN groups or SAA. Using the AMY1H′ and AMY1L′ group assignations, we trained a random forest model with 80% of the SCFA measures to predict the *AMY1* group to which the remaining 20% belonged. The model achieved an accuracy of 70.37% with an MCC of 40.74% and an AUC of 77.8% (Figure S7). In agreement with *AMY1*-CN being positively correlated with SAA, there was a trend for SCFA concentrations to be higher in the AMY1H′ ($CN > 6$) group.

Next, we performed a similar analysis in which we asked if SCFAs could be used to predict subjects with high or low mean SAA measurements. We determined high and low SAA groups by performing k-means clustering for all 25 subjects; all observations from the same subject were labeled with the subject's SAA group. We achieved an accuracy of 83.61% with an MCC of 64.46% and an AUC of 85.03% (Figure 7C). The total concentration of SCFAs was the most informative element for discriminating the high and the low SAA groups, followed by the concentrations of butyrate, valerate, propionate, and acetate (Figure 7D). Using a linear mixed model that included SAA group as a covariate, we confirmed that the concentrations of the SCFAs were higher in subjects with high SAA (adjusted p values: total SCFA concentration = 4.7×10^{-2}, acetate = 6.5×10^{-2}, propionate = 6.5×10^{-2}, and butyrate = 4.7×10^{-2}; valerate, isovalerate, heptanoate, and hexanoate were not significantly different). Assuming equal uptake of SCFAs in the colon across subjects, these results suggest that the higher the host SAA, the greater the SCFA production in the colon. Given that SAA can vary from day to day for a given individual, the observation that SAA is a better predictor of SCFAs than *AMY1*-CN indicates that the microbiome's metabolic output is sensitive to daily SAA variation.

Fecal Transplants from AMY1H Donors into Germ-Free Mice Promote Greater Adiposity Than Those from AMY1L Donors

To gauge differences in function for AMY1H and AMY1L gut microbiomes, we inoculated fecal samples obtained from AMY1H and AMY1L donors, sampled at 5 TPs, into 96 male Swiss-Webster 4- to 6-week-old germ-free mice fed a polysaccharide-rich chow *ad libitum* and single-caged post-transfer. Adiposity was assessed by DEXA after 4–6 weeks. Across all mice, we observed a significantly higher body fat percentage for recipients of the AMY1H compared to the AMY1L microbiomes (linear mixed model; $p = 0.026$). Post hoc pairwise comparisons revealed that TPs 3, 7, and 10 showed a significantly higher final adiposity for the AMY1H compared to the AMY1L treatment groups (Tukey's honest significant difference [HSD] test adjusted $p < 0.05$), whereas TPs 4 and 9 did not (after controlling for weight on the day of inoculation and length of the experiment; Figures 7E–7I). Food intake was not significantly different between high versus low *AMY1*-CN donor groups, and there were no differences in intestinal inflammation (measured by Lipocalin 2 at the end of the experiments using samples from TPs 3 and 10; data not shown). Thus, the AMY1H microbiomes generally drove higher adiposity gains that were unrelated to food intake and metabolic inflammation.

DISCUSSION

Here, we show that variation in the CN of the human salivary amylase gene *AMY1* influences the diversity and function of the human oral and gut microbiome. *AMY1*-CN shapes the carbohydrate milieu in the gut through its dose-dependent effect on salivary amylase production. We observed a strong effect on the composition of the oral microbiome and compositional and functional effects on the gut microbiome that are consistent with the type of complex carbohydrate depletion expected from the host genotype.

We observed that host *AMY1*-CN impacted the oral microbiome in a diet-independent manner. Our month-long intervention study included a 2-week diet standardization that ensured subjects consumed carbohydrate-rich foods daily. The oral microbiomes, first to experience food intake, were unaffected by the standardized diet period. They were, however, sensitive to host *AMY1*-CN status and to SAA. Notably, compared to AMY1L, AMY1H saliva microbiomes exhibited higher proportions of *Porphyromonas spp.* (e.g., *P. endodontalis*). Several members of the genus *Porphyromonas*, including *P. endodontalis*, have been associated with periodontitis (Socransky et al., 1998; Park

Figure 5. Gut Microbiomes Differ between *AMY1*-CN Groups at the OTU Level

(A) OTUs, identified using machine learning, that distinguish AMY1H and AMY1L gut microbiomes. See legend for Figure 3.

(B) Ribbon plot showing the OTUs identified using HARVEST that distinguish AMY1H and AMY1L groups at each time point. See legend for Figure 3. In addition, samples collected at the time points in bold print were also subjected to shotgun metagenomics analysis. See also Figure S5 and Tables S1 and S5.

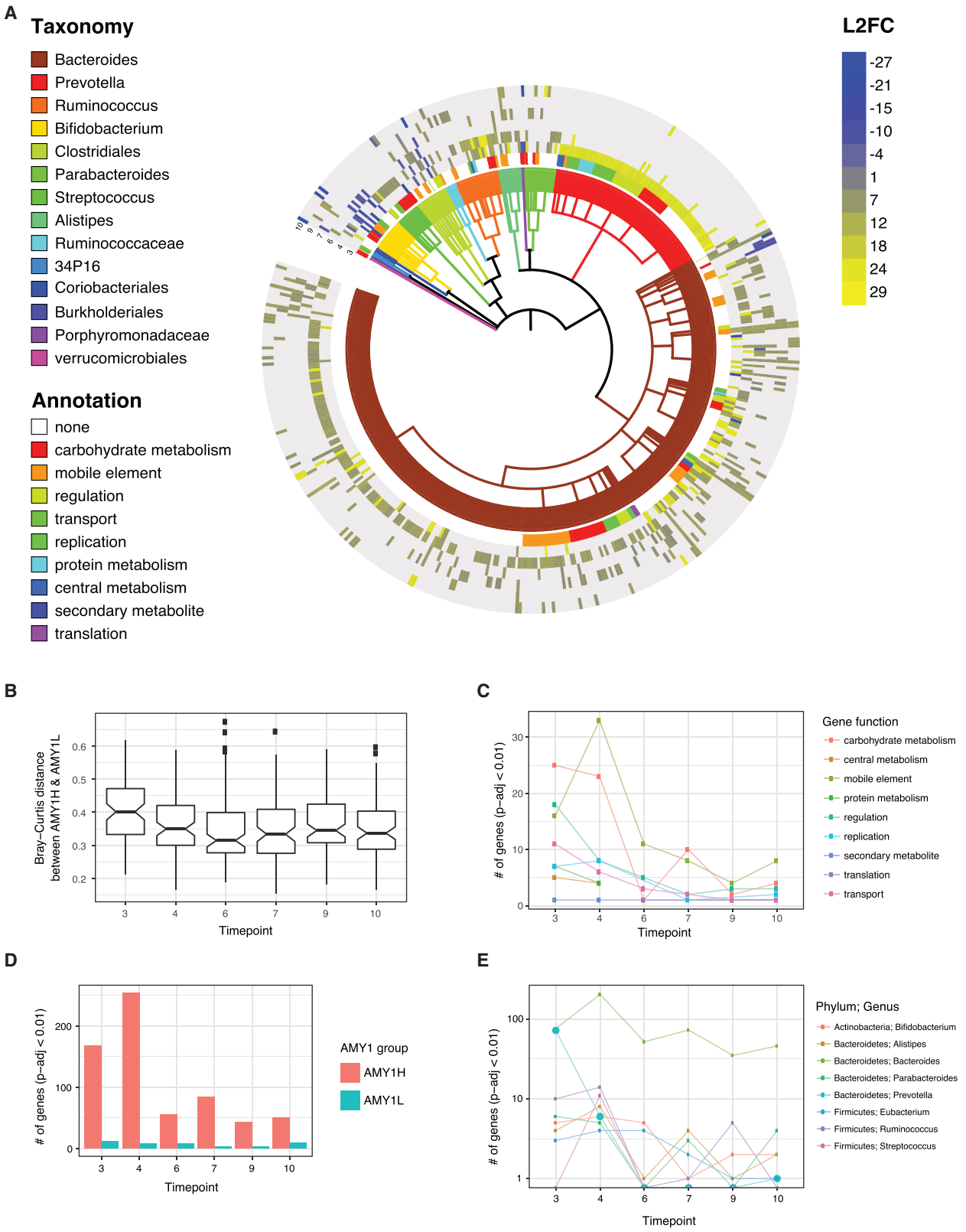


Figure 6. Metagenomes Indicate Convergence at the Gene Level

(A) Heat map displaying each of 481 gene families with abundances differing between AMY1H and AMY1L at one or more of six different TPs (n = 20). The heat map is sorted by taxonomy, annotation group, and gene family. Each concentric circle in the heat map corresponds to a TP. Gene family abundances with

(legend continued on next page)

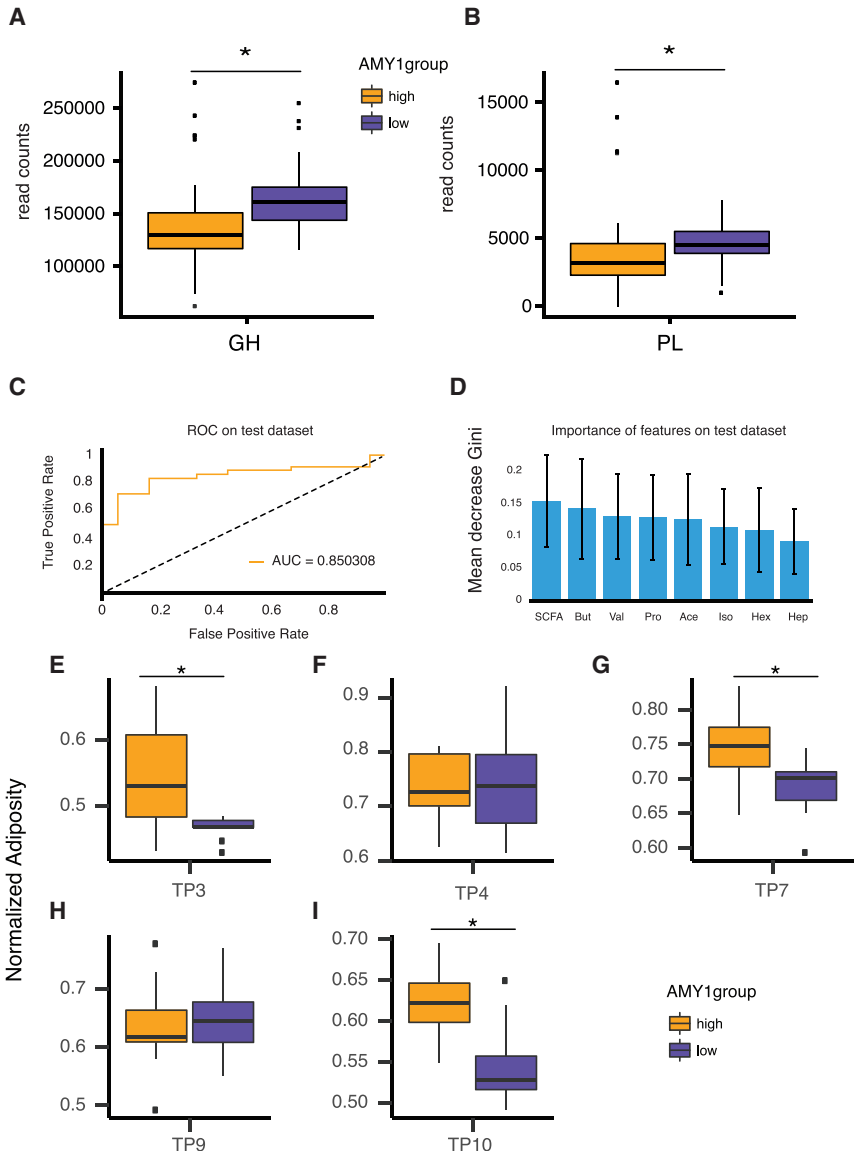


Figure 7. AMY1H and AMY1L Gut Microbiomes Differ in Function

(A and B) Boxplots of the read counts in the AMY1H and AMY1L groups for each of the significantly different carbohydrate-active enzyme classes: glycoside hydrolases (GH; A) and polysaccharide lyases (PL; B) (n = 20).

(C) We used machine learning to assess whether SCFA levels were predictive of SAA. Subjects were clustered according to their salivary amylase activity in two groups (SAA-H and SAA-L) by k-means clustering (n = 25). Shown here is the receiver operating characteristic (ROC) curve of a random forest model used to predict the SAA group using the SCFA measurements in a test dataset.

(D) The features used by the random forest model for the classification of the test samples are shown in decreasing order of importance given by the mean decrease in Gini index. SCFA is the total of all SCFAs. But, butyrate; Val, valerate, Pro, propionate; Ace, acetate; Iso, isovalerate; Hex, hexanoate; Hep, heptanoate.

(E–I) Boxplots of the adiposity measure normalized by baseline weight on the day of inoculation of formerly germ-free mice after humanization with stool collected from the study participants at 5 time points: (E) TP3; (F) TP4; (G) TP7; (H) TP9; (I) TP10. *Tukey's HSD adjusted $p < 0.05$. See also Figures S6 and S7.

et al., 2015; Colombo et al., 2012; Wade, 2013; Griffen et al., 2012; Cao et al., 2012; Lombardo Bedran et al., 2012). Many of the same taxa were discriminatory for host SAA, across all subjects and TPs assayed.

In contrast to the oral microbiomes, the standardized diet period had a noticeable effect on gut microbiomes, particularly at the functional level. Diet-driven convergence of gut microbiomes has been observed previously (Muegge et al., 2011; Clayton et al., 2016; Minot et al., 2011). Diet onset saw a spike in mobile-gene elements similar to what has been described for stress responses (Huddleston, 2014; Lerner et al., 2017). Despite the convergence, differences in gut microbiomes between high and low *AMY1*-CN groups were maintained.

Members of the Ruminococcaceae were enriched in gut microbiomes of ~50 subjects with high predicted *AMY1*-CN and in the intervened population. Members of the Ruminococcaceae family have been reported to ferment resistant starch (Walker et al., 2011; Salonen et al., 2014; Herrmann et al., 2017; Morais et al., 2016; Ze et al., 2012). High *AMY1*-CN hosts may preferentially select for members of the Ruminococcaceae because of an enrichment of resistant starch in chyme.

Host *AMY1*-CN was also related to the functional capacity of the gut microbiome. Our metagenomic analysis revealed enrichment in the AMY1L gut microbiomes of two classes of carbohydrate-active enzymes involved in the breakdown of overall complex carbohydrates, GH, and polysaccharide lyases. These results suggest that for a given diet, the AMY1L distal gut microbiota may be presented with a greater load of complex carbohydrates in general than the AMY1H microbiota. However, as a

significant differences between AMY1 groups were identified using DESeq2, and the $\log_2$ fold difference between AMY1H relative to AMY1L is depicted in the heat map. Higher abundances of gene families in the AMY1H group are colored yellow, while those higher in AMY1L are colored blue. Only gene families with a BH-adjusted $p < 0.01$ and that were assigned taxonomy are shown.

(B) Bray-Curtis distance between AMY1H and AMY1L metagenome samples (reads mapped to gene families; no taxonomic designation).

(C) Number of significantly enriched gene families that could be grouped by function.

(D) Number of significantly enriched gene families at each time point.

(E) Taxonomy of significantly enriched gene families. See also Table S4.

result of greater average host SAA, resistant starch is a greater proportion of the complex carbohydrates in the AMY1H colon, and the corresponding fermenters are proportionally more abundant.

We assessed functional output of AMY1H and AMY1L microbiomes with measures of SCFAs. SCFAs are fermentation products of distal gut microbiota, and their levels in stool are influenced both by production by the microbiome and uptake by the host. We observed that SCFAs in stool were associated more strongly with host SAA levels than with host *AMY1*-CN. Within an individual, SAA varies from day to day. Although *AMY1*-CN is a good predictor of average SAA, it may be a poor predictor of SAA on any given day. A better association of SCFAs with SAA than to *AMY1*-CN indicates that fecal SCFA pools reflect short-term fermentation dynamics in the gut that are affected by fluctuating SAA. Since microbiota known to ferment resistant starch (e.g., Ruminococcaceae) are enriched in the AMY1H subjects, the activity of these microbiota may be driving the higher levels of SCFAs in their stool (Topping and Clifton, 2001).

Another way we tested the functional capacity of the gut microbiomes was to transfer fecal microbiota to germ-free mice and assess adiposity differences for mice receiving microbiomes of AMY1H compared to those receiving AMY1L microbiomes. We transplanted human microbiomes 1:1 into mice and used 5 samples per subject.

The 5 transfer experiments are not exact replicates because we used five separate samples collected at different TPs from each donor, which takes into account daily variability in microbiomes. Overall, we observed a greater adiposity for mice recipients of microbiomes derived from the AMY1H donors. Thus, the daily fluctuation in the SAA and the microbiomes was reflected in the variance in functional output. Our findings suggest that the mice (*AMY1*-CN = 2) consumed a diet rich in complex carbohydrates that human AMY1H-conditioned microbiomes may have accessed better. Within their native human hosts, however, AMY1H microbiomes may not behave the same way since they are not decoupled from their high SAA environment, the way they are when transferred to germ-free mice.

Prospectus

Selection for duplication at the *AMY1* locus and lactase persistence evolved around the Neolithic transition to an agrarian lifestyle approximately 10,000 years ago. The strongest association to date between the gut microbiota and human genetic variation is related to the lactase persistence genotype (Goodrich et al., 2016; Bonder et al., 2016; Blekhman et al., 2015). This study of amylase CN further underscores how adaptation to new diets that drove human genetic variation across populations also underlies differences in modern-day microbiomes. In addition to microbiome composition, the genetic differences between individuals that impact how common foods are digested may therefore be important to take into consideration in the emerging area of personalized nutrition.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human Studies
  - Germfree Mouse Studies
- METHOD DETAILS
  - Initial *AMY1*-CN Screen
  - AMY1-CN Confirmation
  - *AMY2* CN Determination
  - Study Design and Sample Collection
  - Dietary Intervention and Intake Analysis
  - Salivary Amylase Activity
  - Stool Sample Processing
  - Stool Amylase Activity
  - *AMY2* ELISA
  - Short Chain Fatty Acid Measurements
  - Community Composition Based on 16S rRNA Genes
  - Metagenomics Library Sequencing and Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Estimation of *AMY1* CN in British Population
  - Dietary Intervention in Cornell Population
  - Macronutrient Intake Analysis
  - Salivary Amylase Activity Linear Mixed Models
  - Effects of Diet on Distance Metrics
  - Alpha Diversity Linear Mixed Models
  - OTUs with Differential Relative Abundance
  - Stool amylase Activity Linear Mixed Models
  - Metagenomics Analysis
  - Carbohydrate-Active enZYmes Analysis
  - Determination of SAA Groups
  - OTUs Distinguishing SAA Groups
  - Predicting SAA with SCFA Levels in Stool
  - Assessment of Adiposity in Mice
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chom.2019.03.001.

ACKNOWLEDGMENTS

We thank Brenda Daniels-Tibke, Kacie Harrington, Fiona Hoi Yi Chan, Sha Li, Noah Clark, M. Elizabeth Bell, Wei Zhang, Tim DeMarsh, Julia Hildebrandt, Sri Devarakonda, Dorothy Kim, and Vivianna Sanchez. This work utilized Core Services supported by grant DK097153 of the NIH (United States) to the University of Michigan. This work was supported by a David and Lucile Packard Foundation (United States) Fellowship (R.E.L.), the Arnold and Mabel Beckman Foundation (United States) (R.E.L.), a National Science Foundation (United States) Graduate Fellowship (J.K.G.), The Hartwell Foundation (United States) (R.E.L.), and the Max Planck Society (Germany).

AUTHOR CONTRIBUTIONS

R.E.L. supervised the study. R.E.L. and A.C.P. designed the study and oversaw sample collection. A.C.P., J.L.S., and Q.S. generated data. A.C.P. performed the statistical analyses with contributions from J.K.G., L.M.J., N.D.Y., A.R., G.G.L., H.Y.B., M.E.-H., D.H.H., R.E.L., and J.G.B. A.C.P., J.L.W., J.K.G, and Q.S. performed microbiota transfer experiments. A.C.P., J.K.G., N.D.Y., A.R., G.G.L., and R.E.L. prepared the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 3, 2018
Revised: September 10, 2018
Accepted: March 1, 2019
Published: April 10, 2019

REFERENCES

Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., et al. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comp. Biol. *8*, e1002358.

Bar, H.Y., Booth, J.G., and Wells, M.T. (2014). A bivariate model for simultaneous testing in bioinformatics data. J. Am. Stat. Assoc. *109*, 537–547.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4, arXiv. arXiv, 1406.5823 http://arxiv.org/abs/1406.5823.

Lombardo Bedran, T.B., Marcantonio, R.A., Spin Neto, R., Alves Mayer, M.P., Grenier, D., Spolidorio, L.C., and Spolidorio, D.P. (2012). Porphyromonas endodontalis in chronic periodontitis: a clinical and microbiological cross-sectional study. J. Oral Microbiol. *4*.

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. Genome Biol. *16*, 191.

Bonder, M.J., Kurilshikov, A., Tigchelaar, E.F., Mujagic, Z., Imhann, F., Vila, A.V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S.P., et al. (2016). The effect of host genetics on the gut microbiome. Nat. Genet. *48*, 1407–1412.

Bonnefond, A., Yengo, L., Dechaume, A., Canouil, M., Castelain, M., Roger, E., Allegaert, F., Caiazzo, R., Raverdy, V., Pigeyre, M., et al. (2017). Relationship between salivary/pancreatic amylase and body mass index: a systems biology approach. BMC Med. *15*, 37.

Britten, P. (2013). SuperTracker incorporates food composition data into innovative online consumer tool. Procedia Food Sci. *2*, 172–179.

Burnett, G.S. (1962). The effect of damaged starch, amylolytic enzymes, and proteolytic enzymes on the utilisation of cereals by chickens. Br. Poult. Sci. *3*, 89–103.

Cao, H., Qi, Z., Jiang, H., Zhao, J., Liu, Z., and Tang, Z. (2012). Detection of *Porphyromonas endodontalis, Porphyromonas gingivalis* and *Prevotella intermedia* in primary endodontic infections in a Chinese population. Int. Endod. J. *45*, 773–781.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. USA *108*, 4516–4522.

Carpenter, D., Dhar, S., Mitchell, L.M., Fu, B., Tyson, J., Shwan, N.A.A., Yang, F., Thomas, M.G., and Armour, J.A.L. (2015). Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. Hum. Mol. Genet. *24*, 3472–3480.

Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., GTEx Consortium, et al. (2017). The impact of structural variation on human gene expression. Nat. Genet. *49*, 692–699.

Clayton, J.B., Vangay, P., Huang, H., Ward, T., Hillmann, B.M., Al-Ghalith, G.A., Travis, D.A., Long, H.T., Tuan, B.V., Minh, V.V., et al. (2016). Captivity humanizes the primate microbiome. Proc. Natl. Acad. Sci. USA *113*, 10376–10381.

Colombo, A.P., Bennet, S., Cotton, S.L., Goodson, J.M., Kent, R., Haffajee, A.D., Socransky, S.S., Hasturk, H., Van Dyke, T.E., Dewhirst, F.E., et al. (2012). Impact of periodontal therapy on the subgingival microbiota of severe periodontitis: comparison between good responders and individuals with refractory periodontitis using the human oral microbe identification microarray. J. Periodontol. *83*, 1279–1287.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al. (2010). Origins and functional impact of copy number variation in the human genome. Nature *464*, 704–712.

Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2015). Genome-wide association studies of the human gut microbiota. PLoS One *10*, e0140301.

Davenport, E.R., Goodrich, J.K., Bell, J.T., Spector, T.D., Ley, R.E., and Clark, A.G. (2016). ABO antigen and secretor statuses are not associated with gut microbiota composition in 1,500 twins. BMC Genomics *17*, 941.

Falchi, M., El-Sayed Moustafa, J.S., Takousis, P., Pesce, F., Bonnefond, A., Andersson-Assarsson, J.C., Sudmant, P.H., Dorajoo, R., Al-Shafai, M.N., Bottolo, L., et al. (2014). Low copy number of the salivary amylase gene predisposes to obesity. Nat. Genet. *46*, 492–497.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014). Human genetics shape the gut microbiome. Cell *159*, 789–799.

Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., and Ley, R.E. (2016). Genetic determinants of the gut microbiome in UK twins. Cell Host Microbe *19*, 731–743.

Goodrich, J.K., Davenport, E.R., Clark, A.G., and Ley, R.E. (2017). The relationship between the human genome and microbiome comes into view. Annu. Rev. Genet. *51*, 413–433.

Gracia, M.I., Araníbar, M.J., Lázaro, R., Medel, P., and Mateos, G.G. (2003). Alpha-amylase supplementation of broiler diets based on corn. Poult. Sci. *82*, 436–442.

Griffen, A.L., Beall, C.J., Campbell, J.H., Firestone, N.D., Kumar, P.S., Yang, Z.K., Podar, M., and Leys, E.J. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. ISME J. *6*, 1176–1185.

Herrmann, E., Young, W., Rosendale, D., Conrad, R., Riedel, C.U., and Egert, M. (2017). Determination of resistant starch assimilating bacteria in fecal samples of mice by in vitro RNA-based stable isotope probing. Front. Microbiol. *8*, 1331.

Huddleston, J.R. (2014). Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. Infect. Drug Resist. *7*, 167–176.

Iskow, R.C., Gokcumen, O., and Lee, C. (2012). Exploring the role of copy number variants in human adaptation. Trends Genet. *28*, 245–257.

Jiang, H., Lei, R., Ding, S.-W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics *15*, 182.

Jo, J.K., Ingale, S.L., Kim, J.S., Kim, Y.W., Kim, K.H., Lohakare, J.D., Lee, J.H., and Chae, B.J. (2012). Effects of exogenous enzyme supplementation to corn- and soybean meal-based or complex diets on growth performance, nutrient digestibility, and blood metabolites in growing pigs. J. Anim. Sci. *90*, 3041–3048.

Kelley, J.L., and Swanson, W.J. (2008). Positive selection in the human genome: from genome scans to biological significance. Annu. Rev. Genomics Hum. Genet. *9*, 143–160.

Lerner, A., Matthias, T., and Aminov, R. (2017). Potential effects of horizontal gene exchange in the human gut. Front. Immunol. *8*, 1630.

Lim, M.Y., You, H., Yoon, H., Kwon, B., Lee, J.Y., Lee, S., Song, Y.-M., Lee, Kayoung, Sung, J., and Ko, G. (2017). The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. Gut *66*, 1031–1038.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The Carbohydrate-Active Enzymes Database (CAZy) in 2013. Nucleic Acids Res. *42*, D490–D495.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl. Environ. Microbiol. *73*, 1576–1585.

Macfarlane, G.T., and Englyst, H.N. (1986). Starch utilization by the human large intestinal microflora. J. Appl. Bacteriol. *60*, 195–201.

Magurran, A.E. (2004). Measuring Biological Diversity (Blackwell Publishing).

Mandel, A.L., Peyrot des Gachons, C., Plank, K.L., Alarcon, S., and Breslin, P.A.S. (2010). Individual differences in amy1 gene copy number, salivary α-amylase levels, and the perception of oral starch. PLoS One *5*, e13352.

Marcovecchio, M.L., Florio, R., Verginelli, F., De Lellis, L., Capelli, C., Verzilli, D., Chiarelli, F., Mohn, A., and Cama, A. (2016). Low AMY1 gene copy number is associated with increased body mass index in prepubertal boys. PLoS One *11*, e0154961.

Mejía-Benítez, M.A., Bonnefond, A., Yengo, L., Huyvaert, M., Dechaume, A., Peralta-Romero, J., Klünder-Klünder, M., García Mena, J., El-Sayed Moustafa, J.S., Falchi, M., et al. (2015). Beneficial effect of a high number of copies of salivary amylase AMY1 gene on obesity risk in Mexican children. Diabetologia *58*, 290–294.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. *21*, 1616–1625.

Moraïs, S., Ben David, Y., Bensoussan, L., Duncan, S.H., Koropatkin, N.M., Martens, E.C., Flint, H.J., and Bayer, E.A. (2016). Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, ruminococcus champanellensis, reveals a fine-tuned system for cohesin-dockerin recognition. Environ. Microbiol. *18*, 542–556.

Moriyoshi, Y., Takeuchi, T., Shiratori, K., and Watanabe, S.-I. (1991). Fecal isoamylase activity in patients with pancreatic diseases. Pancreas *6*, 70–76.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., and Gordon, J.I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science *332*, 970–974.

Park, O.J., Yi, H., Jeon, J.H., Kang, S.S., Koo, K.T., Kum, K.Y., Chun, J., Yun, C.H., and Han, S.H. (2015). Pyrosequencing analysis of subgingival microbiota in distinct periodontal conditions. J. Dent. Res. *94*, 921–927.

Pedersen, H.K., Gudmundsdottir, V., Nielsen, H.B., Hyotylainen, T., Nielsen, T., Jensen, B.A.H., Forslund, K., Hildebrand, F., Prifti, E., Falony, G., et al. (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. Nature *535*, 376–381.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., et al. (2007). Diet and the evolution of human amylase gene copy number variation. Nat. Genet. *39*, 1256–1260.

Perry, G.H., Kistler, L., Kelaita, M.A., and Sams, A.J. (2015). Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. J. Hum. Evol. *79*, 55–63.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215.

Rukh, G., Ericson, U., Andersson-Assarsson, J., Orho-Melander, M., and Sonestedt, E. (2017). Dietary starch intake modifies the relation between copy number variation in the salivary amylase gene and BMI. Am. J. Clin. Nutr. *106*, 256–262.

Salonen, A., Lahti, L., Salojärvi, J., Holtrop, G., Korpela, K., Duncan, S.H., Date, P., Farquharson, F., Johnstone, A.M., Lobley, G.E., et al. (2014). Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. ISME J. *8*, 2218–2230.

Socransky, S.S., Haffajee, A.D., Cugini, M.A., Smith, C., and Kent, R.L., Jr. (1998). Microbial complexes in subgingival plaque. J. Clin. Periodontol. *25*, 134–144.

Sonnenburg, J.L., and Bäckhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. Nature *535*, 56–64.

Southwood, T.R.E., and Henderson, P.A. (2009). Ecological Methods (John Wiley & Sons).

Topping, D.L., and Clifton, P.M. (2001). Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiol. Rev. *81*, 1031–1064.

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M.S., Kevans, D., Smith, M.I., Guttman, D.S., Griffiths, A., Panaccione, R., Otley, A., et al. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. Nat. Genet. *48*, 1413–1417.

Usher, C.L., Handsaker, R.E., Esko, T., Tuke, M.A., Weedon, M.N., Hastie, A.R., Cao, H., Moon, J.E., Kashin, S., Fuchsberger, C., et al. (2015). Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. Nat. Genet. *47*, 921–925.

Viljakainen, H., Andersson-Assarsson, J.C., Armenio, M., Pekkinen, M., Pettersson, M., Valta, H., Lipsanen-Nyman, M., Mäkitie, O., and Lindstrand, A. (2015). Low copy number of the AMY1 locus is associated with early-onset female obesity in Finland. PLoS One *10*, e0131883.

Wade, W.G. (2013). The oral microbiome in health and disease. Pharmacol. Res. *69*, 137–143.

Walker, A.W., Ince, J., Duncan, S.H., Webster, L.M., Holtrop, G., Ze, X., Brown, D., Stares, M.D., Scott, P., Bergerat, A., et al. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J. *5*, 220–230.

Walter, J., and Ley, R. (2011). The human gut microbiome: ecology and recent evolutionary changes. Annu. Rev. Microbiol. *65*, 411–429.

Yong, R.Y., Mustaffa, S.B., Wasan, P.S., Sheng, L., Marshall, C.R., Scherer, S.W., Teo, Y.Y., and Yap, E.P. (2016). Complex copy number variation of amy1 does not associate with obesity in two East Asian cohorts. Hum. Mutat. *37*, 669–678.

Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012). *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. ISME J. *6*, 1535–1543.

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., et al. (2015). Personalized nutrition by prediction of glycemic responses. Cell *163*, 1079–1094.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
DNA from human lymphoblastoid cell line	Coriell Institute for Medical Research: NHGRI Sample Repository for Human Genetic Research	NA18972
DNA from *Pan troglodytes* cell line	Coriell Institute for Medical Research: NIGMS Human Genetic Cell Repository	NS06006
DNA from human lymphoblastoid cell line	Coriell Institute for Medical Research: NIGMS Human Genetic Cell Repository	NA10472
Saliva and stool samples from 25 participants in month-long study with dietary intervention	This paper	Not applicable
16S rRNA gene data from 1000 TwinsUK subjects	Goodrich et al. (2016)	European Bioinformatics Institute accession number: ERP015317
Human genome SNP data from 1000 TwinsUK subjects	Goodrich et al. (2016)	http://twinsuk.ac.uk/resources-for-researchers/access-our-data/
16S rRNA gene data from the intervention study	This paper	European Bioinformatics Institute accession number: PRJEB27304, Secondary accession: ERP109371
Metagenomic sequence data from the intervention study	This paper	European Bioinformatics Institute accession number: PRJEB27308, Secondary accession: ERP109376
Critical Commercial Assays		
α-Amylase kinetic enzyme assay kit	SALIMETRICS	Cat# 1-1902
ELISA assay specific for *AMY2*	ALPCO	Cat# K 6410
Experimental Models: Organisms/Strains		
Germ Free Swiss Webster male mice	Taconic	SW-M
Oligonucleotides		
AMY1-forward	designed by AC Poole	5'-TGAGAACATTAGGCCACAGCA-3'
AMY1-reverse	designed by AC Poole	5'-TGGAAATCATCTCAATGACCTCT-3'
EIF2B2-forward	designed by AC Poole	5'-GCTCAAAGTGCTTGAGGACC-3'
EIF2B2-reverse	designed by AC Poole	5'-CAAAGCCAAACCCAGACAAT-3'
AMY2A-forward	designed by AC Poole	5'-TGGCGATGGGTTGATATTGCT-3'
AMY2A-reverse	designed by AC Poole	5'-ACAAGCACAGTGAATTCCGC-3'
AMY2B-forward	designed by AC Poole	5'-ACTAATGACCTGTGTTATACTTCCT-3'
AMY2B-reverse	designed by AC Poole	5'-AGCTGTTACGCACAGTTCCA-3'
Taqman Copy Number Assay for *AMY1*, and reference assay for RNase P	Life Technologies	Id Hs07226361_cn and Human, 4403326
Software and Algorithms		
SuperTracker, dietary intake analysis software	Britten (2013)	Discontinued
R, version 3.1.2	Bates et al. (2014)	https://www.r-project.org/
Harvest, bivariate model used to identify OTUs with differential abundances	Bar et al. (2014)	Not applicable
DESeq2	Love et al. (2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ruth E. Ley (rley@tuebingen.mpg.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Studies

We recruited volunteers affiliated with the Cornell University campus in Ithaca, New York by way of flyers and listservs. We used the following exclusion criteria: BMI $\geq$ 35 or $\leq$ 18; age < 18 or > 40; usage within the last six months of amylase inhibitor, systemic antibiotic, corticosteroid, immunosuppressant agent, or probiotic supplement; gastrointestinal disorders. All human-related procedures, sample, and data collection were approved by the Cornell University Institutional Review Board, protocol number 1106002281. The eligibility criteria are listed as Data S1. Informed consent was obtained from all participants.

Germfree Mouse Studies

All mouse-related protocols, sample, and data collection were approved by the Cornell University Institutional Animal Care and Use Committee, protocol number 2010-0065. We inoculated germfree Swiss Webster adult male mice between 4-6.5 weeks of age with stool samples collected at 5 TPs (3, 4, 7, 9, and 10) (Table S2). Each mouse experiment used stool samples from all 9 AMY1L donors. TP 3 had 11 AMY1H donors while all other TPs used sample from 10 AMY1H donors.

At the beginning of each experiment, mice were weight-ranked and then assigned alternately to an AMY1L or AMY1H donor to ensure that the mean weight on the day of inoculation was not different between the 2 groups. Stool suspension was prepared in a Coy anaerobic chamber. Approximately 500 mg frozen stool was solubilized in 10ml of anoxic PBS that contained 2 mM dithiothreitol as a reducing agent, and vortexed at 5 minute intervals until no soluble clumps were visible. Each mouse was orally gavaged with 200 μl stool suspension from one human subject and single-housed. After inoculation, mice were maintained on autoclaved water and autoclaved Teklad diet 7017, NIH-31 (Harlan Laboratories) and kept under a 12-hour light/dark cycle for 33-40 days (TP 3: 40 days; TP 4: 35 days; TP 7: 35 days; TP 9: 33 days; and TP 10: 35 days.) Mouse weight and chow consumption were recorded weekly. At the end of each experiment, mice were DEXA scanned (Lunar PIXImus Mouse, GE Medical Systems, Waukesha, WI) to measure adiposity.

METHOD DETAILS

Initial *AMY1*-CN Screen

We screened 105 individuals on the Cornell University campus for *AMY1*-CN. We collected buccal cells by instructing subjects to swab the inside of their cheek with Epicentre Catch-All Sample Collection Swabs. Genomic DNA was extracted using the Qiagen Gentra Puregene Buccal Cell Kit. qPCR was performed using the same primer sequences for *AMY1* and *TP53* designed by Perry et al. with the following conditions: 2ng of DNA was used in 25μl reactions with Applied Biosystems Power SYBR Green PCR Master Mix on a BioRad MyiQ iCycler Single Color Real-Time PCR Detection System (Perry et al., 2007). The PCR protocol was as follows: initial denaturation at 95°C for 10 minutes and 40 cycles of 95°C for 15 seconds followed by 58°C for 30 seconds. All reactions, including standards, were performed in triplicate. These primers are known to anneal to both human and chimp *AMY1* gene sequence. Two control sample DNAs with known copy number were run on every plate. The control samples, NS06006: chimp DNA with *AMY1*-CN of 2 and NA18972: human DNA with *AMY1*-CN of 18 (Carpenter et al., 2015; Perry et al., 2007), were purchased from the Coriell Institute for Medical Research NIGMS Human Genetic Cell Repository and NHGRI Sample Repository for Human Genetic Research. To calibrate each subject's CN, the ratio of *AMY1* to *TP53* levels was used in a line equation created using the ratios from the two control DNA samples resulting in an adjusted copy number value.

AMY1-CN Confirmation

A second cheek swab was taken from each subject to confirm the *AMY1*-CN. Genomic DNA was isolated as described above. We performed qPCR after designing a new set of primers to amplify the *AMY1* paralogs,

AMY1-forward: 5'-TGAGAACATTAGGCCACAGCA-3' and AMY1-reverse:

5'-TGGAAATCATCTCAATGACCTCT-3'. We also designed primers to use *EIF2B2* as a reference gene, EIF2B2-forward: 5'-GCTCAAAGTGCTTGAGGACC-3', EIF2B2-reverse: 5'-CAAAGCCAAACCCAGACAAT-3'. Primer concentrations were at 0.5 μM, and 5ng of DNA template per reaction were used in 10μl reactions with Roche LightCycler 480 SYBR Green I Master mix. Both *AMY1* and *EIF2B2* were run on the same plate, and the PCR program was as follows: initial denaturation at 95°C for 5 minutes and 40 cycles of 95°C for 10 seconds followed by 60°C for 30 seconds on a Roche LightCycler 480 Real-Time PCR Instrument. A standard curve was made using DNA NA10472. Reactions were performed in triplicate, and the control DNAs NS06006 and NA18972 were run on every plate. To calibrate each subject's CN, the ratio of *AMY1* to *EIF2B2* levels was used in a line equation created using the ratios from the two control DNA samples resulting in an adjusted copy number value.

Digital PCR was performed using Life Technologies Taqman Copy Number Assay Id Hs07226361_cn for the *AMY1* locus and TaqMan Copy Number Reference Assay, RNase P, Human, 4403326 to normalize for total DNA. Reactions were run on a Life Technologies QuantStudio3D Digital PCR system in duplicate. In statistical analyses (below) we used the mean of the values generated by qPCR and by digital PCR for each subject as the *AMY1*-CN value for that subject.

AMY2 CN Determination

We determined the CN of the pancreatic amylase locus, *AMY2*, by performing qPCR with an independent primer pair for each paralog (*AMY2A*, and *AMY2B*), AMY2A-forward: 5'-TGGCGATGGGTTGATATTGCT-3', AMY2A-reverse: 5'-ACAAGCACAGT GAATTCCGC-3', AMY2B-forward:

5'-ACTAATGACCTGTGTTATACTTCCT-3', and AMY2B-reverse:

5'-AGCTGTTACGCACAGTTCCA-3'. We also used the aforementioned primers for *EIF2B2* as a reference gene on the same qPCR plate with each *AMY2* paralog. Primers were at 0.5 μM, and 2ng of DNA template per reaction were used in 10μl reactions with Roche LightCycler 480 SYBR Green I Master mix. *EIF2B2* was run on the same plate with each *AMY2* paralog, and the PCR program was as follows: initial denaturation at 95°C for 5 minutes and 40 cycles of 95°C for 10 seconds followed by annealing temperature for 10 seconds and 72°C for 15 seconds on a Roche LightCycler 480 Real-Time PCR Instrument. Annealing temperature was 58°C for *AMY2A* and 60°C for AMY2B. Reactions were performed in triplicate.

Study Design and Sample Collection

Based on the qPCR results above, we enrolled the 25 people at the upper and lower ends of the distribution to participate in a 4-week study. Metadata about the participants, including gender and age, are described in Table S3. The majority of the participants in this study were female. There are not enough participants of either gender to run the analyses restricted to one gender with the same statistical power. Previous work has suggested that person-to-person variation in microbiomes is not well explained by gender (Human Microbiome Project Consortium 2012). For the first and fourth weeks, participants were instructed to consume their usual diet and record all food and drink intake with approximate amounts. The dietary intervention occurred during weeks 2 and 3. On three days of each of the four weeks, the subjects provided stool and saliva samples for a total of 12 TPs. With few exceptions, all saliva samples were collected on the same day for every subject at each of the twelve TPs. Subjects were instructed to allow saliva to pool in the mouth for three minutes and then express through 5cm drinking straws into a 1.5ml eppendorf tube. Saliva was vortexed and aliquoted into two tubes and chilled on ice until stored at -80°C within 4 hours of collection. Stool samples were collected during a three-day window for each TP. One subject left the study after two weeks, thus only providing samples for TPs 1 through 6. On day 12 of the study, DEXA scanning was performed on 23 of the subjects using a Hologic DEXA, Model: DISCOVERY-A at Cornell University's Human Metabolic Research Unit.

Dietary Intervention and Intake Analysis

During weeks 2 and 3, participants were provided all meals and snacks from a menu designed by a registered dietetic technician. This diet featured healthy meals and snacks and a high-starch food item in every meal. The diets were designed to provide enough calories for an adult male. We collected daily caloric intake estimates from the participants before the beginning of the study. Based on this information, we were prepared to provide an additional portion of each meal as requested by participants. Participants did not all consume the entire volume supplied. During the dietary intervention, participants consumed one meal at Cornell University's Human Metabolic Research Unit dining room in the presence of lab personnel and took two meals away. A researcher from this project took weekday lunches with the participants, and made observations as to the way participants approached the food (e.g., ate very much, very little, avoidance, etc.) On Fridays, participants consumed one meal and took away all meals and snacks packaged for the weekend.

For each of the 25 subjects, individual dietary records of their reported intake for the entire 4 weeks of the study were entered into the diet analysis software SuperTracker (Britten, 2013), and individual reports were generated regarding macronutrient and micronutrient content (Data S2). SuperTracker food nutrition data is based on the Food and Nutrient Database for Dietary Studies (FNDDS), and the Food Patterns Equivalents Database (FPED), both from the USDA/ARS Food Surveys Research Group. The software reports categories entitled carbohydrate, dietary fiber, total sugars, and added sugars. Starch is analyzed using the AOAC method 966.11 or 979.10 (2012) or by a polarometric method (The Feedings Stuffs Regulations 1982), but there is not a separate category reported in the SuperTracker output. Total dietary fiber content is determined by enzymatic-gravimetric methods 985.29 or 991.43 of the AOAC (2012). However, the dietary fiber information provided does not distinguish between specific types of fiber including insoluble, soluble, resistant starches and non-starch polysaccharides.

Salivary Amylase Activity

Salivary amylase activity was measured for each saliva sample in triplicate using the SALIMETRICS α-Amylase kinetic enzyme assay kit (cat 1-1902) as per the instructions with one modification: Instead of 320μl, 300μl of pre-heated substrate was added to the sample. Reactions were performed in triplicate.

Stool Sample Processing

Subjects provided two stool sample aliquots from a single bowel movement in separate tubes and stored them in insulated bags containing frozen ice packs then stored at -80°C. One of these aliquots was later freeze dried prior to DNA extraction, while the other was saved for use in the amylase activity assay and germfree mouse inoculation studies. At each TP from each subject, one of the stool sample aliquots was freeze dried in a 50ml conical tube and then homogenized by roll-milling in the following manner. After

pressing three stainless steel rods (2 different sizes: 9 cm long x 0.9 cm diameter and 9 cm long x 0.3 cm diameter) into the freeze dried sample, the 50ml conical tubes were rolled on a Triple Gallon Tumbler (Covington, cat # 253TUM) for 24-48 hours.

Stool Amylase Activity

We used the SALIMETRICS α-Amylase kinetic enzyme assay kit (cat # 1-1902) to measure amylase activity in frozen samples collected during the human studies. We added approximately 150 mg of frozen stool to MOBIO garnet bead tubes with 0.70mm garnet beads (cat # 13123-50). We added enough Salimetrics kit diluent to obtain a concentration of 0.3g stool/ml diluent. Samples were placed in a BioSpec 1001 Mini-Beadbeater-96 for 2 minutes and then centrifuge at 1500 rcf for 15 minutes. We transferred the supernatant into an eppendorf tube and starting with 25μl of undiluted supernatant, performed three serial dilutions up to 1 in 200 using Salimetrics diluent. Then we proceeded with the assay as described above. We noted that 11 subjects (7 AMY1H and 4 AMY1L) had a median FAA< 10 U/g across all TPs.

AMY2 ELISA

We used an ELISA that employs two monoclonal antibodies to human pancreatic amylase as per the instructions (ALPCO, cat # K 6410). As a control, we used purified α-Amylase from *Bacillus licheniformis* (Krackeler Scientific, cat #A3403-500KU) in both the *AMY2* ELISA and amylase activity assays. Amylase activity from *Bacillus licheniformis* was detected in the amylase activity assay, but the ELISA assay specific for *AMY2* did not detect this microbial amylase.

Short Chain Fatty Acid Measurements

Short chain fatty acid quantification in stool samples was performed by the Metabolomics Core at the University of Michigan using cold extraction of short chain fatty acids, measured by EI GC-MS without derivatization on ~40-60 mg stool from all of the TPs. Short chain fatty acid measurements were normalized to the wet weight of the samples. The short chain fatty acids detected and quantified were acetate, butyrate, propionate, isovalerate, valerate, heptanoate, and hexanoate. Samples were run over two days. There was slight instrumental drift while running the second batch so the LOESS correction method was applied to those data.

Community Composition Based on 16S rRNA Genes

The saliva samples used in enzyme activity measures, from all TPs except 2, 7, and 10, and all fecal samples, were profiled for microbial community diversity and composition (Table S2). Microbial community DNA was extracted from the freeze dried stool and saliva samples using the MO BIO PowerSoil-htp Soil DNA Isolation Kit (MO BIO Laboratories, cat # 12955-4), but instead of vortexing, samples were placed in a BioSpec 1001 Mini-Beadbeater-96 for 2 minutes. We used 10-50 ng of sample DNA in duplicate 50 μl PCR reactions with 5 PRIME HotMasterMix and 0.1 μM forward and reverse primers. We amplified the V4 region of 16S using the universal primers 515F and barcoded 806R and the PCR program previously described (Caporaso et al., 2011) but with 25 cycles. We purified amplicons using the Mag-Bind E-Z Pure Kit (Omega Bio-tek, cat # M1380) and quantified with Invitrogen Quant-iT PicoGreen dsDNA Reagent, and 100 ng of amplicons from each sample were pooled and paired end sequenced (2x250bp) on an Illumina MiSeq instrument. Saliva samples from 3 of 12 TPs were not successfully processed although we did obtain measurements of salivary enzyme activity.

Sequence data were analyzed using the QIIME software package 1.9.0 (Caporaso et al., 2010). Briefly, paired ends were joined using fastq-join, and sequences were demultiplexed and filtered using a Phred quality score threshold of greater than or equal to 25. Open reference OTU picking was performed on all sequence data from oral and gut samples using the uclust method and the August, 2013 Greengenes 16S rRNA Gene Database as reference sequences. We used the QIIME 1.9.0 open reference OTU picking pipeline with all default parameters except the following: max_accepts = 20, max_rejects = 500, stepwords = 20, and word_length = 12. Samples with a sequence count below 10,000 were excluded from downstream analyses. After exclusion, the oral dataset consisted of 216 samples (sequencing was performed on 9 of the 12 TPs collected) yielding 16,030,493 sequences with a median sequence count of 72,107. The fecal data set included 293 samples with a total of 16,421,608 sequences and a median sequence count of 55,165 sequences per sample.

We calculated beta diversity using the unweighted and weighted UniFrac metrics on an OTU table containing 11,146 and 20,133 sequences per sample for the oral and gut datasets, respectively, and performed principal coordinates analysis on the distance matrices (Lozupone et al., 2007). We assessed alpha diversity using Chao 1, Observed Species, Faith's phylogenetic diversity, and Shannon's Index (Magurran, 2004; Southwood and Henderson, 2009) by calculating means from 100 iterations using a rarefaction of 11,146 and 16,848 sequences per sample for the oral and gut datasets, respectively.

Metagenomics Library Sequencing and Analysis

We performed metagenomic analysis on sequences generated from genomic microbial DNA obtained during the DNA extraction method detailed above using the freeze dried stool samples collected at TPs 3, 4, 6, 7, 9, and 10 (Table S2). We prepared metagenomic libraries using 1 ng of DNA input per sample into a Nextera XT DNA Sample Preparation Index Kit as per the instructions (Illumina, cat # FC-131-1096). After purification with Agencourt AMPure XP beads (Beckman Coulter, cat # A63882), samples were normalized and pooled with 20 samples per pool. Size selection was performed on the pools using BluePippin (Sage Sciences, cat # BDF1510) to restrict fragment sizes between 300 to 650 bp. Pools were run on an Illumina HiSeq3000 with 2x150 bp paired end sequencing for a sequencing depth of 14 ± 3.0 Gb (median ± standard deviation).

QUANTIFICATION AND STATISTICAL ANALYSIS

Estimation of *AMY1* CN in British Population

Included in the British genotype data were 7 of the 10 SNPs, rs 6696797, rs 10881197, rs 1999478, rs 11185098, rs 1930212, rs 1566154, and rs 1330403, previously correlated with *AMY1* CN (Usher et al., 2015). Only one randomly chosen twin per pair was included in our analysis. After excluding subjects with a BMI outside of that used to screen our Cornell population, we had data for 994 British subjects. Using the change in copy number values determined for the GoT2D cohort of 2,863 Europeans, we calculated the sum of the change in copy number values for each individual using the methods of Usher et al., wherein each of the seven alleles is associated with a difference in *AMY1* copy number of a given amount (Usher et al., 2015). For each person, we added the values in the table for the alleles present in that person's genome. We then selected only the tail ends of the distribution to include the lowest 5% and highest 5% of individuals for group sizes of 50 each.

Dietary Intervention in Cornell Population

We excluded the AMY1M group from all statistical analyses that include *AMY1* group as a covariate but included them in all analyses with *AMY1*-CN as a covariate. We used logarithm, square root or rank transformation as needed to better meet model requirements of homogeneous variance and normality of residuals, ε. All linear mixed models are described with the notation used in the statistical package lme4 in R, version 3.1.2 (Bates et al., 2014). The Scikit-Learn library in Python was used in the machine learning based modeling.

Macronutrient Intake Analysis

Dietary records from all subjects were manually entered into the nutritional analysis software SuperTracker, which produces nutrient reports that include estimated percentages of macro- and micronutrient content in the food items entered by the user. We fit linear mixed models using each macronutrient percentage as the response variable to determine whether or not dietary intake differed between the *AMY1* CN groups over time. We analyzed two separate models that either included study day or whether or not diet was being provided on that day:

$$y \sim \text{AMY1CNG} + \text{DAY} + \text{AMY1CNG:DAY} + (1|\text{SUBJECT}) + \varepsilon \quad [1]$$

$$y \sim \text{AMY1CNG} + \text{DIET} + \text{AMY1CNG:DIET} + (1|\text{SUBJECT}) + \varepsilon \quad [2]$$

y is the macronutrient percentage. Fixed effects included *AMY1* CN group (**AMY1CNG**) and study day (**DAY**) or whether or not diet was being provided on that day (**DIET**). We also included a random effects term for repeat sampling of subjects (1|**SUBJECT**).

Salivary Amylase Activity Linear Mixed Models

We analyzed two separate models that included either *AMY1* CN or *AMY1* CN group as a predictor:

$$y \sim \textbf{AMY1CNG} + \textbf{TP} + \textbf{AMY1CNG:TP} + (1|\textbf{SUBJECT}) + \varepsilon \quad [1]$$

$$y \sim \textbf{AMY1CN} + \textbf{TP} + \textbf{AMY1CN:TP} + (1|\textbf{SUBJECT}) + \varepsilon \quad [2]$$

y is the salivary amylase activity. Fixed effects included *AMY1* CN group (**AMY1CNG**) or *AMY1* CN (**AMY1CN**; included subjects in the AMY1M group), and TP (**TP**). We also included a random effects term for repeat sampling of subjects (1|**SUBJECT**).

Neither the interaction between *AMY1* CN group and TP nor *AMY1* CN and TP is significant and nor is the effect of TP significant in either model. The effect of both *AMY1* CN ($p = 2.1 \times 10^{-5}$) and *AMY1* CN group ($p = 1.9 \times 10^{-4}$) are significant based on an F-test with a Satterthwaite approximation.

Effects of Diet on Distance Metrics

For 16S rRNA gene sequence data, we compared UniFrac distances (both unweighted and weighted) between individuals in high and low *AMY1* groups during 3 time intervals: pre-diet, on the diet, and post-diet. For shotgun metagenomic sequences, we compared Bray Curtis distances between individuals in high and low *AMY1* groups using the gene family raw counts but only had data for TPs during two time intervals: pre-diet and on the diet. We calculated the non-parametric bootstrap confidence intervals for the difference in population means between the Bray-Curtis distances (just AMY1L versus AMY1H distance values) for pre-diet and during-diet time points. We determined the 95% bootstrap CIs based on 1000 permutations. This approach accounts for non-independence issues caused by repeat sampling from individuals.

Alpha Diversity Linear Mixed Models

For both stool and saliva samples, we used a linear mixed model to assess the effect of *AMY1*-CN or *AMY1*-CN group on alpha diversity:

$$y \sim \textbf{AMY1CN} + \textbf{TP} + \textbf{AMY1CN:TP} + (1|\textbf{SUBJECT}) + \varepsilon \quad [1]$$

$$y \sim \textbf{AMY1CNG} + \textbf{TP} + \textbf{AMY1CNG}{:}\textbf{TP} + (1|\textbf{SUBJECT}) + \varepsilon \quad [2]$$

y is the alpha diversity metric. Fixed effects included *AMY1* CN group (**AMY1CNG**; excluded AMY1M) or *AMY1* CN (**AMY1CN**; included subjects in the AMY1M group) and TP (**TP**). We also included a random effects term for repeat sampling of subjects (1|**SUBJECT**). Whenever the interaction term of the linear mixed model was significant, we identified the affected TPs by performing post-hoc pairwise comparisons between the TPs using Tukey's HSD method to adjust for multiple comparisons.

OTUs with Differential Relative Abundance

We used a bivariate model called Harvest (Bar et al., 2014) to identify OTUs with differential means or variances in relative abundance between the AMY1H and AMY1L groups at each TP separately. For this analysis we omitted OTUs not present in at least half of the samples in either the AMY1L or the AMY1H group at the TP being considered. We adjusted p values using the Benjamini-Hochberg procedure to account for all OTUs tested at a given TP.

Stool amylase Activity Linear Mixed Models

We used a linear mixed model to assess the effect of *AMY1* CN, *AMY2* CN, or *AMY1*

CN group on stool amylase activity using the following models:

$$y \sim \text{AMY2CN} + \text{TP} + \text{AMY2CN:TP} + (1|\text{SUBJECT}) + \varepsilon \quad [1]$$

$$y \sim \text{AMY1CN} + \text{TP} + \text{AMY1CN:TP} + (1|\text{SUBJECT}) + \varepsilon \quad [2]$$

$$y \sim \text{AMY1CNG} + \text{TP} + \text{AMY1CNG:TP} + (1|\text{SUBJECT}) + \varepsilon \quad [3]$$

y is the stool amylase activity. Fixed effects included *AMY1* CN (**AMY1CN**; included subjects in the AMY1M group), *AMY2* CN (**AMY2CN**; included subjects in the AMY1M group), or *AMY1* CN group (**AMY1CNG**; excluded AMY1M group), and TP (**TP**). We also included a random effects term for repeat sampling of subjects (1|**SUBJECT**).

In model [3] we determined that stool amylase activity at TP 6 was significantly greater than TP 12 by performing post-hoc pairwise comparisons between the TPs using Tukey's HSD method to adjust for multiple comparisons.

Metagenomics Analysis

For metagenomic read quality control, we used skewer v0.2.2 (Jiang et al., 2014) to trim the 3' ends until quality scores reached $\geq$15, and end-trimmed reads <100 bp were removed. The software skewer v0.2.2 was also used to remove Illumina adapter contamination, and bbmap v37.78 was used to filter out human host genome reads. Post-QC reads were subsampled at 20,000,000 paired-end reads per sample to normalize for sequencing depth and reduce downstream processing time. Across the entire read length used in the analysis, the median Phred score was greater than 30. We used the HMP Unified Metabolic Analysis Network (HUMAnN2 v0.11.1) pipeline to classify the reads against the ChocoPhlAn and UniRef90 databases (Abubucker et al., 2012). For our statistical analyses, we used the gene families file with the abundances normalized to reads per kilobase (RPKs). However, DESeq2, the software that we used to identify gene families with differential abundances between AMY1 groups, requires integer values, or counts that have not been normalized with respect to library size, as a requirement of the statistical model because the DESeq2 software adjusts for differences in library size internally (Love et al., 2014). To accommodate this requirement, we multiplied the RPKs reported by HUMAnN2 by the gene lengths and total number of reads in order to obtain integers unadjusted for library size. The gene families file produced by HUMAnN2 is stratified; for each gene family there are one or more rows with the first row being the total number of reads assigned to that gene family and the additional rows corresponding to the number of reads assigned to each of the different taxonomy, i.e. species, when known. Therefore, the raw data contains entries that are not independent. We removed the entry reporting the sum of the mappings assigned to the gene family prior to analysis and kept the mappings to species including unclassified.

We also filtered out gene families not present in at least half of the samples in either the AMY1L or the AMY1H group in the dataset. Then we identified the differentially abundant gene families at each TP using DESeq2 and used the $\log_2$ fold change between AMY1H and AMY1L for each gene family to create a heatmap. We adjusted p values using the Benjamini-Hochberg procedure to account for all gene families tested at a given TP and displayed gene families with BH-adjusted p values < 0.01. Furthermore, the heatmap includes only gene families with assigned taxonomy and is sorted by taxonomy. When a gene family is not significant at a TP, the corresponding heatmap cell is colored gray. The heatmap in Figure 6 was created using the software iTOL with gene families ordered by taxonomy.

Carbohydrate-Active enZYmes Analysis

We used hmmscan to query the gene families with HMMs from dbCAN (release 6.0), and used an e-value cutoff of 1e-18 to positively identify CAZymes.

We used linear mixed models to determine whether the number of read counts from any of the CAZyme classes differed between the AMY1H and AMY1L groups (N=20).

$$y \sim \text{AMY1CNG} + \text{CAZYCLASS} + \text{TP} + \text{AMY1CNG:TP} + \text{AMY1CNG:CAZYCLASS} + \text{CAZYCLASS:TP} + \text{AMY1CNG:CAZYCLASS:TP} + (1|\text{SUBJECT}) + \varepsilon$$

y represents the number of read counts, and fixed effects are *AMY1* CN group (**AMY1CNG**), TP (**TP**), and CAZyme class (**CAZYCLASS**). We also included a random effects term for repeat sampling of subjects (1|**SUBJECT**).

Determination of SAA Groups

Each of the 25 participants were labeled as a member of either SAA-H or SAA-L group. These labels were assigned using the KMeans clustering module in the Scikit-Learn library in Python, which implements the k-means algorithm, and was parameterized to identify two clusters within the mean salivary amylase activity measurements.

OTUs Distinguishing SAA Groups

For both saliva and stool samples, using phyloseq v1.22.3, we filtered out OTUs with an average relative abundance below 0.001%. This screening produced two data sets composed of 216 salivary and 283 fecal samples with 672 and 900 taxa, respectively. Samples were labeled according to the salivary amylase activity level assigned to its subject of origin. For the machine learning analysis, the OTUs were considered features, and the salivary amylase activity level (high or low) as the response variable. We used caret v6.0 to create a partition with 80% of the samples for training, and the remaining 20% for testing purposes. Then, we executed a random forest model on the training data set using the randomForest v4.6 package, adjusting the number of trees to 200 and enabling the flag to calculate the feature importances. Feature selection was done using Boruta v5.2.0 with default parameters on the training data set (after excluding, or not, OTU abundance measures coming from subjects originally placed in the AMY1M group). All the OTUs confirmed as important or tentative were treated as relevant, resulting in 113 for saliva and 301 for feces. To examine the predictive power of the model, we used the caret package and trained two random forest models with default parameters, one removing the relevant OTUs, and the other with only same relevant OTUs.

Predicting SAA with SCFA Levels in Stool

The dataset was divided in two subsets, one for the training process with 266 records, and another with 54 records for testing purposes. A RandomForestClassifier object was configured to use 250 estimators to predict the SAA group of each sample. Once the classifier was trained, its predictive performance was assessed using the following statistics: the F1 score, which represents the harmonic average of the precision and recall, and the Matthews correlation coefficient. To visualize the classification results, we generated a confusion matrix and a receiver operating characteristic (ROC) curve. Finally, the importance of features was calculated from the Gini impurity criteria used by the classifier to evaluate the quality of the resulting decision trees.

For the linear mixed models, SCFA concentrations were used as response variables and the SAA group as the predictor. The subjects and TPs were included as random effects (due to the fact that they were not taken into account in the random forest analysis). An ANOVA was then performed on each model to evaluate the differences in concentrations between SAA groups. The models were the following:

$$y \sim \textbf{SAA-group} + (1|\ \textbf{TP/SUBJECT}) + \varepsilon$$

with y being the concentration of acetate, propionate, butyrate, valerate, iso-valerate, heptanoate, and hexanoate alone, and then their sum:

$$\text{Total SCFA concentrations} = [\text{But}]+[\text{Val}]+[\text{Pro}]+[\text{Ace}]+[\text{Iso}]+[\text{Hex}]+[\text{Hep}]$$

The p-values from all of the models were corrected using the BH correction method.

Assessment of Adiposity in Mice

We used a linear mixed model to determine whether adiposity differed between AMY1H and AMY1L microbiome recipients using the following equation:

$$y \sim \text{AMY1CNG} + \text{TP} + \text{AMY1CNG:TP} + w_0 + \text{duration} + (1|\text{DONOR}) + \varepsilon$$

y is the percent fat determined by DEXA, and effect terms include *AMY1* CN group, time point, weight on the day of inoculation, and duration or length of experiment (**AMY1CNG**, **TP**, $\mathbf{w_0}$, **duration**), and we included a random term for human donor. We identified the affected TPs by performing post-hoc pairwise comparisons between the TPs using Tukey's HSD method to adjust for multiple comparisons.

DATA AND SOFTWARE AVAILABILITY

The 16S rRNA gene and metagenomic sequence data obtained from the intervention study have been deposited in the European Nucleotide Archive with accession IDs PRJEB27308 and PRJEB27304.