Diet leaves a genetic signature in a keystone member of the gut microbiota

Tanja Dapa[1], Miguel F. Pedro[1,2,4], Ricardo S. Ramiro[1,3,4], Isabel Gordo[1], Karina Bivar Xavier[1,5]*.

[1]Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal

[2]Present address: LEAF – Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, 1349-017, Lisbon, Portugal. (mfpedro@isa.ulisboa.pt)

[3]Present address: InnovPlantProtect, Estrada de Gil Vaz, Apartado 72, 7350-999, Elvas, Portugal (ricardo.ramiro@iplantprotect.pt)

[4]These authors contributed equally and are listed alphabetically.

[5]Lead Contact

*Correspondence: kxavier@igc.gulbenkian.pt

SUMMARY

Dietary switch from a low-fat and high-fiber diet to a Western-style high-fat and high-sugar diet is a common cause of microbiota imbalances underlying a variety of pathological conditions (i.e. dysbiosis). Although the effects of such dietary changes on microbiota composition and functions are well documented, their putative impact in gut bacterial evolution remains unexplored. Here we followed the emergence of mutations in *Bacteroides thetaiotaomicron*, a prevalent fiber-degrading microbiota member, upon colonization of the murine gut under different dietary regimens. *B. thetaiotaomicron* evolved rapidly to the gut and Western-style diet selected for mutations that promote the degradation of mucin-derived glycans. Periodic changes in diet led to fluctuations in the frequency of such mutations and were associated with metabolic shifts, resulting in the maintenance of higher intra-species genetic diversity compared to constant dietary regimens. Finally, our results suggest that *B. thetaiotaomicron* genetic diversity can be a biomarker for dietary differences among individuals.

KEYWORDS

Microbiota; High-Fat High-Sugar Diet; Western-style diet; Microbiota Evolution; Gut Dysbiosis; *Bacteroides thetaiotaomicron;* Bacteroidetes; Gut Metabolon; Gut Ecology; Multi-omics analyzes.

INTRODUCTION

The mammalian gut is inhabited by a diverse community of microbes, the gut microbiota, which establishes a symbiotic relationship with the host. This community is composed of hundreds of different species with vast genomic and metabolic repertoires that complement many host functions. Therefore, the gut microbiota can contribute to host health and physiology by influencing host nutrition, maturation of the immune system, and direct protection against infections (Baumler and Sperandio, 2016; Dominguez-Bello et al., 2019; Flint et al., 2012; Gilbert et al., 2018; Hooper et al., 2012; Rakoff-Nahoum et al., 2004; Ubeda et al., 2017). One of the most important functions provided by gut microbiota relates to its metabolic capabilities, which supply nutrients and energy via the degradation of resources that cannot be digested by the host, such as dietary plant fibers (Sonnenburg and Bäckhed, 2016). Degradation of these long-chain polysaccharides is performed by thousands of lytic enzymes, produced by members of gut microbiota, that depolymerize and ferment dietary polysaccharides into host-absorbable short-chain fatty acids (SCFAs) (El Kaoutari et al., 2013). Given the chemical diversity of dietary plant polysaccharides, the high genetic diversity provided by a high microbial diversity in the gut microbiota is linked to robust metabolism and improved host health, while a low gut microbial diversity is generally associated with disease (Le Chatelier et al., 2013; Cotillard et al., 2013; Turnbaugh et al., 2009a). Many factors, such as drugs, lifestyle, and inflammatory responses to infections can perturb microbiota composition, and often result in decreased diversity (Relman, 2020).

The host diet changes the microbiota, and the impacts go beyond the known direct effect on body weight and metabolic imbalances (Kreuzer and Hardt, 2020; Sonnenburg and Sonnenburg, 2014). Many members of the microbiota rely on dietary fibers for their metabolic functions. Therefore, shifts from a diet low in fat and rich in plant fibers (hereby Standard Diet, SD) to a diet high in fat and simple sugars and low in polysaccharides from plant fibers (hereby Western-style Diet, WD), cause dramatic changes in the composition of gut microbiota in humans and rodents (David et al., 2014; Desai et al., 2016; Faith et al., 2011; McNulty et al., 2013; Rey et al., 2013; Turnbaugh et al., 2009b; Zhang et al., 2012). A dietary

change that extends over several host generations can cause the irreversible loss of several important taxa from the microbiota, the most affected being the Bacteroidetes (Sonnenburg et al., 2016).

The members of the Bacteroidetes phylum, which includes the genus *Bacteroides*, are the most abundant Gram-negative gut symbionts in urban human populations, and in many humans can constitute 50-80% of the microbiota (The Human Microbiome Project Consortium*, 2012). The *Bacteroides* are among the microbiota members with the larger repertoire of lytic enzymes capable of degrading chemically diverse long-chain polysaccharides and can be considered gut specialists in fiber degradation (Porter and Martens, 2017). While most *Bacteroides* degrade complex polysaccharides from dietary plant fibers, some can also metabolize host glycans (Salyers et al., 1977a, 1977b). As a result, diets depleted in fiber lead not only to a decrease in members of *Bacteroides* that rely mainly on the digestion of plant fibers, but also in shifts in gene expression and enzyme production in *Bacteroides* capable of accessing both plant and host-polysaccharides (Sonnenburg et al., 2016, 2005). The increase in consumption of host glycans by microbiota can have negative consequences for the host as it can decrease the thickness of the mucus layer and accelerate disease progression induced by the pathogen *Citrobacter rodentium* (Desai et al., 2016; Kreuzer and Hardt, 2020). Additionally, a diet lacking microbiota-accessible carbohydrates favors *Clostridioides difficile* infections (Hryckowian et al., 2018), and a high-fat diet can favor *Salmonella* Typhimurium gut colonization (Wotzka et al., 2019).

The negative impacts of a WD on the microbiota are well documented, WD changes: microbiota composition, patterns of gene expression in members of these community, and gut metabolome (Albenberg and Wu, 2014; Desai et al., 2016; Turnbaugh et al., 2009a). However, the potential evolutionary implications for species present in the microbiota remain unexplored (Crook et al., 2019; Ghalayini et al., 2019). The evolution of bacteria in laboratory cultures is strongly influenced by the nutrients available in the medium. For example, the type and abundance of the carbon source affects both which mutations will be adaptive and the mutational spectrum (Maharjan and Ferenci, 2017; Turner et al., 2018). Host diet affects

the nutrients available for bacteria in the gut, but knowledge is very limited about how that affects gut bacterial evolution. Yilmaz, et. al., recently showed that, in mice colonized with a defined microbiota community, different variants and sub-strains can coexist within individual taxa and that changes in diet affect not only the proportion of the different taxa, but also the proportion of the variants and sub-strains (Yilmaz et al., 2021). Experimental evolution approaches with *Escherichia coli* have revealed that its evolutionary adaptation to the mouse gut selects for mutations related to changes in metabolic capabilities (Barroso-Batista et al., 2014; Conway and Cohen, 2015; Giraud et al., 2008; Lescat et al., 2017; De Paepe et al., 2011), and that those mutational patterns are shaped by gut ecology (Barroso-Batista, Pedro et al., 2020). In humans, different dietary habits have been associated with genetic diversity in *Prevotella copri*, leading to genetic variants with functional consequences, namely strains from non-Western subjects showing higher potential for complex plant fiber-degradation (Fehlner-Peach et al., 2019; De Filippis et al., 2019). Also in humans, there is evidence that gut microbes can accumulate genetic differences over host-relevant timescales (Garud et al., 2019; Zhao et al., 2019). However, given the complexity of monitoring human gut microbiota populations, it is difficult to accurately measure intra-species evolutionary dynamics and determine the contribution of emergence of *de novo* mutations versus invasion of new strains due to inter-host transmission. Furthermore, possible strong colonization bottlenecks, and non-controlled factors, make the understanding of the mechanisms shaping intra-species genetic diversity within human hosts difficult.

We reasoned that the microbiota-dependent functional consequences observed in diet-induced dysbiosis could be caused not only by changes in microbiota composition and in gene/metabolic regulation, but also by evolutionary changes, via the emergence of *de novo* mutations, which can lead to intra-species changes. To address this hypothesis, we determined the effects of diet on mutational changes in genes at the single species level by monitoring evolutionary adaptation to the gut of mice undergoing different dietary regimens. We focused on *Bacteroides thetaiotaomicron* (hereafter referred as *B. theta*) due to its predominance in the mammalian gut. *B. theta* is a strict anaerobe, and is among the fiber

degrading *Bacteroides* that in the absence of dietary plant polysaccharides can consume host glycans (Salyers et al., 1977a; Sonnenburg et al., 2005). This bacterium shows phenotypic plasticity by gene and metabolic regulatory mechanisms, which enable it to prioritize usage of carbon sources (Schwalm et al., 2017), and to consume host glycans only in the absence of dietary complex polysaccharides (Kashyap et al., 2013; Martens et al., 2008). By following the evolutionary dynamics of *B. theta* over a 3-month timescale, we observe genetic signatures resulting from diet-specific evolution, fluctuating rapidly as the diet consumed by the mice alternates from high in fat and sugar and low in fiber (Western-style Diet (WD)) to standard high-fiber chow diet (Standard Diet (SD)). We show that adaptation under WD specifically favors the emergence of mutations advantageous in consumption of mucin *O*-glycans. This supports the hypothesis that intra-species evolution can influence the microbiota-dependent phenotypes observed upon dietary changes. Finally, through an integrative multi-omic analysis, combining metabolomic and microbiota profiling with the *B. theta* mutational profile, we show that intra-species mutational diversity is a powerful biomarker of dietary differences between individuals.

RESULTS

B. theta *evolves rapidly to the mammalian gut under different diets*

We studied evolutionary adaptation dynamics of a prevalent member of the gut microbiota, *B. theta* (VPI-5482) (Consortium et al., 2010; Qin et al., 2010), to the mouse gut in animals fed with different diets. To do this, antibiotic-treated mice, susceptible to colonization, were gavaged with a 1:1 mixture of *B. theta* carrying either mCherry or sfGFP constitutively expressed (Whitaker et al., 2017), enabling us to monitor frequency changes of these markers as a proxy for evolutionary adaptation (Barroso-Batista et al., 2014; Hegreness et al., 2006). Mice were stably colonized with *B. theta* 48 hours after the gavage (Table S1). At that point the mice were either kept on the same diet (*i*) SD: rich in microbiota-accessible carbohydrates (*i.e.* plant polysaccharides and fibers) and low in fat and simple sugars; or the

diet was changed to (*ii*) WD: rich in fat and simple sugars, but poor in microbiota-accessible carbohydrates (Fig. 1A). As expected, mice on WD gained more weight, had larger fat pads, and tended to have increased inflammation (Fig. S2A-C). Changes in microbiota assessed through 16S rRNA amplicon sequence variants (ASVs) revealed a tendency for lower relative levels of phylum Bacteroidetes in the WD group, and higher levels of phylum Verrucomicrobia, which includes the mucus degrader specialist *Akkermansia* (Fig. S2D), consistent with previous observations (Carmody et al., 2015; Desai et al., 2016)

We observed rapid shifts in the frequency of the fluorescent markers of *B. theta* populations in fecal samples from mice fed with either SD (e.g., lines SD2/6/9/14, Fig. 1B, see also Fig. S1, left panel) or WD (e.g., WD2/6/8/11, Fig. 1C, see also Fig. S1, center panel). The rapid shifts suggest that the accumulation of adaptive mutations can occur as quickly as 2 weeks. In both the SD and WD groups, the changes in fluorescent markers of *B. theta* populations in certain mice are consistent with clonal interference (CI): quick reversals in marker frequency indicates that distinct beneficial mutations appeared in either background with emerging adapted clones competing for fixation (e.g., SD6 and WD8, Fig 1B-C). Absolute abundance of *B. theta* (CFUs/g) did not change significantly throughout the experiment (Table S1), indicating that the strong shifts in marker frequency were due to the acquisition of beneficial mutations rather than to population bottlenecks caused by potential deleterious effects of the gut environment.

To confirm that the changes in marker frequency were indeed caused by evolutionary adaptation, we asked if populations that evolved in either diet (SD_{evol} or WD_{evol}) have higher fitness than the ancestral strain (ANC) when competing for colonization. We observed that evolved populations largely outcompete ancestral clones in newly colonized mice, irrespective of the dietary regime used during competitions (Fig. 1D-E, Fig. S2E-H). These results confirmed that *B. theta* had acquired strong-effect mutations that are adaptive in the mouse gut independently of the diet. To determine if *B. theta* had acquired diet-specific adaptive mutations, we next competed SD_{evol} against WD_{evol} populations on each diet. Most SD_{evol} populations outcompeted WD_{evol} populations when mice were fed SD (Fig. 1F,

frequency of SD_{evol} in Fig. S2I-L, left panels), but were outcompeted by WD_{evol} when mice were fed WD (Fig. 1F, frequency of WD_{evol} in Fig. S2I-L, right panels), confirming that *B. theta* does accumulate diet-specific beneficial mutations during colonization of the mouse gut.

Alternating diet leads to fluctuating selection

We next sought to analyze a more realistic and complex regimen, where diet changes across time. This was accomplished by alternating each week between standard and western diets, *i.e.*, Alternation Diet (AD) group in Fig. 1A. With this regimen we observed strong fluctuation of the neutral markers in response to the weekly changes in diet (Fig. 1G, see also Fig. S1, right panel). For example, lineages AD1 and AD4 show clear fluctuations, where markers alternate in dominance reaching frequencies ~100% in one week and ~0% in the next. We quantified the shifts in marker frequency across all 3 dietary regimes by calculating ΔMarker frequency ($t_{x=0}$-$t_{x=-1}$) (Fig. 1H) and observed that AD causes stronger marker shifts than the SD and WD groups (Fig.1H - insert, AD vs. SD $^{**}p = 0.0015$; AD vs. WD $^{***}p = 0.0006$; SD vs. WD non-statistically significant, $p = 0.9240$; one-way ANOVA with Tukey's multiple comparisons test).

All together these results provide evidence of fast evolutionary adaptation of *B. theta* to the mouse gut in all three nutritional regimens, with a higher number of fluctuations in marker frequency in the AD regimen, indicating the emergence of diet-specific beneficial mutations.

Mutations underlying B. theta *evolutionary adaptation in different diets*

To identify mutations potentially selected in the different diets, we sequenced populations of *B. theta* by pooling clones sampled from each mouse after 12 weeks of colonization. We determined *de novo* mutations by mapping the sequencing reads against our assembled ancestral *B. theta* reference (see methods).

The number of mutations accumulated in most *B. theta* populations was between 20 and 70 mutations per mouse (Fig. 2A, Z-axis – bars and Table S2). Interestingly, three of the evolved populations (mice SD1, SD2 and WD3, Fig 2A) had much higher numbers of mutations, harboring 274, 230 and 156 mutations, respectively. This is likely due to the emergence of clones carrying mutations that increase the mutation rate (mutators). Indeed, in population SD1, a mutation in the promoter region of the *mutS* gene, a known target for mutator phenotypes (Radman et al., 1995), was detected at 0.7% frequency. Additionally, the majority of the mutations in mice SD1, SD2 and WD3 are at low frequency, as is often observed for the mutations arising in mutator backgrounds (Giraud et al., 2001; Ramiro et al., 2020). Specifically, 250 mutations in SD1, 205 mutations in SD2, and 86 mutations in WD3 are below 2% (Fig. 2A, Y-axis – dots).

Among all mutations analyzed from all sequenced populations, single-nucleotide polymorphisms (SNPs) are the most prevalent, followed by small insertions and deletions (indels) (Fig. S3A).

As adaptive mutations appearing in different mice emerged independently, a proxy for *bona fide* adaptive mutations is parallel acquisition. We identified 73 parallel mutations (Fig. 2B and Table S2), defined as any gene or intragenic region (mutational target) carrying mutations in populations from at least two mice, and with the sum of mutation frequencies in each target of >5% in at least one mouse. As most mutations identified fit our criteria of parallel mutations, the majority of mutations are likely adaptive (Fig. 2B and S3B). No bias regarding genome position was observed for parallel mutations, as these are scattered across the bacterial genome of the populations from all groups of mice (Fig. 2C). These results suggest that *B. theta* has a large target for beneficial mutations when adapting to the mouse gut ecosystem.

Alternation diet leads to increased intra-species genetic diversity in B. theta

To understand the effect that diet has on intra-species diversity in *B. theta,* we compared the frequencies of parallel mutations of the different groups and observed that the AD shows a tendency to have lower maximum frequencies than SD (Fig. 3A) and is lower than WD (AD vs. WD *p = 0.0226; one-way ANOVA with Tukey's multiple comparisons test). Interestingly, the Shannon index (here used to measure intra-species richness) for parallel mutations is significantly higher in the AD group than in WD and SD groups (Fig. 3B, AD vs. *p = 0.0240; AD vs. WD **p = 0.0044; SD vs. WD non-statistically significant, p = 0.7636; one-way ANOVA with Tukey's multiple comparisons test), indicating that the AD regimen supports higher polymorphism.

Next, to identify potential diet-specific mutational targets we used a generalized linear model (GLM) to identify mutations differentially prevalent across dietary regimens (p<0.05 and false discovery rate (FDR) of q<0.05), and/or differentially enriched in a specific dietary regimen (pairwise post-hoc tests between regimens per mutational target) (Fig. 3C). Mutations that did not pass our criteria for diet-specific selection (GLM p>0.05), are listed in Table S2. Mutations, which are present in a majority of the mice, irrespectively of the diet, are plausibly generally adaptive mutations to the mouse gut. The occurrence of general evolutionary adaptation to the gut is consistent with the higher fitness observed in populations evolved under SD or WD in comparison with the ancestral strain, regardless of diet (Fig. 1D-E).

Regarding the diet-specific mutations (Fig. 3C), BT1754 was identified as a single mutational target exclusively in mice from the SD group. This gene encodes a transcriptional regulator previously shown to be induced by the presence of plant dietary polysaccharides (Lynch and Sonnenburg, 2012). In contrast, two of the four mutational targets enriched in the WD group (BT4246 and BT4247) are SusCD-like genes, which are part of polysaccharide utilization locus (PUL) 78 involved in degradation of mucin *O*-glycans from the host (Kashyap et al., 2013). Additionally, mutations in the intergenic region between genes involved in capsular polysaccharide biosynthesis (BT1725/1726) were enriched in the WD group;

expression of this locus has been previously observed to be among the dominant capsular loci in the low-fiber diet (Porter et al., 2017). These mutations enriched in either diet likely result from *B. theta* adaptation to the general gut environment.

We also identified 3 mutational targets present in both the SD and AD groups (BT0623, BT0370 and BT0317), and one present in both the WD and AD groups (BT3702). Mutations in these targets are likely to be more beneficial in either SD or WD diets, respectively, but are still being maintained in the AD regimen, where the mice are sequentially exposed to both diets.

Interestingly, we found more mutations specific to the AD regimen (6) than mutations specific to SD and WD (1 and 4, respectively) (Fig. 3C). These AD-specific targets likely contain evolutionary adaptations to a fluctuating regimen, and again include genes involved in the utilization of polysaccharides, in this case both, dietary and host glycans. Overall, these results show that the fluctuating diet favors maintenance of a higher intra-species polymorphism among the newly emerged genetic variations.

B. theta *evolution is more strongly influenced by the metabolome than by the microbiota composition*

We next wanted to determine how different dietary regimens affect the interplay between *B. theta* evolution, gut microbiota composition, and the metabolic environment in the gut. We analyzed the microbiota compositions and the metabolite composition in fecal samples from the last day of the experiment. We observed a tendency towards lower relative levels of members of the phylum Bacteroidetes and higher levels of phylum Verrucomicrobia in the microbiota of mice from the WD group than with the SD and AD groups (Fig. 4A, Table S4). In contrast, significant differences in metabolites were observed between the groups. In particular, the levels of the major fermentation products of the microbiota SCFAs (namely, acetate and propionate) and the majority of simple sugars present in the WD diet (e.g., sucrose and its derivative, fructose), were drastically different in the WD group compared

with SD and AD groups, with SCFAs being lower and these sugars higher in WD (Fig. 4B). Taurine, an important component of bile acids is also higher in the WD (Fig. S4A-C and Table S3, shown together with remaining metabolites). Bile salts are necessary for fat digestion; the increase in bile acids due to the high fat diet was associated with promoting gut colonization by pathogenic bacteria like *Salmonella* Typhimurium (Wotzka et al., 2019).

To identify significant correlations among mutational profiles, microbiota compositions or metabolomes, we performed Procrustes analysis based on principal component analysis (PCA) of Aitchinson distances for each dataset - mutations, metabolites and microbiota composition (Fig. 4C-E, PCA for each dataset are presented in Fig. S5A-C). This analysis detected a significant correlation between the mutational profile and the metabolome, but not between the mutational profile and the microbiota or between the metabolome and the microbiota (Fig. 4F). Similar results were obtained with a Mantel test (Fig. S4D). Therefore, both analyses provide support for a strong correlation between the metabolome and the mutations, greater than the correlation between the metabolome and the microbiota compositions.

We next used the tool HallA (Hierarchical All-against-All association testing, (Rahnavard et al., 2017)) to identify correlations between features of the mutational profile, microbiota and metabolomes. The results from this analysis were then used to create the correlation network (Fig. 4G). The network recapitulates some expected associations, such as (*i*) the positive correlation between frequency of mutations in the SusC/D-like genes enriched in the WD regimen, BT4246-7, with the concentration of simple sugars present in WD (sucrose and fructose), (*ii*) the negative association between these mutational targets and metabolites enriched in SD diet (arabinose or xylose); (*iii*) the positive associations between glutamate and mutations in BT0317 and BT0370, which are enriched in SD and AD regimens, (the BT0370 gene has homology to galactokinase, an indication that it could be involved in galactan metabolism, a polysaccharide present in SD); and (*iv*) the positive correlations between taurine and mutations in the intergenic region BT1725/BT1726 involved in capsular polysaccharide biosynthesis (Porter et al., 2017), which are enriched in WD.

Interestingly, we also found associations that could not be predicted from the independent analysis of each dataset, e.g., (*i*) negative correlations between frequency of mutations in BT3045 and tyrosine, or BT0172/BT0173 and creatine; and (*ii*) the positive correlations between BT3045 (predicted gene for arabinan degradation (Martens et al., 2011)) and *Blautia*. Importantly, consistent with the Procurstes analysis, this correlation network also supports a stronger correlation of mutations with metabolites than of mutations with microbiota, as it has a higher number of network edges between the mutational profile and metabolome (36 edges) than between mutations and the microbiota (4) or between microbiota and the metabolites (17) (Fig. 4G). Overall, these results indicate that bacterial evolution is more strongly influenced by the metabolite environment rather than by the gut microbiota composition.

Mutational targets are good predictors of host diet

The gut microbiota and the metabolome are often mined for their potential as biomarkers (Krautkramer et al., 2021; Manor et al., 2020). However, given the limited data available on genetic diversity within individual microbiota species, the predictive power of within-species mutational profiles has been less explored. To determine the power of genetic diversity of *B. theta* to predict the different dietary regimens in comparison with the predictive power of microbiota and/or metabolome, we used DIABLO (Data Integration Analysis for Biomarker discovery using Latent variable approaches for 'Omics studies (Singh et al., 2019)), which is based on sparse generalized canonical correlation analyzes. We carried out an integrative analyzes of the mutational profile, the metabolome and the microbiota and tested the ability of each dataset to distinguish between the different dietary regimens. DIABLO identified the three main latent variables, the first two are shown in Fig. 5A-C as an ordination plot (for variable 3 see also Fig. S5E). Interestingly, the mutation dataset separated the different dietary regimens into three clusters (Fig. 5A), whereas metabolites strongly separate WD from AD and SD, but not the latter two (Fig. 5B), likely a consequence of the fact that the AD

samples were from a time point when the AD group was on SD diet. Moreover, while microbiota composition clusters the AD samples tightly, the clustering of SD and WD is poor, thus causing a poor overall separation among the three dietary regimens when using the microbiota dataset (Fig. 5C).

We then evaluated the ability of the DIABLO model to correctly predict the dietary regimen of each sample based on the separate datasets. We obtained cross-validation error rates for the 3 datasets, for which lower error rates indicate a better prediction (Fig. 5D). Interestingly, the mutation dataset is the only one that always achieved a misclassification of ≤20% for all three dietary regimes, being the best at identifying SD samples and as good as metabolites at identifying WD samples. Conversely, while the microbiota dataset could better identify AD samples, it was the poorest for the other two regimens, with error rates >40% (Fig. 5D). The fact that the mutation dataset was a good diet predictor is further support for the existence of a specific genetic signature for AD, also suggested by the larger number of AD-specific mutations identified (Fig. 3C). Taken together, our results support the conclusion that diet can leave specific genetic signatures in *B. theta* and provide evidence that intra-species mutational diversity can be a powerful biomarker of dietary differences between individuals, being at least on par with the microbiota composition and the metabolome.

Weekly nutritional changes drive genetic fluctuation in B. theta

Our identification of diet-specific mutations at 12 weeks shows that diet shapes the evolution of *B. theta* in the mouse gut (Fig. 3). Since we observed strong fluctuation of the fluorescent marker under AD dietary regimen (Fig. 1G-H) we hypothesized that the weekly changes in diet could cause fluctuating selection on mutations, *i.e.*, differential benefits in the two diets alternated in the AD group. To test this hypothesis, we performed time-course population WGS for the two populations from the AD group that showed the most intense fluctuation in marker dynamics, i.e., AD1 and AD4. We sequenced populations from Day 0 of evolutionary experiment, and each week on Day 6 post diet shift, from week 2 until the end of the

experiment. Fig. 6A shows the frequency of all mutations across time. We found mutations that fluctuated up to 4-5 times, these being the ones that displayed the stronger absolute changes in frequency (Fig. S6A, left panel).

Among the mutations fluctuating more strongly we highlight two mutational targets detected in both mice: BT0867 (encoding a SusC-like protein) and BT2689 (encoding a hypothetical protein) (Fig. 6A, colored lines, Table S5). The mutation in BT0867 was first detected on week 3, under a western diet week (AD_{WD}) and fluctuated in frequency between >80% in AD_{WD} weeks and <10% on AD_{SD} weeks, reaching frequencies of 100% in the former and becoming undetectable in the latter (Fig. 6A). The mutations in the gene BT2689 showed very similar dynamics, but with the frequency of the mutations peaking on AD_{SD} (Fig. 6A). This analysis confirms that changes in diet can lead to strong fluctuations in the frequency of mutations.

To understand if weekly dietary changes lead to fluctuations in the gut metabolome and microbiota in a similar way, we analyzed metabolites and the microbiota composition of AD1 and AD4 samples from week 4 until the end of the experiment. Similar to mutations, it was possible to detect fluctuations for metabolites and microbiota composition (Fig. 6B-D and S6).

The metabolites showing the strongest fluctuation patterns were sugars (Fig. 6B). Their dynamics clearly reflects the diet consumed in the week analyzed, with sugars derived from polysaccharides, which are part of SD composition (e.g., arabinose/arabinan, xylose/xylan and galactose/galactan), peaking on AD_{SD} week, and simple sugars present in WD (e.g., sucrose and its derivative fructose), peaking on AD_{WD} week (Fig. 6C). Levels of glucose did not fluctuate significantly through the entire course of the experiments, perhaps because its concentration is regulated by the host. Among the SCFAs, acetate showed the strongest oscillation pattern and, together with the organic acid galactonate (abundant building block in fiber polymers), peaked in the AD_{SD} week (Fig. 6C, for remaining metabolites Fig. S6B-C and Table S3). Regarding the microbiota in AD4, Verrucomicrobia

phylum (which include the ASVs corresponding to *Akkermansia*) fluctuated with peaks in frequency corresponding to the AD_{WD} weeks (Fig. 6D, for individual ASVs see Table S4).

Next, to compare the rate of fluctuation across the different datasets (mutations, metabolites and microbiota composition) each of the datasets was first filtered to reduce sparsity (see methods). We then applied the centred log ratio (CLR) transformation to each of the datasets to account for compositionality (Fernandes et al., 2014; Quinn et al., 2019). Using this data we calculated the Aitchinson distance between consecutive time points for each mouse. The distance between consecutive time points was the highest for mutations, followed by metabolites and 16S variants (Fig. 6E and Fig. S7). Furthermore, PCAs generated based on the Aitchinson distance for each dataset, showed that for mutations and metabolites the time points clearly cluster according to the diet consumed in the indicated week (Fig. 6F). Conversely, for microbiota dataset there is no clear separation between AD_{SD} and AD_{WD} weeks (Fig. 6F, right panel). Therefore, this analysis shows that mutations and metabolites are subject to stronger fluctuation than microbiota composition.

Overall, our results demonstrate that changes in diet lead to rapid changes in the metabolic environment. Interestingly, these fluctuations in diet lead to even clearer evolutionary dynamics observed at the gene level, supporting that the mutational landscape is a reflection of the host diet. Additionally, these results suggest that mutations follow diet fluctuations more closely than the microbiota composition dynamics.

Western-style diet selects for fluctuating mutations involved in consumption of mucin O*-glycans*

Our results showing fluctuations in the frequency of mutations indicate that the fitness effect of mutations can change with weekly dietary changes. We analyzed the frequency of the highly fluctuating mutational targets BT0867 and BT2689 on the last day of the experiment (week 12). Intriguingly, mutations in BT0867 were highly prevalent in all the groups under 3 regimens (SD, WD and AD), reaching up to 60% in frequency (Fig. 7A). This was in contrast

with the mutations in BT2689, which peaked in the AD_{SD} weeks and at 12 weeks was detected only in the SD and AD groups (Fig.7A, Table S2). Detailed analysis of the different mutations in BT0867 at the last time point of the experiment revealed 5 different alleles. Interestingly, out of the five, the only allele enriched in the AD group is the one generating the amino acid substitution T756I (Fig. 7B), which is the SNP detected in BT0867 that fluctuates over time in both AD1 and AD4 populations (Fig. 6A). Importantly, the same allele is also enriched in the WD group. However, the R988C substitution, is also prevalent in WD group (Fig. 7B).

Because the allele T756I peaks in the AD_{WD} weeks, and is highly prevalent in the WD group, we hypothesized that this mutation could be beneficial under the WD conditions. As the gene BT0867 is part of PUL12, which has been associated with the utilization of mucin *O*-glycans (Bjursell et al., 2006), we reasoned that this mutation could be benefiting *B. theta* by optimizing its capacity to consume these compounds. To test this hypothesis, we took advantage of the strong fluctuation of the T756I allele throughout the experiment, and isolated clones from the AD1 and AD4 populations at time points of extreme frequencies of T756I; we isolated 8 clones at week 4, when frequency of this allele is ~0% (AD_{SD}), and 8 clones at week 5, when its frequency is ~100% (AD_{WD}). Then the phenotype of these clones with respect to *O*-glycan consumption was tested in laboratory cultures. We used a modified carbohydrate-free medium, which enables testing growth enhancement by carbohydrate supplementation. We tested the isolated clones in competition against the ancestral clone (ANC), either in medium with no added carbohydrate or in medium supplemented with either glucose or purified mucin *O*-glycans. We observed that neither the AD_{SD} nor AD_{WD} clones showed advantage against ANC in medium supplemented with glucose or in medium with no carbohydrates added. However, the AD_{WD} evolved clones outcompeted the ANC in medium supplemented with purified mucin *O*-glycans (Fig. 7C). Moreover, AD_{WD} clones showed a higher growth rate than the AD_{SD} clones when grown in the presence of purified mucin *O*-glycans, but not when supplemented with glucose or when no carbohydrate was added to the medium (Fig. 7D). The growth advantage of the evolved clones in the presence of *O*-

glycans indicates that these clones might be poised for better for consumption of the mucus layer, a phenotype previously observed for *B. theta* in animals under WD (Sonnenburg et al., 2005).

The results showing that the T756I allele of BT0867 strongly fluctuates in frequencies upon changes from the AD_{WD} week to the AD_{SD} weeks demonstrate that *B. theta* mutations selected during the AD regimens can have different fitness effects in the two diets. Moreover, our results indicate that the T756I allele is strongly beneficial in WD, where utilization of mucus *O*-glycans is expected to be advantageous, a phenotype that is consistent with its advantage in *O*-glycan utilization (Fig. 7C and D). Therefore, our results support the notion of selection for *de novo* mutations (i.e., emergence of genetic diversity) as an additional mechanism to explain microbiota-dependent phenotypic alterations observed upon changes in diet.

DISCUSSION

Host diet is an important factor influencing the functions and composition of the gut microbiome (Sonnenburg and Sonnenburg, 2014). Diets low in complex polysaccharides, such as the WD used here, lead to microbiota-dependent negative effects on the host, including decreased thickness of the protective mucus layer and increased susceptibility to infection and inflammation (Bäckhed et al., 2005; Desai et al., 2016; Hryckowian et al., 2018; Wotzka et al., 2019). Many studies have shown that such consequences of diet-induced dysbiosis relate to impacts in gut ecology characterized by changing both the microbiota composition and the expression of genes by members of this community (Carmody et al., 2015; Desai et al., 2016; Martens et al., 2008, 2011; Sonnenburg et al., 2005; Turnbaugh et al., 2008). These mechanisms are not mutually exclusive, and both change the gut metabolic environment (Barroso-Batista, Pedro et al., 2020; Desai et al., 2016). The metabolic environment in turn changes the fitness landscape experienced by the gut commensals and can change their evolutionary path. Evidence that evolution by *de novo* mutations occurring

during bacterial colonization of the gut is just now starting to emerge (Fehlner-Peach et al., 2019; De Filippis et al., 2019; Garud et al., 2019; Yilmaz et al., 2021; Zhao et al., 2019).

Here we investigated how diet shapes the evolution of *B. theta* and the emergence of intra-species genetic diversity in this prevalent member of the microbiota. We found that this bacterium evolves quickly to the mouse gut and that different dietary regimens leave a distinct genetic signature. Our approach illustrates the power of experimental evolution with trackable microbiota members to understand the microbiota-mediated mechanisms involved in functional alterations observed in the mammalian gut in response to environmental perturbations. Our findings show that intra-species evolution can happen within host-relevant timescales in response to dietary changes.

We observed that *B. theta* accumulated a similar number of parallel mutations under SD and WD. Competitive fitness assays in newly colonized mice showed that evolved clones outcompete their ancestors, confirming that those mutations are beneficial mutations. Population sequencing of the evolved clones revealed the main mutational targets in both groups. This supports the notion that many mutations constitute diet-independent evolutionary adaptations, which is likely a consequence of *B. theta* being a human isolate, and therefore mouse gut presents a novel environment for it. Interestingly, when evolved populations were competed against each other, the SD_{evol} populations outcompeted the WD_{evol} in mice fed SD, but lost the competition when mice were fed WD. This indicates that *B. theta* did not acquire only diet-independent mutations, but also diet-specific adaptive mutations. Accordingly, we identified mutational targets that were selected only in SD or only in WD. As the majority of mutations identified are non-synonymous SNPs, which are less likely to lead to loss of functions than deletions or stop codons, we reason that such mutations are likely to result in changes of gene expression or protein activity. These diet-specific mutations include mutations in genes that were previously shown to be important under conditions that are relevant for either the SD or WD diet. The BT1754 gene mutated in SD is essential for growth on fructans (fructo-oligosaccharide, inulin and levan) present in plant fibers (Deutschbauer and Chen, 2021; Sonnenburg et al., 2010). Two of the mutational

targets enriched in WD, SusCD-like loci BT4246/7, belong to the polysaccharide utilization loci PUL78, previously described as one of the PULs involved in consumption of host mucin *O*-glycans (Kashyap et al., 2013; Martens et al., 2008). Gut mucus has high polysaccharide content, with mucin *O*-glycans presenting up to 80% of total biomass (Johansson et al., 2013). However, only a minority of gut microbiota species is able to use it as a nutrient source (Hoskins and Boulding, 1981; Png et al., 2010), with *B. theta* being one of them (Sonnenburg et al., 2005). These glycans are not *B. theta´*s primary nutrient choice, but rather a last resort (Kashyap et al., 2013; Pudlo et al., 2015). Therefore, the appearance of mutations in SusC-like (BT4247) and SusD-like (BT4246) genes suggests that an environment poor in polysaccharides, creates a selective pressure for *B. theta* to evolve towards increased host mucin consumption. Additionally, among the targets enriched in WD we identified mutations targeting a locus involved in capsular polysaccharide biosynthesis (BT1725/1726), previously shown to be one of dominant capsular loci under low-fiber diet (Porter et al., 2017). Altogether, the appearance of these mutations indicates that *B. theta* evolves to the metabolically distinct gut environment generated by SD and WD diets. Some of the diet-specific mutations selected are in genes of unknown function, but the fact that these genes are under selection in these experiments strongly supports their role in gut colonization. Understanding the factors that selected for mutations in these targets will help identify the functions encoded by them.

Our results are consistent with previous findings that metabolites and microbiota composition are both affected by diet. We found new associations between mutations in certain genes and metabolites and/or microbiota members which were enriched in the same diet where these mutations were selected. This is evident for correlations between mutations and diet-specific sugars, as well as for certain mutations and *Akkermansia*, a microbiota mucus degrader often seen to expand in diets low in fiber (Desai et al., 2016). We also observed correlations between metabolites and microbiota composition, which are consistent with previous studies showing that metabolic changes induced by diet shifts lead to changes in microbiota composition (Desai et al., 2016). Interestingly, the overall correlation between

mutations and metabolites is stronger than those between metabolites and microbiota composition, or between mutations and microbiota. These results show that diet and the gut metabolic environment shape the adaptive landscape of *B. theta* more than the microbiota composition.

Analyzing the effect of periodic alterations in diet (AD group) allowed us to determine how evolution of *B. theta* proceeds in a more realistic scenario, where the host experience periodic changes in diet. We observed that the AD regimen selects for a similar number of targets as did continuous exposure to either SD or WD. However, maximum frequencies reached by prevalent mutants are smaller in the AD group than in SD or WD, revealing less population dominance in AD. This together with a higher level of intraspecies genetic richness in the AD group shows that this group maintains a richer genetic repertoire, and thus demonstrates the importance of dietary alternation in maintaining intra-species diversity.

Importantly, the number of diet-specific targets of selection was higher in the AD regimen than in SD and WD, suggesting that alternations in diet require different mechanisms and introduces different specific selective pressures. This was particularly obvious when strong fluctuations in the frequency of mutations were observed. A particular mutation showing strong fluctuations in AD is a single SNP in the gene BT2689, coding for a hypothetical protein. The function of BT2689 seems to be strictly related to consumption of resources from SD, since mutation in this target never appeared in mice on WD. The other mutation showing strong fluctuation in AD is in BT0867, a gene previously shown to be induced by mucin *O*-glycans (Bjursell et al., 2006), where a high prevalence of mutations occurred across dietary groups. Importantly, the observed fluctuations in mutations can have a functional impact as we show that the fluctuating clones with the SNP T7561 in BT0867 peaking in AD_{WD} have a growth advantage in mucin *O*-glycans. This enrichment for mutants with enhanced growth in mucin *O*-glycans in AD_{WD} weeks, correlates well with other studies describing the effects of WD in the decrease in thickness of the mucus barrier (Desai et al., 2016; Sonnenburg et al., 2005). Therefore, we reason that selection for mutants with enhanced mucus degradation capabilities is one of the mechanisms involved in the observed

negative effect of WD dysbiosis. More generally, we propose that in responses to environmental perturbations, microbiota functions can change either by the previously described mechanism of gene regulation (Desai et al., 2016), or by rapid selection of mutations. Importantly, changes resulting in mutations will have more permanent consequences than the responses at the level of gene regulation, as the later are reversible, while mutations are not. Therefore, the results obtained here with the AD, showing increased richness for mutational targets and fluctuation of mutations with the weekly changes in diet, demonstrate that periodic variations in dietary regimens are important to avoid irreversible fixation of specific mutations and to enable maintenance of a higher genetic polymorphism. Therefore, one can in principle use diet supplementations to manipulate intra-species genetic diversity. Overall, these results suggest that the AD regimen selects for a specific genetic signature, resulting in functionally distinct evolutionary paths. Moreover, the results from the weekly alterations in diet revealed stronger fluctuations in mutations than in microbiota composition, thus highlighting the importance of investigating potential changes in functional properties of the microbiota, even under conditions where alterations on the species composition are not obvious.

The observed fluctuations in frequencies of mutations correlating with periodic alternations in diet in the AD group is consistent with the occurrence of antagonistic pleiotropy (Van den Bergh et al., 2018). A mutation under antagonistic pleiotropy will increase in frequency in a week where diet favors it, with an expansion of the linked fluorescent marker, but will decrease in frequency when the diet changes as the mutation becomes temporally disfavored (Chen and Zhang, 2020).

Our findings showing that *B. theta* undergoes fast evolutionary adaptation to the mouse gut are consistent with studies on the evolution of *E. coli* in the mouse gut (Barroso-Batista et al., 2014, 2015; Barroso-Batista and Pedro et al., 2020; Fabich et al., 2011; Ghalayini et al., 2019; Giraud et al., 2001, 2008; Leatham et al., 2005; Lescat et al., 2017; Lourenço et al., 2016; De Paepe et al., 2011; Poulsen et al., 1995; Welling et al., 1980), where selection of mutations related to metabolic capabilities also occurs (Barroso-Batista et

al., 2014; Sousa et al., 2017). While host immunity can affect *E. coli* evolution in the mouse gut (Barroso-Batista et al., 2015), the pattern of *E. coli* evolution more strongly reflects the gut metabolic environment, which in turn is shaped by the gut ecology (Barroso-Batista and Pedro et al., 2020). The studies with *E. coli* raised the question of whether the quick evolution of *E. coli* to changing metabolic environments would be possible only given the genetic and metabolic versatility of *E. coli*, a facultative anaerobe capable of fermenting diverse sources of simple compounds as carbon and nitrogen sources. Our results, beyond addressing the role of diet in microbiota adaptation, show that quick evolution also occurs in *B. theta,* a more abundant and well-adapted strict fermentative anaerobe member of the microbiota, specialized in consumption of complex polysaccharides. By studying the evolution of *B. theta* our results show that *B. theta* indeed has a high degree of genetic versatility and metabolic plasticity. Therefore, our current findings in *B. theta* are likely applicable to other members of the gut microbiota and should inspire future studies to assess the importance of within species evolution in response to dietary changes and other gut perturbations.

Through integration of mutational data, metabolomics and microbiota composition, we show that that recently emerged mutations (<3 months) are a stronger signature of the past host diet than the microbiota composition and are at least on par with the metabolite composition in the feces. Different dietary regimens easily differentiated the genetic signatures of evolutionary adaptation. Our results suggest that microbial evolution is an overlooked biomarker and cause for the microbiota-mediated functional effects of diet which go beyond the microbiota composition. Therefore, the power of intra-specific genetic diversity as a biomarker for host diet should be investigated for other environmental factors affecting the gut ecology. Taken together, our findings emphasize the need to consider the evolution of the microbiota as an important mechanism involved in shaping microbiota functions. Such information needs to be taken into account to understand the microbiota-dependent host responses to diet and most likely to other environmental perturbations.

ACKNOWLEDGMENTS

We thank Joao Xavier and Roberto Balbontín for critical reading of the manuscript. We further thank the members of the Bacterial Signalling Lab an Evolutionary Biology lab for discussions throughout this work. In particular, we thank Anka Konrad and members of the IGC Genomics Facility for help with sequence analysis. We thank J. L. Sonnenburg for sending plasmids for strains construction. This work was supported by Portuguese national funds from Fundação para a Ciência e Tecnologia (FCT) under the project PTDC/BIA-MIC/30487/2017 in addition to support by ONEIDA (LISBOA-01-0145-FEDER-016417) and CONGENTO (LISBOA-01-0145-FEDER-022170) co-funded by FEEI (Fundos Europeus Estruturais e de Investimento from Programa Operacional Regional Lisboa 2020.) K.B.X. is supported by FCT-Investigator IF/00831/2015. The NMR data were acquired at CERMAX, ITQB-NOVA, Oeiras, Portugal with equipment funded by FCT, project AAC 01/SAICT/2016. Whole Genome Sequencing data have been deposited in the NCBI Sequence Read Archive with accession PRJNA749657, while 16S rRNA sequences are available in the public data repository Zenodo (doi:10.5281/zenodo.5137478).

AUTHOR CONTRIBUTIONS

T.D. and K.B.X. conceived and designed the experiments. T.D. and M.F.P. performed the experiments. R.S.R. processed sequencing analysis, optimized and executed the protocols for correlation and prediction analysis. T.D., M.F.P., R.S.R., I.G. and K.B.X. analyzed the data. K.B.X. contributed reagents/materials/analysis tools. T.D. and K.B.X. wrote the original draft of the paper. R.S.R. and I.G. contributed to the discussion and writing of the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE TITLES AND LEGENDS

Figure 1. Rapid evolutionary adaptation of *Bacteroides thetaiotaomicron* (*B. theta*) to the mouse gut under different dietary regimens. (**A**) Schematic representation of experimental design. Green triangles indicate start and end of the dietary regimes. White triangles indicate sample collection for measuring marker frequency, whereas sample collection for Whole-Genome population sequencing (WGS), metabolomics and microbiota profiling are indicated by orange triangles, and by the triangles with black and yellow stroke, respectively. (**B**-**C**) Dynamics of neutral fluorescent marker are shown for mice on Standard Diet (SD) regimen (n=14), or on Western-style Diet (WD) regimen (n=13), respectively. (**D**-**F**) Competitions during mouse gut colonization. Competitions of ancestral bacteria (ANC) against clones evolved on SD (SD_{evol}, n=10) (**D**), on WD ($WD_{evol,}$ n=12) (**E**), or between clones evolved in each diet (n=12) (**F**). For all competitions, competitive indices of evolved populations are shown, and blue background indicates competitions performed in mice fed SD, while red background indicates competitions in mice fed WD. One-sample *t*-test was used to determine whether each dataset is significantly different than 1, *i.e.,* has advantage over the ANC (**D** and **E**). Mann-Whitney *U*-test was used to compare the competitions between the evolved clones (**F**). Each data point indicates an individual biological replicate (individual mouse), and horizontal lines indicate the median of each group. not significant $p > 0.05$, $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{****}p \leq 0.0001$. Time course for all the competitions, frequencies of the fluorescent marker and CFU/g of feces, are presented in Fig. S2 E-L and in Table S1, respectively. (**G**) Dynamics of neutral fluorescent marker are shown for mice on Alternation Diet (AD) regimen (n=14). For (**B-C**) and (**G**) each line shows frequencies of mCherry ($\frac{CFU_{mCherry}}{CFU_{total}}$) in each mouse through the period of 12 weeks (84 days). Lines highlighted in color represent examples of the different dynamic profiles in marker frequency observed in each dietary environment. Separate lines for each mouse, corresponding to 4 independent experiments (four mice cohorts, 8-12 mice each) are shown in Supplement Fig. S1A-D. (**H**) Changes in maker frequency, calculated as ∆Marker frequency = Frequency (Day x) –

Frequency (Day (x-1)), plotted through time (12 weeks). Bold lines present the average of each group, while shadowed surrounding area represents +/- standard deviation. Accumulative marker frequency ($|\Delta t|$) for each group are presented in the insert (***$p < 0.05$; **$p < 0.01$; one-way ANOVA with Tukey's multiple comparisons test).

Figure 2. Genetic basis of *B. theta* evolutionary adaptation to the mouse gut under different dietary regimens. (**A**) Frequencies of mutations (Y axis, dots) and total number of mutations (Z axis, bars) for all the mice (SD group n=14; WD group n=13; and AD group n=14). (**B**) Number of parallel mutations per mouse, each data point indicates an individual biological replicate (individual mouse), and horizontal lines indicate the median for each group. (**C**) Graphical representation of the location of *de novo* mutations. The three circles represent the *B. theta* chromosome, while different colors indicate the dietary regimens of the mice (SD blue; WD red; AD grey). Parallel mutations with frequency >5% are shown in each circle plot. The size of the bars indicates the number of mice in which that particular target appears mutated. See also Table S2, for the complete list of mutations (high and low confidence).

Figure 3. Diet-specific mutational targets of evolutionary adaptation in *B. theta*. (**A**) Maximum frequencies of parallel mutations identified per mouse. (**B**) Shannon index to measure intra-species richness for parallel mutations. ANOVA, followed by a Tuckey test for pairwise comparison, was used to compare the frequencies of parallel mutations and intra-species richness. *$p \leq 0.05$, **$p \leq 0.01$. (SD n=14; WD n=13; AD n=14). (**C**) Mutational targets (genes or intergenic regions) differentially prevalent across dietary regimens ($p<0.05$). Genes that passed a false discovery rate (FDR) of $q<0.1$ are marked in bold. Vertical diet labels (on the left) refer to each dietary regimen(s) for which a given mutational target is enriched, according to pairwise post-hoc test between regimens. For each mutational target predicted functions and substrates are listed. Green and red rectangles around the gene names highlight genes known to be induced in the presence of mucin *O*-glycans, plant polysaccharides or both (Lynch and Sonnenburg, 2012; Martens et al., 2008, 2011). Each

column on the heatmap indicates a mouse, whereas each row indicates a mutational target. Heatmap colors indicate the frequency at which a given mutational target shows mutations in each mouse. Parallel mutations, which did not show diet-specific selection (GLM $p > 0.05$) are listed in Table S2.

Figure 4. Effects of dietary regimen on microbiota composition, metabolites, and interplay correlation network between the three datasets. (**A**) Relative abundance (16S rRNA sequences – ASVs variants) of different phyla from fecal samples collected on the last day of evolutionary experiment for all mice (SD n=14; WD n=13; AD n=14). The relative abundances of each phylum present at >3% are shown (see also Table S4). (**B**) Fecal metabolite concentrations (NMR) on the last day of evolutionary experiment for all the mice. The plots show the absolute concentrations of SCFAs and sugars (µmol/g of feces). For remaining metabolites (organic acids, other metabolites and amino acids) see Fig. S4A-C and Table S3. Post hoc Mann-Whitney *U*-test with Holm's correction for multiple comparisons was used; $*p \leq 0.05$, $***p \leq 0.005$. In all the graphs each data point indicates an individual biological replicate (individual mouse), and horizontal lines indicate the median for each group. Principal component analyses (PCA) for each dataset are presented in Fig. S5A-C. To correlate different datasets: (**C**) mutations versus metabolites, (**D**) mutations versus microbiota composition (16S variants - ASVs), and (**E**) metabolites versus microbiota composition (16S variants - ASVs), we used Procrustes analysis. Plots represent two PCAs, for the two different datasets (mutations - green, metabolites - yellow, and 16S variants - pink), with individual samples being connected by a line. (**F**) Procrustes m^2 statistic (lower values indicate stronger correlation). See also Mantel test (Fig. S4D). (**G**) Multi-omics correlation analysis between (*i*) mutation frequency in parallel mutations; (*ii*) metabolite concentrations; and (*iii*) microbiota composition (ASVs abundance) was carried with HAllA (Hierarchical All-against-All discovery). Areas of the network, whose nodes have >3 edges, are shaded. To generate the correlation network we kept only correlations with FDR≤0.1 and

absolute Spearman correlation *rho>0.5.* Heatmap, representing all correlations for which FDR $p \leq 0.1$, is shown in Fig. S5D.

Figure 5. Mutation can strongly predict dietary regime. (**A-C**) The DIABLO method identifies a set of latent variables, which explain more variance across the datasets. Plots from Variate 1 vs Variate 2 are presented for (**A**) mutations dataset (mutation frequency), (**B**) metabolites dataset (concentration), and (**C**) microbiota composition (16S variant - ASVs abundances). A total of 3 latent variables were included in the DIABLO model (See also Fig. S5E). (**D**) Cross-validation error rate of the DIABLO model. This method allows us to assess the performance of the DIABLO model in correctly classifying the diet to which each sample belongs (lower error indicates better performance).

Figure 6. Changes in dietary regimens influence intra-species evolution and metabolite profile more strongly than microbiota composition in the mouse gut. (**A**-**D**) Dynamics of mutations, metabolites, and microbiota are shown for two populations: AD1 (left panels) and AD4 (right panels). Pink background on the plots indicates the weeks when mice were on WD diet. **(A)** Each line corresponds to one mutation, and all mutations are shown. Two selected mutational targets, BT0867 and BT2689, are highlighted in red and blue, respectively. See also Table S5 for the full list of mutations. (**B-C**) Temporal dynamics for fecal metabolites, sugars and organic acids, respectively. For other metabolites see also Fig. S6B-C, and Table S3. (**D**) Relative abundance of microbiota composition – phyla levels are shown (see also Table S4 for individual ASVs). (**E**) The box plots show the Aitchinson distance between consecutive time points (for AD1 and AD4, top and bottom panel, respectively). mutations vs. microbiota composition ***p <.0001; metabolites vs. microbiota composition ***p = 0.0001; mutations vs. metabolites non-statistically significant, p = 0.8674; one-way ANOVA with Tukey's multiple comparisons test). (**F**) Principal component analysis (PCA) of temporal dynamics for mice AD1 (top panels) and AD4 (bottom panels). PCA were done on Aitchinson distances for each dataset: mutations (left panel), metabolites (central panel) and microbiota composition (16S variants - ASVs, right panel). Lines connect points of

the same mouse in temporal sequential order. Numbers inside the dots correspond to the week, to which a sample belongs, with week 4 being the first time point of the analysis and 12 being the last. An arrow below the dots indicates the first time of point of the analysis (week 4), and an arrow above the dot indicates the last time point (week 12). Each PCA was calculated per mouse separately and represent PC1 versus PC2 data. For the dynamics of PCA1 and PCA2 through time, and consecutive Aitchinson distance see also Fig. S7A-C.

Figure 7 – Mutations selected under Western-style diet have advantage in consumption of mucin *O*-glycans. (**A**) Number of populations with mutations in BT0867 and BT2689 in SD, WD and AD groups (bars, Y-axis), and their frequencies (dots, Z-axis) at the last timepoint of evolutionary experiment. In case of multiple mutations in a target gene, the sum of all was considered as total frequency (See also Table S2). (**B**) Number of populations with different mutations, which generate the amino acid substitution, in BT0867 (at the last time point) for SD, WD and AD groups (bars, Y-axis) and their frequencies (dots, Z-axis) (See also Table S2). (**C**) Competitions of evolved clones selected from AD1 and AD4 at week 4 (SD week = AD_{SD}) and week 5 (WD week = AD_{WD}) against the ancestral wild type (ANC) under laboratory conditions. Selection coefficients per generation of evolved clones are shown. One-sample *t*-test was used to determine whether each dataset is significantly higher than 0, *i.e.,* has advantage over the ANC. $^{*}p \leq 0.05$. (**D**) Ratio of growth rates of clones from WD week versus SD week (AD_{WD} / AD_{SD}). Unpaired *t*-test was used to compare the competitions between the evolved clones. $^{**}p \leq 0.01$, $^{***}p \leq 0.001$. For all experiments the bars represent the median, and error bars +/-SEM. Experiments were repeated 4 independent times, n=96.

STAR methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents may be directed to, and will be fulfilled by, the Lead Contact, Karina Bivar Xavier (kxavier@igc.gulbenkian.pt).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Animals used in this study were bred at the Instituto Gulbenkian de Ciência (IGC) rodent facility, and maintained in biosafety level-2 animal barrier facility under strict specific pathogen-free (SPF) conditions. Experiments with animals were approved by the Institutional Ethics Committee and the Portuguese National Entity (Direção Geral de Alimentação e Veterinária; Ref. number 008958), which complies with European Directive 86/609/EEC of the European Council.

Animals were arbitrarily assigned to experimental groups. None of the animal experiments were performed blinded. Sample size was chosen according to institutional directives and in accordance with the guiding principles underpinning humane use of animals in research. No statistical analyses were performed to predetermine the sample sizes. All of the experiments were performed independently at least three times.

Mice on Standard Diet regimen consumed autoclaved pellets Rat and Mouse No.3 Breeding (Special Diets Services), while mice on Western-Style regimen consumed γ-irradiated High cholesterol diet for mice, Western – butter fat – diet (ssniff, TD.88137).

Bacterial strains and culture conditions

All strains used were derived from *Bacteroides thetaiotaomicron* VPI-5482, strain DSMZ 2079 from German Collection of Microorganisms and Cell Cultures.

B. theta was routinely cultured in a supplemented brain heart infusion (BHIS) growth medium (Bacic and Smith, 2008) (per litre: combine 37g Brain Heart Infusion Broth powder (Sigma-Aldrich); 1g L-cysteine (free base, Sigma); 10ml hemin solution (dissolve 100mg of hemin powder in 2ml 1M NaOH and 198ml dH2O; Sigma); if required 15g agar (Invitrogen); and upon autoclaving add 20ml 10% NaHCO3 (Sigma) and 200µl vitamin K (Sigma)). Competitions in laboratory conditions were done in modified meat extract media – media with

lower amount of meat extract, to allow assessment of carbon source (modified from (Rakoff-Nahoum et al., 2014). Per litre: combine 4.375g Meat Extract (Sigma); 10g Peptone (BD Bacto); 5g NaCl (Merck); 2.5g Na_2HPO_4 (Merck); 1g L-cysteine (free base, Sigma); 10ml hemin solution (as described above; Sigma); and upon autoclaving add 20ml 10% NaHCO3 (Sigma) and 200µl vitamin K (Sigma)).

All *B. theta* growths were done in anaerobic box (atmosphere after 24h: O_2<1%, CO_2 7-15%, Oxoid AnaeroGen) at 37°C. When required, appropriate antibiotics were added to the medium (30mg/mL gentamicin, and 25 mg/mL erythromycin). For competitions in laboratory conditions anaerobic boxes were incubated with shaking. Liquid culture of *B. theta* was normally grown in 20h, while on agar plates in 48h.

Plasmids pWW3452 and pWW3515, carrying sfGFP and mCherry (Whitaker et al., 2017), respectively, were conjugated into *B. theta* with a modified protocol from previously described protocol (Mimee et al., 2015). Briefly, an overnight culture of a donor *Escherichia coli* S17 λpir strain (grown in Lysogeny Broth (LB) with 100mg/mL ampicillin (Amp), with aeration), carrying the plasmid, was diluted 1000x in LB Amp, and grew to exponential phase (OD_{600}~0.4). Overnight culture of a recipient strain *B. theta* was diluted 250x in BHIS, and grew to OD_{600}~0.3. For mating, 900µl of *E.coli* culture was spin down, and resuspended with 100µl of *B. theta* culture (to a final ratio of 1:10 recipient:donor). A mating mixture of *E.coli- B. theta* was spotted onto BHIS agar (~20µl per spot), and incubated upright in a 37°C incubator aerobically, for 5h. Afterwards the cells were collected by scraping and resuspended in 1ml PBS. Serial dilutions were plated on BHIS Gen Ery, and incubated anaerobically for 48h at 37°C. Obtained colonies were re-isolated on BHIS Gen Ery before further use.

METHOD DETAILS

Antibiotic treatment protocol

Since wild-type strains of *Bacteroides* spp. are unable to stably colonize SPF mice (Cullen et al., 2015; Lee et al., 2013), we used a modified antibiotic treatment protocol (Mimee et al.,

2015) to allow colonization with *B. theta* VPI-5482 strain. Littermate female C57BL/6J littermates mice (with at least 20g of weight, ~10- to 12-week old animals) were separated and moved into individually ventilated cages with high-efficiency particulate air filters, with access to sterilized food and water *ad libitum*. Mice were administered with 0.625g/l of ciprofloxacin HCl and 1g/l of metronidazole in drinking water for 7 days *ad libitum*. To prevent usage of sugar-sweetener in the drinking water, but to ensure mice would not reject the antibiotic water (unsweetened water with pH lower than 6.5 is rejected by the animals), we ensure antibiotic-supplemented water would not have pH lower than 6.5. We diluted 0.625g of ciprofloxacin HCl in 20ml 0.1N HCl, and added it to 980ml of drinking water, which already contained 1g of metronidazole. We replaced antibiotic-supplemented water daily, due to the low percent of acid in the water ciprofloxacin HCl would precipitate faster. After 7 days we administered 100mg/kg of metronidazole by oral gavage for 3 days (every 24h), while keeping ciprofloxacin HCl (0.625g/l) dissolved in drinking water. After the third gavage with metronidazole mice were provided with non-supplemented drinking water for 48h. Animals were transferred to a clean cage on the third and eighth day of the antibiotic treatment.

Mouse colonization, sample collection and fluorescent marker dynamics

Overnight cultures of two isogenic strains of *B. theta* VPI-5482, labelled with sfGFP or mCherry fluorescent proteins, were washed in PBS, OD_{600} normalized (measured by NanoDrop 2000), and mixed at 1:1 ratio. 100µL of a suspension of ~10^8 colony forming units (CFUs) was used to gavage SPF-antibiotic treated mice (48h post antibiotic treatment). At 48h post gavage with bacteria, a period which allowed *B. theta* to colonize the mouse gut, the dietary regimen occurred was changed for the first time.

To follow marker frequency, fecal pellets were collected for 12 weeks on the first (day 0) and last day of the experiment (day84), and on each week on the 1st, 4th, and 6th day after changes in dietary regimen (corresponds to 1 day prior to following changes).

Fecal pellets for studying microbiota composition (3 pellets per time point) were collected on the first (day 0), prior to antibiotic treatment, and last day of the experiment (day

84), and on the 6th day post dietary regime change. Samples for studying metabolomics (3 pellets per time point) were collected on the 6th day after change in dietary regime, from week 4 onwards.

Samples for inflammation state of the animals (1 pellet per time point) were collected on the last day of evolutionary experiment (day 84, week 12). Fresh fecal pellets for following marker frequency were collected directly from mice, while feces microbiota composition and metabolomics (6 pellets in total) were harvested from each mouse by placing them in a sterile plastic cage until it defecated. The mass of all collected pellets was recorded for further analysis.

To follow marker frequency, fresh fecal pellets were diluted in PBS, and serial dilutions were plated on BHIS Gen Ery. Plates were incubated anaerobically for 48h at 37°C. Prior to determining the fluorescence color of the colonies, plates were exposed to atmospheric oxygen for at least 1h at 4°C, to allow fluorescent proteins to fold efficiently. Afterwards, the frequencies of sfGFP- and mCherry-labelled bacteria were assessed by counting fluorescent colonies with LED Transilluminator (Nippon Genetics Europe, GmbH). A sample of each collected fecal pellet was also cryopreserved in 25% glycerol at -80°C for further analysis.

Evolutionary experiment was repeated in 4 independent experiments, for a total of 41 mice (SD group 14 mice, WD group 13 mice, AD group 14 mice).

Competitions in animals

For competitions in animals, 30 colonies were isolated from 4 bacterial populations from the last day of evolutionary experiment (day 84, week 12). A dilution of a fecal pellet glycerol stock from SD6, SD14, WD6 and WD11 was plated on BHIS Gen Ery. 30 sfGFP colonies for SD6 and WD6, and 30 mCherry colonies for SD14 and WD11 were isolated, inoculated in liquid BHIS Gen Ery and anaerobically grown overnight at 37°C. Overnight cultures were washed in PBS and normalized by measuring OD_{600} with NanoDrop 2000. An equal mixture of clones (1:30) for 4 populations was then stored in 25% glycerol at -80°C. Wild type *B.*

theta, carrying sfGFP or mCherry, were grown from one colony for each color, prepared and stored in the same way as mixture of colonies.

After the antibiotic treatment (as described above), the mice were gavaged with the glycerol stock mixtures of (i) population with WT; or (ii) population with population, to a ratio of 1:1. 100µL of a suspension of ~10^8 CFUs was used to gavage the mice. Upon gavage, the feed was changed for the group of mice for WD dietary regime.

Fresh fecal pellets were collected daily, for 5 days, and fluorescence marker frequency was determined as described above. Competitions were repeated from 2 to 4 independent times.

Measurement of gut inflammation

We measured gut inflammation via the concentration of lipocalin-2 as previously described (Barreto et al., 2020). Briefly, frozen feces were thawed and homogenized in the appropriate volume of 1X PBS to reach a final concentration of 100mg/mL. After extensive vortexing (5 min) the samples were centrifuged for 30min at 18000g and 4°C. From the recovered supernatant we assayed the concentration of lipocalin-2 by using a Mouse lipocalin-2/NGAL DuoSet ELISA (R&D Systems), as indicated by the manufacturer.

Sequencing and analysis

Ancestral genomes sequencing

One colony of the ancestral clone *B. theta* VPI-5482 (without fluorescent protein tag), one colony of *B. theta* tagged with sfGFP, and one colony of *B. theta* tagged with mCherry, were grown in BHIS at 37°C, anaerobic conditions, for 20h. Bacterial genomic DNA was isolated by following the classic phenol-chloroform extraction method as previously described (Wilson, 2001). Library was constructed at the IGC Genomics facility, using a low-volume Nextera protocol that uses the Tn5 transposase for tagmentation-based library construction (Baym et al., 2015). Genomes were paired-end sequenced using Illumina MiSeq Sequencers, which produced datasets of 250bp read pairs.

Hybrid de novo *genome assembly*

DNA of untagged *B. theta* was additionally sequenced by Oxford Nanopore MinION system, a long-read DNA sequencing technology. The genome was then assembled using a combination of Oxford Nanopore MinION and Illumina Miseq reads. Nanopore reads were first trimmed with PoreChop (v0.2.4) (Wick et al., 2017) and then assembled with the Flye (Version 2.6, with the plasmid option) (Kolmogorov et al., 2019). This was followed by four rounds of polishing with Racon (v1.4.1) (Vaser et al., 2017) and one with Medaka (v0.11.5) (Oxford Nanopore Technologies Ltd., 2018) using the Nanopore reads. Finally, the Illumina MiSeq data was used for two rounds of polishing with Racon (v1.4.1) followed by pilon (v1.23). To evaluate the quality of the assembly, we used CheckM, which showed 99.26% completeness, 0.01% contamination and 0% strain heterogeneity, exactly the same values as for two closely related assemblies: *B. theta* VPI-5482 (GCA_000011065.1 (Washington University Department of Molecular Biology and Pharmacology, 2003)) and *B. theta* DSM 2079 (GCA_014131755.1 (UFZ, 2020)). The latter is PacBio+Illumina assembly for the parent strain of the strain used here.

The final genome assembly included 4 circular contigs, with sizes 6272442bp, 59187bp, 21263bp and 20321bp, with the first two corresponding to the chromosome and the plasmid known to be present in this *B. theta* strain. We then remapped the trimmed Illumina MiSeq reads against the assembly (using *breseq* in clonal mode (Barrick et al., 2009)) and observed that coverage for the two smaller contigs was highly non-homogeneous, with coverage failing to fit a negative binomial distribution for the smallest contig. Thus, we discarded these contigs from further analysis. The plasmid contig also showed non-homogeneous coverage, with most of its length being covered in repetitive regions. This, together with the fact that the two NCBI assemblies mentioned above had plasmid sizes of 33036-33038bp, led us to suspect that the plasmid assembly was incorrectly duplicated (moreover, with different assembly parameters we could recover the plasmid with 33Kbp smaller size, albeit at a cost of having a non-contiguous chromosome). We indeed could align the ~30Kbp of *B. theta* VPI-5482 plasmid at two locations in our plasmid contig,

further reinforcing the idea that the plasmid is misassembled. Thus, we next combined the *de novo* assembly of the chromosome with the sequence of the *B. theta* VPI-5482 plasmid and remapped the trimmed Illumina MiSeq reads against it (*breseq* in clonal mode). In this case, coverage was homogeneous along the length of both the plasmid and the chromosome. While *breseq* did not identify any differences between the reads and the plasmid sequence, it still identified 6 indels relative to the chromosome, which we corrected using the gdtools APPLY function from *breseq.* Our final chromosome assembly has 6272440bp, similar to the 6260361bp and 6271157bp for the chromosomes of *B. theta* VPI-5482 and DSM 2079 strains, with which it had a high Average Nucleotide Identity (*de novo* assembly versus DSM 2079: 99.992%; *de novo* assembly versus VPI-5482: 99.982%; calculated with fastANI v1.32) (Jain et al., 2018).

We used prokka (v1.14.6) (Seemann, 2014), with standard settings, to annotate the *de novo* assembled chromosome of *B. theta* and identified 4975 features, including 4889 CDS (CoDing Sequence). To facilitate understanding the functional relevance of the mutated targets, we used BLAST between genes identified in our assembly and the genes annotated in the *B. theta* VPI-5482 assembly (GCA_000011065.1 (Washington University Department of Molecular Biology and Pharmacology, 2003)), whose gene names are commonly referred in the literature. We attributed the gene names from *B. theta* VPI-5482 if % identity was >97%, query cover ≥97%, and subject cover was ≥97% or ≤103% (genes that fell outside of these thresholds kept the gene name assigned by prokka, i.e. genes named TDAxxxx).

Population sequencing

To obtain DNA from bacterial populations we used a mixture of clones (~1000-5000 colonies). Clones were collected from the plates (BHIS Ery+Gen) by scrapping and resuspending in PBS. To extract the DNA, we used MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Illumina). Additionally, DNA of *B. theta* tagged with sfGFP and mCherry were also sequenced in population analysis mode (~1000-5000 colonies obtained by plating serial dilutions of glycerol stock from frozen vials). For all bacterial populations, and the two populations of ancestral clones tagged with sfGFP or mCherry, the libraries were

prepared as described above and genomes were paired-end sequenced by NextSeq500 (Illumina), which produced datasets of 150bp read pairs.

Genome re-sequencing and mutation inference for evolved populations

All sequenced populations were analyzed with *breseq* 0.34.1 (Barrick et al., 2009), using the following settings: (i) polymorphism frequency cut-off 0; (ii) polymorphism minimum variant coverage each strand 3, (iii) polymorphism bias cut-off 0.05; (iv) minimum mapping quality 20; (v) base quality cutoff 30. These settings were chosen to allow *breseq* to identify all mutations (independent of their frequency), but a mutation is only identified if: (i) at least 3 reads align to each strand; (ii) there is no significant bias ($p<0.05$) in the number of reads aligned to each strand; (iii) the minimum mapping quality is >20; and (iv) base quality is >30. The identified mutations were then filtered to remove low confidence mutations, defined as mutations that: (i) were detected in the ancestral clones and (ii) appeared in regions that had ≥3 mutations in ≤1 Kb within a population and for which the same region was detected across ≥2/3 of the mice. Both high and low confidence mutations are included in Table S2 and S5, but low confidence mutations were not considered for further analysis because these can be sequencing artefacts or mutational hot spots.

Identification of parallel mutations at day 84

Parallel mutations were defined as any gene or intergenic region that was mutated in at least 2 mice (irrespective of diet) and for which the summed mutation frequency at that target was ≥5% in at least one of them. When multiple mutations occur in the same gene or intergenic region, we summed their frequency and use the sum as a final percent.

Identification of diet-specific parallel mutations

We used a Poisson Generalized Linear Model (GLM; using the *brglm2* package (Kosmidis, 2021)) to identify mutational targets that were differentially prevalent across dietary regimens ($p<0.05$) and that had a false discovery rate $q<0.1$. For significant mutational targets, we then applied a *post-hoc* Tukey test (*emmeans* package (Lenth, 2020)) to define whether a mutational target was enriched or depleted for one or two dietary regimens. For example, the post-hoc test indicated that mutations in BT4246 were significantly more prevalent ($p<0.05$)

in the WD group when compared against either the AD or the SD groups. Thus, mutations in this gene were identified as being enriched in the WD group. On the other hand, mutations in gene BT3702 were significantly less prevalent in the SD group versus either the WD or the AD group, but there was no significant difference between WD and AD. Thus, mutations in this gene were identified as being enriched in the SD and AD groups.

Microbiota composition in fecal samples (16S rRNA sequencing)

Fecal samples were collected during the evolutionary experiment and stored at −80°C utill processed. DNA extraction was performed as previously described (Thompson et al., 2015). Briefly, samples were disrupted by using 0.3g of 0.1-mm glass beads (Scientific Industries SI-BG01) and bead-beaten using a QIAGEN Tissuelyser II (Retsch) for 2X 1min with 30 rev/s pulses. Next, DNA extraction was done using the QIAamp DNA Stool Mini Kit (QIAGEN) according to the manufacturer's instructions. Final samples were diluted to a final volume of 100µl in ATE buffer and stored at -20°C. 16S rRNA gene was amplified in triplicate using the 515F/806R (V4 regions) primer pairs recommended by the Earth Microbiome Project under the following PCR cycling conditions: 94°C for 3min; 35 cycles of 94°C for 60s, 50°C for 60s and 72°C for 105s; and extension at 72°C for 10min (Caporaso et al., 2011, 2012). Library preparation and sequencing (2 × 250bp sequencing by Illumina MiSeq Benchtop Sequencer) was performed at the IGC Genomics Unit.

Qiime2 (v2018.8) (Bolyen et al., 2019) was used to analyze the raw sequencing data, using the DADA2 (Callahan et al., 2016) plug-in for quality filtering/trimming, merging of paired-end reads, filtering of chimeras, identification of amplicon sequence variants (ASV) and quantification of their abundances. ASVs with less than 100 reads across the entire dataset were filtered out. Taxonomic assignment was done against the SILVA (Quast et al., 2013) database (version 132) and ASV tables with taxonomic classification were then used as input for the analysis described below.

Metabolomics of gut contents

We performed ^{1}H-NMR (proton nuclear magnetic resonance) analysis on fecal contents to evaluate the composition of the metabolic environment in the mouse gut. Fecal samples were diluted in 1ml of deuterated water (D_2O, 99,9% Sigma-Aldrich). 0.3g of 0.1mm glass beads (Scientific Industries SI-BG01) were added to each tube and samples were bead-beaten using a QIAGEN Tissuelyser II (Retsch), 2X 1min with 30rev/s pulses. To remove large debris and the glass beads, the samples were pelleted by centrifugation at 22000g for 30min at 4°C. Supernatant was collected and filtered through a 0.22mm filter (Milipore). Next, samples were filtered with 3KDa filters (Vivaspin 500) by centrifugation at 15000g and 4°C, for 3h (or until 150mL of filtrate was obtained). Filtered samples were stored at -80°C until spectrum acquisition.

For spectrum acquisition, samples were thawed at room temperature and mixed with 60µl of 350mM phosphate buffer (pH 7.09 with 2% NaN3, 10µl of a 0.05% (w/v) 3-(Trimethylsilyl)propionic-2,2,3,3-d4 (TSP-d4, Sigma-Aldrich) solution, and 380µl of D_2O) to a total volume of 600µl. The mixture was transferred to a 5mm glass NMR tube. All solutions were prepared with D_2O. Samples were homogenized by inversion and the spectra were acquired after pH measurement. Acquisitions were performed on a Bruker AVANCE II+ 500MHz instrument equipped with Cryo TCI (F) (Prodigy) 5mm probehead with z-gradients.

^{1}H-NMR spectra were acquired using 1D NOESY pulse sequence with pre-saturation (noesypr1d) under the following conditions: 90° pulse for excitation mixing time 100ms, acquisition time 4s, and relaxation delay 1s. All spectra were acquired with 200 scans at 25°C, with 48k data points and 6002 Hz (12 ppm) spectral width (Chenomx acquisition parameters). The recorded ^{1}H-NMR spectra were phase corrected using Bruker TopSpin 3.2 and spectra were then analyzed using Chenomx NMR Suite 8.1. Compounds were identified by manually fitting reference peaks to spectra in database Chenomx 500MHz Version 10. Quantification was based on internal standard peak integration (TSP-d4).

Integrative analysis of microbiota, metabolome and mutational profiles

In order to perform the analysis addressed in this section, each of the datasets were first filtered to reduce sparsity. Thus, for the mutational profile, we kept only parallel mutations (as defined above), for the microbiota we kept only the ASVs that were detected in at least 2 mice with a frequency ≥0.05 in at least one mouse (i.e. the same criteria used for parallel mutations), for metabolome we kept all metabolites as sparsity was low. This meant that we kept 73 mutational targets, 24 ASVs, and all 45 metabolites. We applied the centered log ratio (CLR) transformation to each of the datasets (through the microbiome R package (Ernst et al., 2020)), which is a commonly used transformation for integration of omics datasets, as it can account for compositionality (Fernandes et al., 2014; Quinn et al., 2019).

Using this data, we estimated the Aitchinson distance between samples (within each dataset) and carried both Procrustes and Mantel tests with the three pairwise combinations of datasets in order to understand how well one dataset can predict the other. Both tests were performed with the *vegan* R package (Oksanen et al., 2020), with *p* values determined from 999 permutations.

Additionally, to identify correlations between pairs of specific features from different datasets, we used the tool hierarchical all-against-all association (HallA v0.8.17 (Rahnavard et al., 2017)), which is a method to identify associations, specifically designed for high-dimensional data. The network shown in Fig. 4 was then obtained by filtering the pairwise correlations to keep only those with FDR $q<0.1$ and $|rho|>0.5$ and was generated with the R packages ggraph (Pedersen, 2020a), igraph (Csardi and Nepusz, 2006) and tidygraph (Pedersen, 2020b).

In order to understand which datasets could better distinguish between the different diets we used the R package mixOmics (Rohart et al., 2017), which implements the DIABLO method (Data integration analysis for biomarker discovery using latent components), that is based on sparse generalized canonical correlation analysis (Singh et al., 2019)). In order to run DIABLO, we first tested which was the best number of latent variables (using functions *block.splsda* and *perf,* with a max of 5 latent variables). We kept 3 latent variables per dataset and tested which was the number of features per dataset and component that would

be most appropriate, using function *tune.block.splsda.* This identified 6, 25 and 45 features for components 1, 2 and 3 for the mutation dataset; 20, 14 and 12 features for the microbiota and 7, 5 and 7 features for the metabolome. We fitted the final model with the *block.splsda* function and evaluated its performance with the *perf* function, using 5-fold cross validation, repeated 50 times. This allowed us to estimate the cross-validation error rate for classifying samples to each diet.

Purification of Mucin *O*-Glycans

Mucin *O*-glycans were purified from porcine gastric mucus as previously described (Marcobal et al., 2011) with modifications. Briefly, porcine gastric mucin (Type III, SigmaAldrich) was suspended at 2.5% w/v in 100 mM Tris (pH 7.4). The solution was autoclaved for 10 min at 121°C. The solubilized mucin was cooled to 60°C. Proteinase K (GRiSP, Lda.) was added to a final concentration of 0.1mg/ml and proteolysis of mucin was allowed to happen by incubation at 55°C, for 20h. The suspension was mixed well for 1min, and ultracentrifuged for 30min, 21000g (room temperature) to remove insoluble material. The supernatant was collected and NaOH and $NaBH_4$ were carefully added (to prevent formation of bubbles) to final concentrations of 0.1M and 1M, respectively. To release *O*-glycans from the mucin glycoproteins the solution was incubated at 55°C for additional 20h with slow shaking. The pH was subsequently neutralized to 7.4 and solution was dialyzed exhaustively against water (1kDa MW cut-off). To remove remaining insoluble materials an additional step of ultracentrifugation in Ultracentrifuge RC50 with SL1500 rotor at 20412g, 21°C, 30min was performed. Supernatant was collected and neutralized with HCl. The mass yieldwas determined by complete evaporation (with SpeedVac Vacuum Concentrators) of one aliquot of purified mucin *O*-glycan (final concentration 0.3-0.5% w/v).

Experiments in laboratory conditions

All experiments in laboratory conditions were done in modified meat extract media supplemented with 0.25% purified mucin *O*-glycans, 0.1% glucose, or with no saccharides

added. Isolated evolved clones were first isolated from their respective frozen vials (coming from evolutionary experiment), and then individual colonies (3 per isolate) were inoculated separately in BHIS medium without antibiotics and incubated anaerobically overnight (approximately 20h). The following day the cultures were washed in 1XPBS and number of cells in each culture was determinaed by NanoDrop. OD_{600} was normalized to 0.1 for all the cultures. For all clones three independent replicates were used and all the experiments were repeated at least twice.

Competitions between evolved clones and ancestral clone

Evolved clones were mixed with the ancestral clone of the opposite color in a final 1:1 mixture, and 2µl of mixture was added to 148µl medium in a 96-well plate. Plates were incubated anaerobically with shaking for 12h, at 37°C. The initial and final frequencies of the strains were obtained by counting their cell numbers in the Flow Cytometer mentioned below. The relative fitness (selection coefficient per generation) of evolved clones was measured by competitive growth against an ancestral clone *B. theta*. Selection coefficient was calculated as follow:

$$s = \frac{\left(ln\frac{Nevf}{Nancf}\right) - \left(ln\frac{Nevi}{Nanci}\right)}{\left(ln\frac{Nancf}{Nanci}\right)}$$

with $Nevi$ and $Nevf$ being the initial and the final number of evolved bacteria, respectively, and $Nanci$ and $Nancf$ being the initial and the final number of ancestral bacteria.

A BD LSR FortessaTM SORP flow cytometer was used to quantify bacteria, using a 96-well plate High-Throughput Sampler (HTS). To accurately determine cell concentration we also measure volumes using SPHERO fluorescent spheres (AccuCount 2.0 µm blank particles). Final number of bacterial cells was calculated based on the counts of fluorescently labelled bacteria. The instrument was equipped with a 488nm laser used for scatter parameters and sfGFP detection, with a forward scatter (FSC) detector in a photomultiplier tube (PMT) to detect bacteria, and a 561nm laser for mCherry detection. sfGFP and mCherry were measured using bandpass filters in the range of 540/30 nm and 630/75nm, respectively.

The samples were acquired using FACSDiVa (version 9) software, and analyzed using FlowJo (version 10). All Flow Cytometry experiments were done at the Flow Cytometry Facility of the Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal.

Growth rates of the evolved clones

An aliquot of 2µl of evolved clones (as above) was added to 148µl medium in a 96-well plate. Growth was followed by measuring OD_{600} every 2h. To obtain a full growth curve, with minimal disruption of growth (exposure to O_2) all samples were inoculated in two parallel 96-well plates (incubated in two separate anaerobic boxes), and OD_{600} was measured alternately every 2h, in a way that each plate was only exposed to O_2 every 4h. OD_{600} was measured using a microplate reader (VICTOR3™ - Perkin). Growth rate was calculated from the maximum slope of linear regression of $ln\,(OD_{600})$ increase over time.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed in Graphpad Prism 8.2.1 or R software: https://www.r-project.org/, using the procedures described in the previous sections. Statistical details for all tests performed can be found in the figure legends.

DATA AND CODE AVAILABILITY

Whole Genome Sequencing data have been deposited in the NCBI Sequence Read Archive with accession PRJNA749657, while 16S rRNA sequences are available in the public data repository Zenodo (doi:10.5281/zenodo.5137478). Code will be available at https://github.com/ramiroricardo/BthetaDietEvolution.

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

Figure S1 – Shifts in marker frequency as a proxy for evolutionary adaptation of *B. theta* to the mouse gut. (*See also Fig. 1*) Marker frequency of red-labelled *B. theta* carrying

mCherry marker, in each mouse separately - Standard Diet (SD) (n = 14, left column); Western-style Diet (WD) (n = 13, middle column); and Alternation Diet (AD) (n = 14, right column). (**A**) First evolutionary experiment. (**B**) Second evolutionary experiment. (**C**) Third evolutionary experiment. (**D**) Forth evolutionary experiment. Names of each mouse are indicated on the corresponding graph. Each evolutionary experiment represents a biologically independent experiment: (i) independent initial bacterial ancestral clones were used for the gavage; (ii) and independent mice cohort.

Figure S2 – Physiological state of the animals (weight, fat pads and inflammation), microbiota composition, and competitions. (*See also Fig. 1*) (**A**) Average mouse weight (grams) during the period of 12 weeks (84 days). Mean values for each group ± SD are plotted. (*p<0.05; **p<0.01; ***p<0.005; two-way ANOVA). Blue stars indicate the difference between SD and WD groups, while grey stars indicate difference between WD and AD groups. (**B**) Inguinal fat pads (grams) from 5 arbitrarily selected animals from each group (*p≤ 0.05, **p≤ 0.01; Mann-Whitney *U*-test). (**C**) Inflammation state of the host animal. Levels of Lipocalin-2 measured in the fecal samples of all the animals (bars SD, WD, and AD) at the last day of the evolutionary experiment, week 12 (Day 84). (**D**) Fecal bacterial microbiota compositions measured by 16S rRNA amplicon sequence variants (amplicon sequence variants – ASVs). Plots represent average of the relative abundance of each taxon (the lower taxonomic levels) existent at >3%. All other taxon were combined in the category “Others”. Samples were analyzed for all the mice (for SD group n=14, for WD group n=13, for AD group n=14) across time (starting from before the antibiotic treatment – UT – untreated). (**E-L**) *De novo* mutations accumulated during evolutionary adaptation are diet specific. (**E-H**) Frequency of a fluorescent marker of a population of 30 clones of the same color from the day 84 (winning color in the evolutionary experiment, Fig.1 and Fig.S1) were competed against the ancestral clone (ANC) of the opposite color. (**E**) Clones from population SD6 (sfGFP winning population) were competed against ANC tagged with mCherry (Ma1-5). (**F**) Clones from population SD14 (mCherry winning population) were competed against ANC

tagged with sfGFP (Mb1-5). (**G**) Clones from population WD6 (sfGFP winning population) was competed against ANC tagged with mCherry (Mc1-6). (**H**) Clones from population WD11 (mCherry winning population) were competed against ANC tagged with sfGFP (Md1-6). (**I-L**) Competition of SD_{evol} against WD_{evol}. (**I** and **K**) 30 clones from population SD6 (sfGFP winning population) were competed against 30 clones from population WD11 (mCherry winning population) (Ma1-6 and Mc1-6). (**J** and **L**) 30 clones from population SD14 (mCherry winning population) were competed against 30 clones from population WD6 (sfGFP winning population) (Mb1-6 and Md1-5). Background of each graph indicates the diet consumed by the animals during the competition – white background stands for SD diet, pink background from WD diet. Each line presents competition in individual mouse and the color of the line indicates which frequency is presented: Blue line - Frequency of a fluorescent marker of SD_{evol} is presented; Red line - Frequency of a fluorescent marker of a WD_{evol} is presented. Each panel (from **E** to **L**, two graphs for each – white and pink background) present individual experiment.

Figure S3 – Characterization of *de novo* mutations from the last day of the evolutionary experiment. (*See also Fig. 2*) (**A**) Mutation spectra. Each bar presents mutations in every mouse. Y-axis indicates the frequency of mutation type. Mutation types are indicated with different colors (color panel on the right). Black dots indicate the frequency of SNPs, and white dots indicate the frequency of mutations which occur in the coding region. Indel – insertion/deletion; IS - Insertion sequence; SNP – single nucleotide polymorphism. (**B**) Box plot is presenting the fraction of parallel mutations among all mutations (median indicated).

Figure S4 - Fecal metabolites and microbiota. (*See also Fig. 4*). (**A**-**C**) The concentrations of metabolites (µmol/g) in the feces were measured using NMR. The segments represent the absolute concentrations of (**A**) organic acids, (**B**) other metabolites, and (**C**) amino acids (*See also* Table S3). For all analysis Post hoc Mann-Whitney U-test with Holm's correction for multiple comparisons was used; *$p \leq 0.05$, ** $p \leq 0.01$ ***$p \leq 0.005$. All the samples from

the last day of the evolutionary experiment, Day 84 – week 12 (SD=14 – blue dots, WD=13 – red dots, and AD=14 – grey dots) were analyze. In all the graphs presented, each data point indicates an individual biological replicate (individual mouse), and horizontal lines indicate the median for each group. (**D**) Statistics obtained with the Mantel test are presented. From Mantel test we obtain the *r* statistic, for which larger values indicate stronger correlation.

Figure S5 – Multi-omics analysis of the three different datasets. (*See also Fig. 4 and Fig. 5*). (**A-C**) Principal component analysis (PCA) on Aitchinson distances for each dataset. (**A**) Mutation frequency per gene for parallel mutations (*i.e.*, genes that were mutated in at least two mice and for which the frequency >5% in at least one of the mice); (**B**) Metabolites (raw data); and (**C**) Microbiota composition (filtered ASVs, which appeared in at least two mice at >5% relative abundance). For all datasets the following statistical analysis were performed: (i) Permanova and *betadisper* on the Aitchinson distances; and (ii) ANOVA on the values of each principal component followed by a post-hoc Tukey test. (**A**) For mutations, both Permanova and *betadisper* are significant and thus it is not possible to determine if the difference between the diets is on the centroid or the dispersion of the points. ANOVA for each PC shows significant differences (p<0.05) in PC1 (WD is significantly different from all other diets) and PC2 (SD is significantly different from all other diets). (**B**) For metabolites, Permanova test is significant, while *betadisper* is not, which indicates an effect of treatment on the centroid of the Aitchinson distances. ANOVA for each PC shows significant differences in PC1 (WD is significantly different from all other diets) and no significant effect on PC2. (**C**) For microbiota, both Permanova and *betadisper* are significant. ANOVA for each PC shows no significant differences in PC1 and a significant effect on PC2 (AD is significantly different from all other diets). (**D**) Heatmap showing all correlation values obtained from the analysis with HAllA, where FDR *p*-value<0.1 (*See also Fig. 4G*). Names of genes with mutations, metabolites, and microbiota composition (ASVs), are shown. Colors indicate Spearman *rho*. (**E**) Ordination plots for DIABLO latent Variate 1 vs Variate 3 (top

panel), and Variate 2 vs Variate 3 (bottom panel), for the following datasets: mutations (left), metabolites (center), and microbiota composition (ASV variants) (right) (*see also Fig. 5A-C*).

Figure S6 - Analysis of temporal dynamics sequences. (*See also Fig. 6*). Global summary of temporal dynamics for the three datasets: relative abundance of mutations, metabolites and microbiota composition. The following graphs quantitatively summarize the dynamics. The X-axis shows the number of fluctuations that each dataset undergoes. The Y-axis shows the maximum frequency change that each dataset undergoes. This indicates that even when majority of mutations seem to fluctuate only once, the mutations that change the most in frequency are those which fluctuate 4 times. For all plots: Mouse AD1 – yellow line; Mouse AD4 – green line. (**B**-**C**) Temporal dynamics for fecal metabolic compositions for two mice, Mouse AD1 (left panels) and Mouse AD4 (right panels). The concentrations of metabolites (μmol/g of feces) were measured using NMR. The panels represent the absolute concentrations of (**B**) amino acids, and (**C**) other metabolites. (*See also Table S3*).

Figure S7. Aitchinson distances. (*See also Fig. 6*). Temporal dynamics for (**A**) principal component 1 (PC1), and (**B**) principal component 2 (PC2) across time for mutations (green line), metabolites (yellow line), and microbiota composition (ASVs, pink line). PCAs for AD1 - top panel; and PCAs for AD4 - bottom panel. Mutations, which appear in at least twice (in two different time points), were used for these analyzes. (**C**) To see the consecutive distances between samples, consecutive Aitchinson distances for Mouse AD1 (top panel) and Mouse AD4 (bottom panel) were calculated. The plots show the Aitchinson distance between consecutive time points. Pooled data are presented in Fig. 6E.

REFERENCES

Albenberg, L.G., and Wu, G.D. (2014). Diet and the Intestinal Microbiome: Associations, Functions, and Implications for Health and Disease. Gastroenterology *146*, 1564–1572.

1283 Bacic, M.K., and Smith, C.J. (2008). Laboratory Maintenance and Cultivation of Bacteroides
1284 Species. Curr Protoc Microbiol. *Chapter 13*.

1285 Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-
1286 Bacterial Mutualism in the Human Intestine. Science (80-.). *307*.

1287 Barreto, H.C., Sousa, A., and Gordo, I. (2020). The Landscape of Adaptive Evolution of a
1288 Gut Commensal Bacteria in Aging Mice. Curr. Biol. *30*, 1102-1109.e5.

1289 Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and
1290 Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with
1291 Escherichia coli. Nature *461*, 1243–1247.

1292 Barroso-Batista, J., Sousa, A., Lourenço, M., Bergman, M.-L., Sobral, D., Demengeot, J.,
1293 Xavier, K.B., and Gordo, I. (2014). The First Steps of Adaptation of Escherichia coli to the
1294 Gut Are Dominated by Soft Sweeps. PLoS Genet. *10*, e1004182.

1295 Barroso-Batista, J., Demengeot, J., and Gordo, I. (2015). Adaptive immunity increases the
1296 pace and predictability of evolutionary change in commensal gut bacteria. Nat. Commun. *6*,
1297 6:8945.

1298 Barroso-Batista, J., Pedro, M.F., Sales-Dias, J., Pinto, C.J.G., Thompson, J.A., Pereira, H.,
1299 Demengeot, J., Gordo, I., and Xavier, K.B. (2020). Specific Eco-evolutionary Contexts in the
1300 Mouse Gut Reveal Escherichia coli Metabolic Versatility. Curr. Biol. *30*, 1049-1062.e7.

1301 Baumler, A.J., and Sperandio, V. (2016). Interactions between the microbiota and pathogenic
1302 bacteria in the gut. Nature *535*, 85–93.

1303 Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., and Kishony, R.K.
1304 (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS One
1305 *10*, e0128036.

1306 Van den Bergh, B., Swings, T., Fauvart, M., and Michiels, J. (2018). Experimental Design,
1307 Population Dynamics, and Diversity in Microbial Experimental Evolution. Microbiol. Mol. Biol.
1308 Rev. *82*, e00008-18.

1309 Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006). Functional genomic and metabolic
1310 studies of the adaptations of a prominent adult human gut symbiont, Bacteroides
1311 thetaiotaomicron, to the suckling period. J. Biol. Chem. *281*, 36269–36279.

1312 Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A.,
1313 Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive,
1314 scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. *37*, 852–
1315 857.

1316 Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P.
1317 (2016). dada2: high-resolution sample inference from illumina amplicon data. Nat. Methods
1318 *13*, 581–583.

1319 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh,
1320 P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of
1321 millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. *108*, 4516–4522.

1322 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens,
1323 S.M., Betley, J., Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial
1324 community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. *6*, 1621–1624.

1325 Carmody, R.N., Gerber, G.K., Luevano, J.M., Gatti, D.M., Somes, L., Svenson, K.L.,
1326 Turnbaugh, P.J., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2015). Diet Dominates
1327 Host Genotype in Shaping the Murine Gut Microbiota. Cell Host Microbe *17*, 72–84.

1328 Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M.,
1329 Arumugam, M., Batto, J.-M., Kennedy, S., et al. (2013). Richness of human gut microbiome

1330 correlates with metabolic markers. Nature *500*, 541–546.

1331 Chen, P., and Zhang, J. (2020). Antagonistic pleiotropy conceals molecular adaptations in
1332 changing environments. Nat. Ecol. Evol. *4*, 461–469.

1333 Consortium, H.M.J.R.S., Nelson, K.E., Weinstock, G.M., Highlander, S.K., Worley, K.C., Huot
1334 Creasy, H., Jennifer, R.W., Douglas B, R., Mitreva, M., Sodergren, E., et al. (2010). A
1335 Catalog of Reference Genomes from the Human Microbiome. Science *328*, 994–999.

1336 Conway, T., and Cohen, P.S. (2015). Commensal and Pathogenic Escherichia coli
1337 Metabolism in the Gut. Microbiol Spectr. *3*, 10.1128/microbiolspec.MBP-0006–2014.

1338 Cotillard, A., Kennedy, S.P., Kong, L.C., Prifti, E., Pons, N., Le Chatelier, E., Almeida, M.,
1339 Quinquis, B., Levenez, F., Galleron, N., et al. (2013). Dietary intervention impact on gut
1340 microbial gene richness. Nature *500*, 585–588.

1341 Crook, N., Ferreiro, A., Gasparrini, A.J., Pesesky, M.W., Gibson, M.K., Wang, B., Sun, X.,
1342 Condiotte, Z., Dobrowolski, S., Peterson, D., et al. (2019). Adaptive Strategies of the
1343 Candidate Probiotic E. coli Nissle in the Mammalian Gut. Cell Host Microbe *25*, 499-512.e8.

1344 Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network
1345 research, InterJournal, Complex Systems 1695.

1346 Cullen, T.W., Schofield, W.B., Barry, N.A., Putnam, E.E., Rundell, E.A., Trent, M.S., Degnan,
1347 P.H., Booth, J.C., Yu, H., and Goodman, J. (2015). Antimicrobial peptide resistance mediates
1348 resilience of prominent gut commensals during inflammation. Science *347*, 170–175.

1349 David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling,
1350 A. V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly
1351 alters the human gut microbiome. Nature *505*, 559–563.

1352 Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo,

N.A., Kitamoto, S., Terrapon, N., Muller, A., et al. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell *167*, 1339-1353.e21.

Deutschbauer, A., and Chen, Y. (2021). Functional genetics of human gut commensal Bacteroides thetaiotaomicron reveals metabolic requirements for growth across environments. Cell Rep. *34*, 108789.

Dominguez-Bello, M.G., Godoy-Vitorino, F., Knight, R., and Blaser, M.J. (2019). Role of the microbiome in human development. Gut *68*, 1108– 1114.

Ernst, F.G.M., Huang, R., Shetty, S., Borman, T., Braccia, D.C., Bravo, H.C., and Leo, L. (2020). R/Bioc microbiome ecosystem with SummarizedExperiment.

Fabich, A.J., Leatham, M.P., Grissom, J.E., Wiley, G., Lai, H., Najar, F., Roe, B.A., Cohen, P.S., and Conway, T. (2011). Genotype and phenotypes of an intestine-adapted Escherichia coli K-12 mutant selected by animal passage for superior colonization. Infect. Immun. *79*, 2430–2439.

Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. (2011). Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice. Science *333*, 101–104.

Fehlner-Peach, H., Magnabosco, C., Raghavan, V., Scher, J.U., Tett, A., Cox, L.M., Gottsegen, C., Watters, A., Wiltshire-Gordon, J.D., Segata, N., et al. (2019). Distinct Polysaccharide Utilization Profiles of Human Intestinal Prevotella copri Isolates. Cell Host Microbe *26*, 680-690.e5.

Fernandes, D., A., Reid, J., Macklaim, M., J., McMurrough, T.A, Edgell, D.R., et al. (2014). Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome *2*, 1–13.

De Filippis, F., Pasolli, E., Tett, A., Tarallo, S., Naccarati, A., De Angelis, M., Neviani, E., Cocolin, L., Gobbetti, M., Segata, N., et al. (2019). Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets. Cell Host Microbe *25*, 444-453.e3.

Flint, H.J., Scott, K.P., Louis, P., and Duncan, S.H. (2012). The role of the gut microbiota in nutrition and health. Nat. Rev. Gastroenterol. Hepatol. *9*, 577–589.

Garud, N.R., Good, B.H., Hallatschek, O., and Pollard, K.S. (2019). Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol. *17*, (1): e3000102.

Ghalayini, M., Magnan, M., Dion, S., Zatout, O., Bourguignon, L., Tenaillon, O., and Lescat, M. (2019). Long-term evolution of the natural isolate of Escherichia coli 536 in the mouse gut colonized after maternal transmission reveals convergence in the constitutive expression of the lactose operon. Mol. Ecol. 00:1–16.

Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S.V., and Knight, R. (2018). Current understanding of the human microbiome. Nat Med *24*, 392–400.

Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., and Taddei, F. (2001). Costs and benefits of high mutation rates: Adaptive evolution of bacteria in the mouse gut. Science (80-.). *291*, 2606–2608.

Giraud, A., Arous, S., Paepe, M. De, Gaboriau-Routhiau, V., Bambou, J.-C., Rakotobe, S., Lindner, A.B., Taddei, F., and Cerf-Bensussan, N. (2008). Dissecting the Genetic Components of Adaptation of Escherichia coli to the Mouse Gut. PLoS Genet. *4*, e2.

Hegreness, M., Shoresh, N., Hartl, D., and Kishony, R. (2006). An equivalence principle for the incorporation of favorable mutations in asexual populations. Science *311*, 1615–1617.

Hooper, L. V, Littman, D.R., and Macpherson, A.J. (2012). Interactions Between the Microbiota and the Immune System. Science *336*, 1268–1273.

1401 Hoskins, L.C., and Boulding, E.T. (1981). Mucin degradation in human colon ecosystems.
1402 Evidence for the existence and role of bacterial subpopulations producing glycosidases as
1403 extracellular enzymes. J. Clin. Invest. *67*, 163–172.

1404 Hryckowian, A.J., Treuren, W., Smits, S.A., Davis, N.M., Gardner, J.O., Bouley, D.M., and
1405 Sonnenburg, J.L. (2018). Microbiota-accessible carbohydrates suppress Clostridium difficile
1406 infection in a murine model. Nat. Methods *3*, 662–669.

1407 Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High
1408 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat
1409 Commun. *9*, 5114.

1410 Johansson, M.E. V., Sjövall, H., and Hansson, G.C. (2013). The gastrointestinal mucus
1411 system in health and disease. Nat. Rev. Gastroenterol. Hepatol. *10*, 352–361.

1412 El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013). The
1413 abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat.
1414 Rev. Microbiol. *11*, 497–504.

1415 Kashyap, P.C., Marcobal, A., Ursell, L.K., Smits, S.A., Sonnenburg, E.D., Costello, E.K.,
1416 Higginbottom, S.K., Domino, S.E., Holmes, S.P., Relman, D.A., et al. (2013). Genetically
1417 dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the
1418 gut microbiota. Proc. Natl. Acad. Sci. U. S. A. *110*, 17059–17064.

1419 Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone
1420 reads using repeat graphs. Nat. Biotechnol. *37*, 540–546.

1421 Kosmidis, I. (2021). brglm2: Bias Reduction in Generalized Linear Models. R package
1422 version 0.7.1.

1423 Krautkramer, K.A., Fan, J., and Bäckhed, F. (2021). Gut microbial metabolites as multi-
1424 kingdom intermediates. Nat Rev Microbiol. *19*, 77–94.

Kreuzer, M., and Hardt, W.D. (2020). How Food Affects Colonization Resistance against Enteropathogenic Bacteria. Annu. Rev. Microbiol. *74*, 787–813.

Leatham, M.P., Stevenson, S.J., Gauger, E.J., Krogfelt, K.A., Lins, J.J., Haddock, T.L., Autieri, S.M., Conway, T., and Cohen, P.S. (2005). Mouse intestine selects nonmotile flhDC mutants of Escherichia coli MG1655 with increased colonizing ability and better utilization of carbon sources. Infect. Immun. *73*, 8039–8049.

Lee, M.S., Donaldson, G.P., Mikulski, Z., Boyajian, S., Ley, K., and Mazmanian, S.K. (2013). Bacterial colonization factors control specificity and stability of the gut microbiota. Nature *501*, 426–431.

Lenth, R. V. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.3.

Lescat, M., Launay, A., Ghalayini, M., Magnan, M., Glodt, J., Pintard, C., Dion, S., Denamur, E., and Tenaillon, O. (2017). Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an Escherichia coli natural isolate in the streptomycin-treated mouse gut. Mol. Ecol. *26*, 1802–1817.

Lourenco, M., Ramiro, R.S., Guleresi, D., Barroso-Batista, J., Xavier, K.B., Gordo, I., and Sousa, A. (2016). A Mutational Hotspot and Strong Selection Contribute to the Order of Mutations Selected for during Escherichia coli Adaptation to the Gut. PLOS Genet. *12*, e1006420.

Lynch, J.B., and Sonnenburg, J.L. (2012). Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a Bacteroides hybrid two-component system. Mol. Microbiol. *85*, 478–491.

Maharjan, R.P., and Ferenci, T. (2017). A shifting mutational landscape in 6 nutritional states: Stress-induced mutagenesis as a series of distinct stress input–mutation output

relationships. PLOS Biol. *15*, e2001477.

Manor, O., Dai, C.L., Kornilov, S.A., Smith, B., Price, N.D., Lovejoy, J.C., Gibbons, S.M., and Magis, A.T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun. *11*, 5206.

Marcobal, A., Barboza, M., Sonnenburg, E.D., Pudlo, N., Martens, E.C., Desai, P., Lebrilla, C.B., Weimer, B.C., Mills, D.A., German, J.B., et al. (2011). Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell Host Microbe *10*, 507–514.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. Cell Host Microbe *4*, 447–457.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., et al. (2011). Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. PLoS Biol. *9*, e1001221.

McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C., Pudlo, N.A., Muegge, B.D., Henrissat, B., Hettich, R.L., et al. (2013). Effects of Diet on Resource Utilization by a Model Human Gut Microbiota Containing Bacteroides cellulosilyticus WH2, a Symbiont with an Extensive Glycobiome. PLoS Biol. *11*, e1001637.

Mimee, M., Tucker, A.C., Voigt, C.A., Lu Correspondence, T.K., and Lu, T.K. (2015). Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. Cell Syst. *1*, 62–71.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2020). vegan: Community Ecology Package. R package version 2.5-7.

Oxford Nanopore Technologies Ltd. (2018). medaka: Sequence correction provided by ONT Research.

De Paepe, M., Gaboriau-Routhiau, V., Rainteau, D., Rakotobe, S., Taddei, F., and Cerf-Bensussan, N. (2011). Trade-Off between Bile Resistance and Nutritional Competence Drives Escherichia coli Diversification in the Mouse Gut. PLoS Genet. *7*, e1002107.

Pedersen, T.L. (2020a). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. R package version 2.0.4.

Pedersen, T.L. (2020b). tidygraph: A Tidy API for Graph Manipulation. R package version 1.2.0.

Png, C.W., Linden, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., McGuckin, M.A., and Florin, T.H.J. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. Am. J. Gastroenterol. *105*, 2420–2428.

Porter, N.T., and Martens, E.C. (2017). The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. Annu. Rev. Microbiol. *71*, 349–369.

Porter, N.T., Canales, P., Peterson, D.A., and Martens, E.C. (2017). A Subset of Polysaccharide Capsules in the Human Symbiont Bacteroides thetaiotaomicron Promote Increased Competitive Fitness in the Mouse Gut. Cell Host Microbe *22*, 494-506.e8.

Poulsen, L.K., Licht, T.R., Rang, C., Krogfelt, K.A., and Molin, S. (1995). Physiological state of Escherichia coli BJ4 growing in the large intestines of streptomycin-treated mice. J. Bacteriol. *177*, 5840–5845.

Pudlo, N.A., Urs, K., Kumar, S.S., German, J.B., Mills, D.A., and Martens, E.C. (2015). Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. MBio *6*, e01282-15.

1497 Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N.,
1498 Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by
1499 metagenomic sequencing. Nature *464*, 59–65.

1500 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Rg Peplies, J., and
1501 Glö Ckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data
1502 processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.

1503 Quinn, T.P., Erb, I., Gloor, G., Notredame, C., Richardson, M.F., and Crowley, T.M. (2019). A
1504 field guide for the compositional analysis of any-omics data. Gigascience *8*, 1–14.

1505 Radman, M., Matic, I., Halliday, J.A., and Taddei, F. (1995). Editing DNA replication and
1506 recombination by mismatch repair: from bacterial genetics to mechanisms of predisposition
1507 to cancer in humans. Philos. Trans. R. Soc. B Biol. Sci. *347*, 97–103.

1508 Rahnavard, G., Franzosa, E.A., McIver, L.J., Schwager, E., Lloyd-Price, J., Weingart, G.,
1509 Moon, Y.S., Morgan, X.C., Waldron, L., and Huttenhower, C. (2017). High-sensitivity pattern
1510 discovery in large multi'omic datasets.

1511 Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. (2004).
1512 Recognition of commensal microflora by toll-like receptors is required for intestinal
1513 homeostasis. Cell *118*, 229–241.

1514 Rakoff-Nahoum, S., Coyne, M.J., and Comstock, L.E. (2014). An ecological network of
1515 polysaccharide utilization among human intestinal symbionts. Curr. Biol. *24*, 40–49.

1516 Ramiro, R.S., Durão, P., Bank, C., and Gordo, I. (2020). Low Mutational Load Allows for High
1517 Mutation Rate Variation in Gut Commensal Bacteria. PLoS Biol. *18*, e3000617.

1518 Relman, D.A. (2020). Thinking about the microbiome as a causal factor in human health and
1519 disease: philosophical and experimental considerations. Curr. Opin. Microbiol. *54*, 119–126.

Rey, F.E., Gonzalez, M.D., Cheng, J., Wu, M., Ahern, P.P., and Gordon, J.I. (2013). Metabolic niche of a prominent sulfate-reducing human gut bacterium. Proc. Natl. Acad. Sci. U. S. A. *110*, 13582–13587.

Rohart, F., Benoıˆt Gautier, B., Singh, A., and Cao, K.-A.L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration.

Salyers, A.A., Vercellotti, J.R., West, S.E., and Wilkins, T.D. (1977a). Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. Appl. Environ. Microbiol. *33*, 319–322.

Salyers, A.A., West, S.E., Vercellotti, J.R., and Wilkins, T.D. (1977b). Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. Appl. Environ. Microbiol. *34*, 529–533.

Schwalm, N.D., Townsend, G.E., and Groisman, E.A. (2017). Prioritization of polysaccharide utilization and control of regulator activation in Bacteroides thetaiotaomicron. Mol. Microbiol. *104*, 32–45.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069.

Singh, A., Shannon, C.P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S.J., and Cao, K.A.L. (2019). DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics *35*, 3055–3062.

Sonnenburg, E.D., and Sonnenburg, J.L. (2014). Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. Cell Metab. *20*, 779–786.

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

1544 Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N.,
1545 Sonnenburg, J.L., Coutinho, P.M., Minx, P., Latreille, P., et al. (2010). Specificity of
1546 polysaccharide use in intestinal bacteroides species determines diet-induced microbiota
1547 alterations. Cell *141*, 1241–1252.

1548 Sonnenburg, E.D., Smits, S.A., Tikhonov, M., Higginbottom, S.K., Wingreen, N.S., and
1549 Sonnenburg, J.L. (2016). Diet-induced extinctions in the gut microbiota compound over
1550 generations. Nature *529*, 212–215.

1551 Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.-H., Westover, B.P., Weatherford, J., Buhler,
1552 J.D., and Gordon, J.I. (2005). Glycan Foraging in Vivo by an Intestine-Adapted Bacterial
1553 Symbiont. Science *307*, 1955–1959.

1554 Sousa, A., Frazão, N., Ramiro, R.S., and Gordo, I. (2017). Evolution of commensal bacteria
1555 in the intestinal tract of mice. Curr. Opin. Microbiol. *38*, 114–121.

1556 The Human Microbiome Project Consortium* (2012). Structure, function and diversity of the
1557 healthy human microbiome. Nature *486*, 207–214.

1558 Thompson, J.A., Oliveira, R.A., Djukovic, A., Ubeda, C., and Xavier, K.B. (2015).
1559 Manipulation of the quorum sensing signal AI-2 affects the antibiotic-treated gut microbiota.
1560 Cell Rep. *10*, 1861–1871.

1561 Turnbaugh, P.J., Bä, F., Fulton, L., and Gordon, J.I. (2008). Diet-Induced Obesity Is Linked
1562 to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. Cell Host Microbe
1563 *3*, 213–223.

1564 Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin,
1565 M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009a). A core gut microbiome in obese
1566 and lean twins. Nature *457*, 480–484.

1567 Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., and Gordon, J.I. (2009b).

The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Sci Transl Med *1*, 6ra14.

Turner, C.B., Marshall, C.W., and Cooper, V.S. (2018). Parallel genetic adaptation across environments differing in mode of growth or resource availability. Evol. Lett. *2*, 355–367.

Ubeda, C., Djukovic, A., and Isaac, S. (2017). Roles of the intestinal microbiota in pathogen protection. Clin. Transl. Immunol. *6*, e128.

UFZ, H.C. for E.R.- (2020). ASM1413175v1.

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746.

Washington University Department of Molecular Biology and Pharmacology (2003). ASM1106v1.

Welling, G.W., Groen, G., Tuinte, J.H., Koopman, J.P., and Kennis, H.M. (1980). Biochemical effects on germ-free mice of association with several strains of anaerobic bacteria. J. Gen. Microbiol. *117*, 57–63.

Whitaker, W.R., Shepherd, E.S., Sonnenburg Correspondence, J.L., and Sonnenburg, J.L. (2017). Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome. Cell *169*, 538–546.

Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genome *3(10):e000*.

Wilson, K. (2001). Preparation of Genomic DNA from Bacteria. In Current Protocols in Molecular Biology, p. 2.4.

Wotzka, S.Y., Kreuzer, M., Maier, L., Arnoldini, M., Nguyen, B.D., Brachmann, A.O., Berthold, D.L., Zünd, M., Hausmann, A., Bakkeren, E., et al. (2019). Escherichia coli limits

Salmonella Typhimurium infections after diet shifts and fat-mediated microbiota perturbation in mice. Nat. Microbiol. *4*, 2164–2174.

Yilmaz, B., Mooser, C., Keller, I., Li, H., Zimmermann, J., Bosshard, L., Fuhrer, T., Gomez de Aguero, M., Fernandez Trigo, N., Tschanz-Lischer, H., et al. (2021). Resource Long-term evolution and short-term adaptation of microbiota strains and sub-strains in mice. Cell Host Microbe *29*, 1–14.

Zhang, C., Zhang, M., Pang, X., Zhao, Y., Wang, L., and Zhao, L. (2012). Structural resilience of the gut microbiota in adult mice under high-fat dietary perturbations. ISME J. *6*, 1848–1857.

Zhao, S., Lieberman, T.D., Poyet, M., Kauffman, K.M., Gibbons, S.M., Groussin, M., Xavier, R.J., and Alm, E.J. (2019). Adaptive Evolution within Gut Microbiomes of Healthy People. Cell Host Microbe *25*, 656–667.

Figure 1

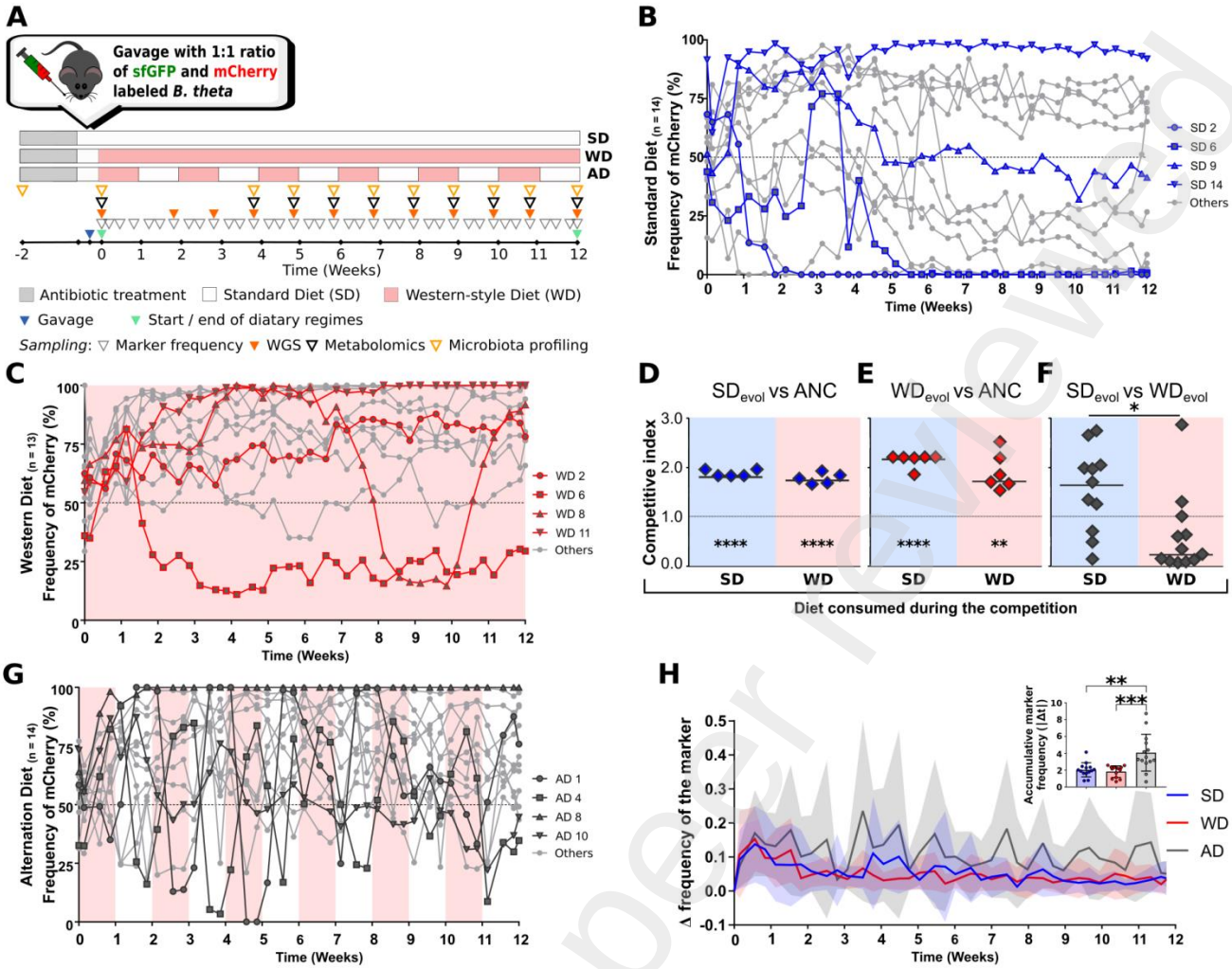

Figure 2

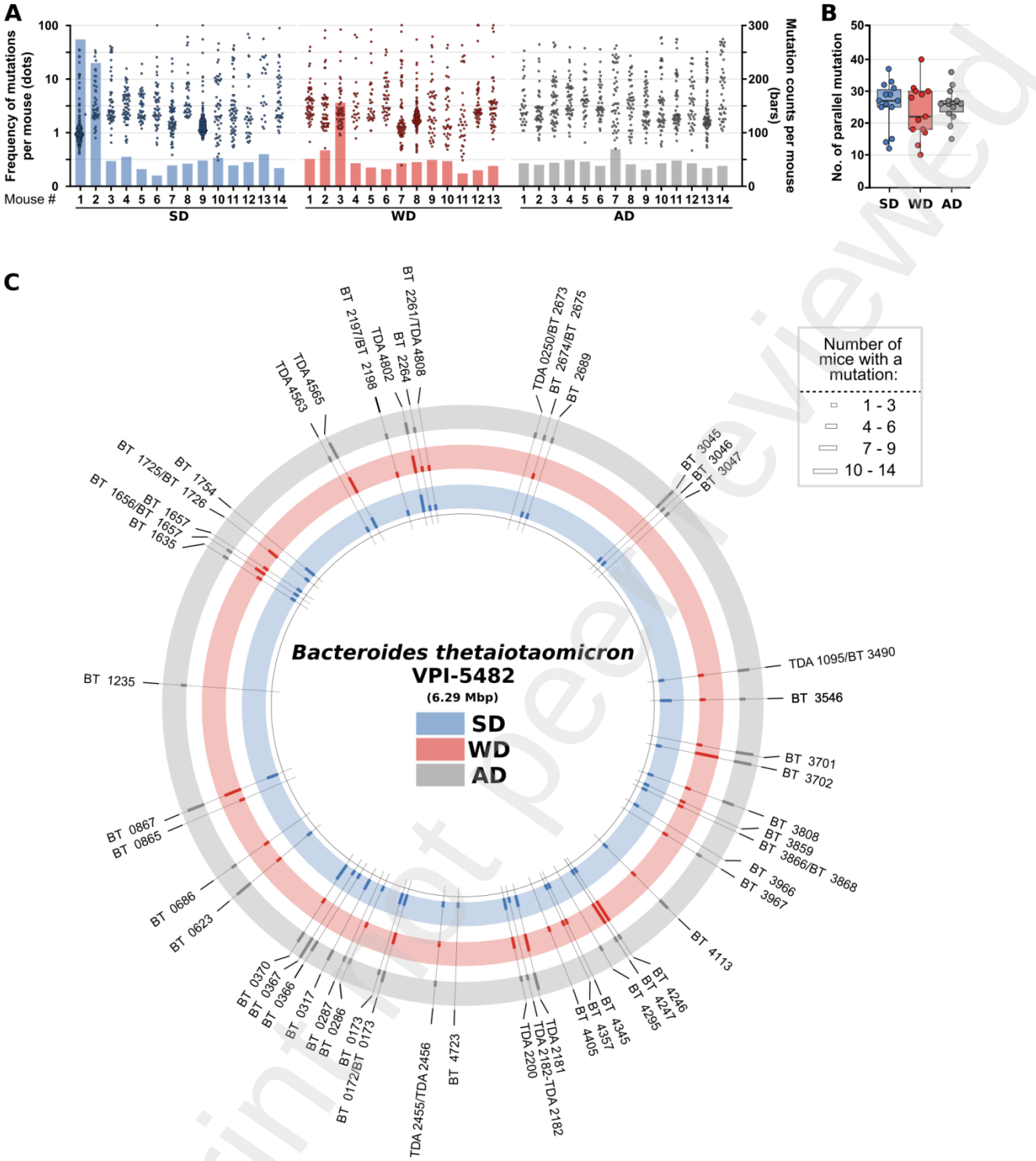

Figure 3

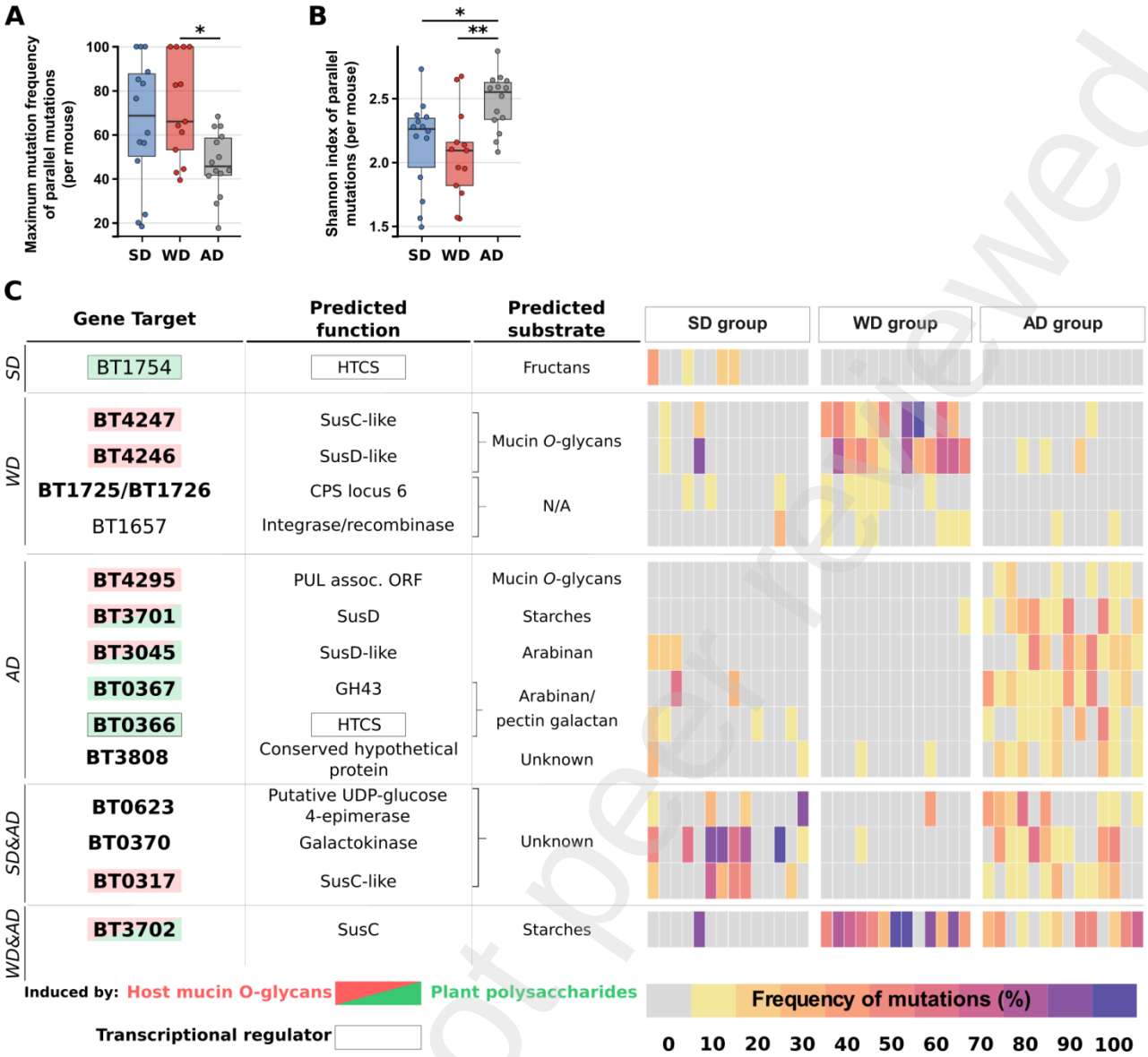

Figure 4

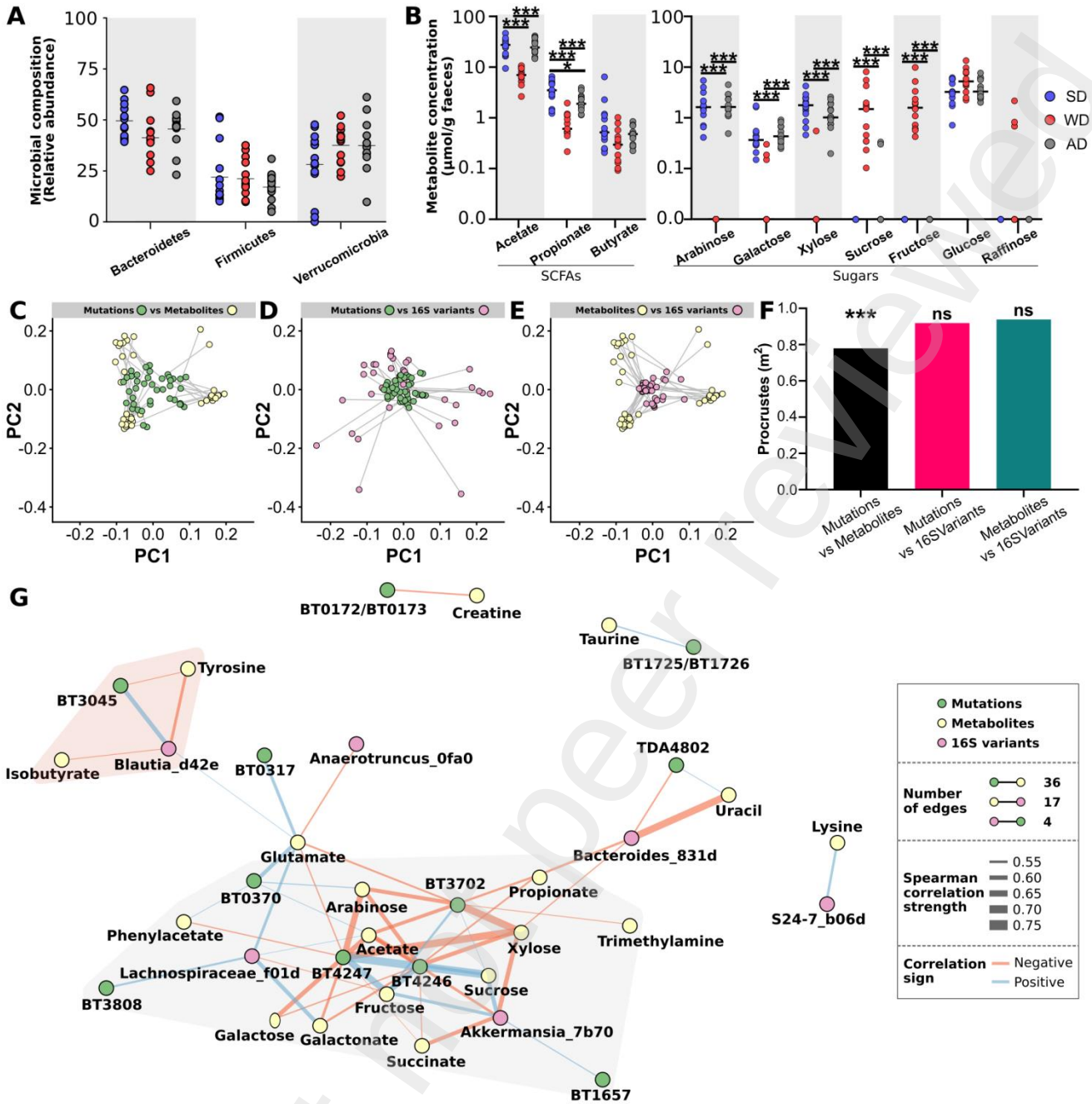

Figure 5

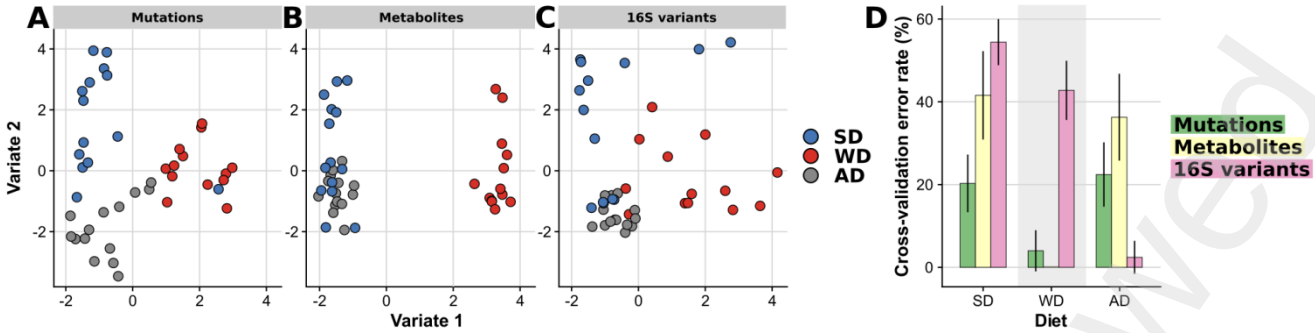

Figure 6

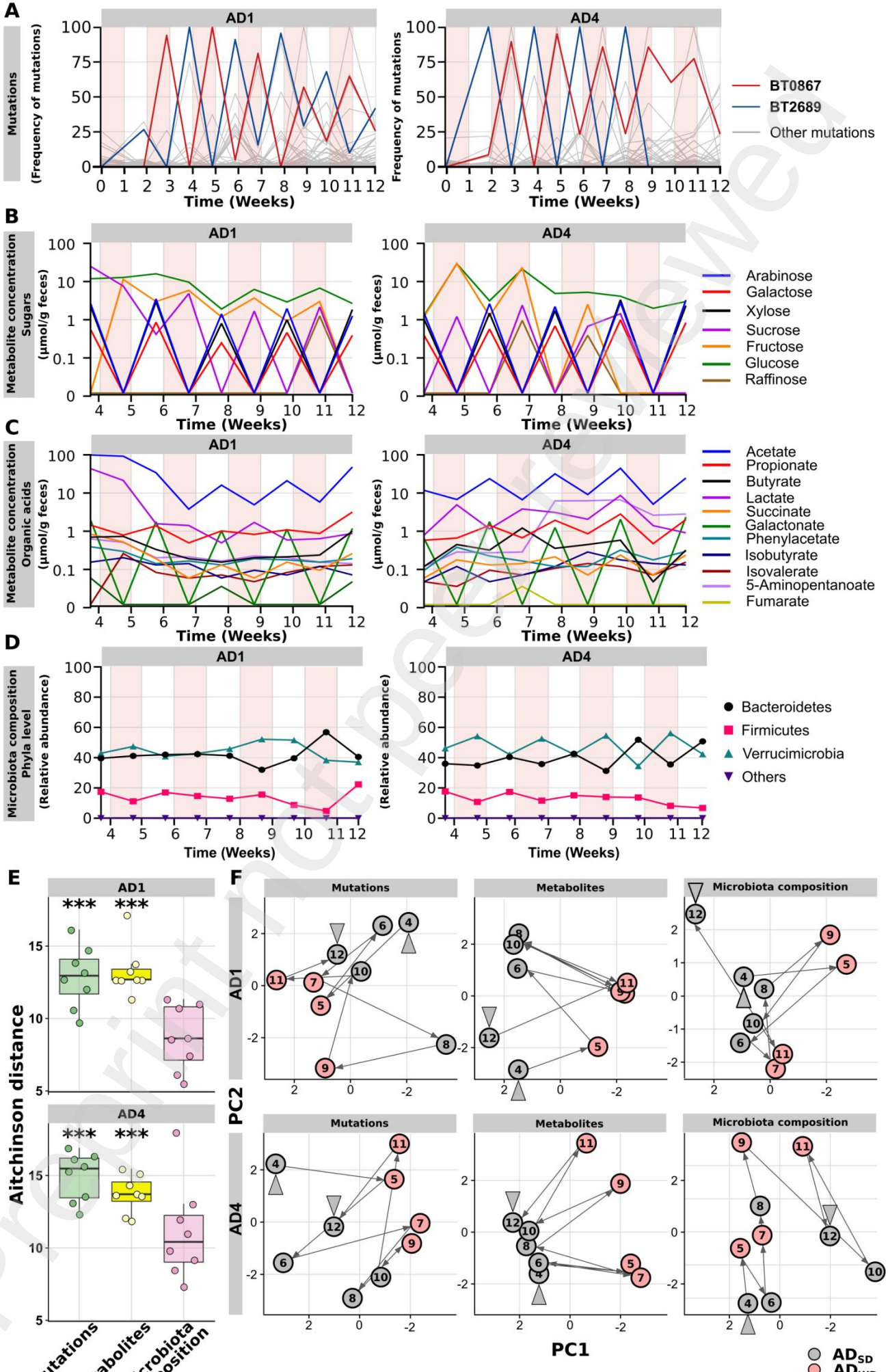

Figure 7

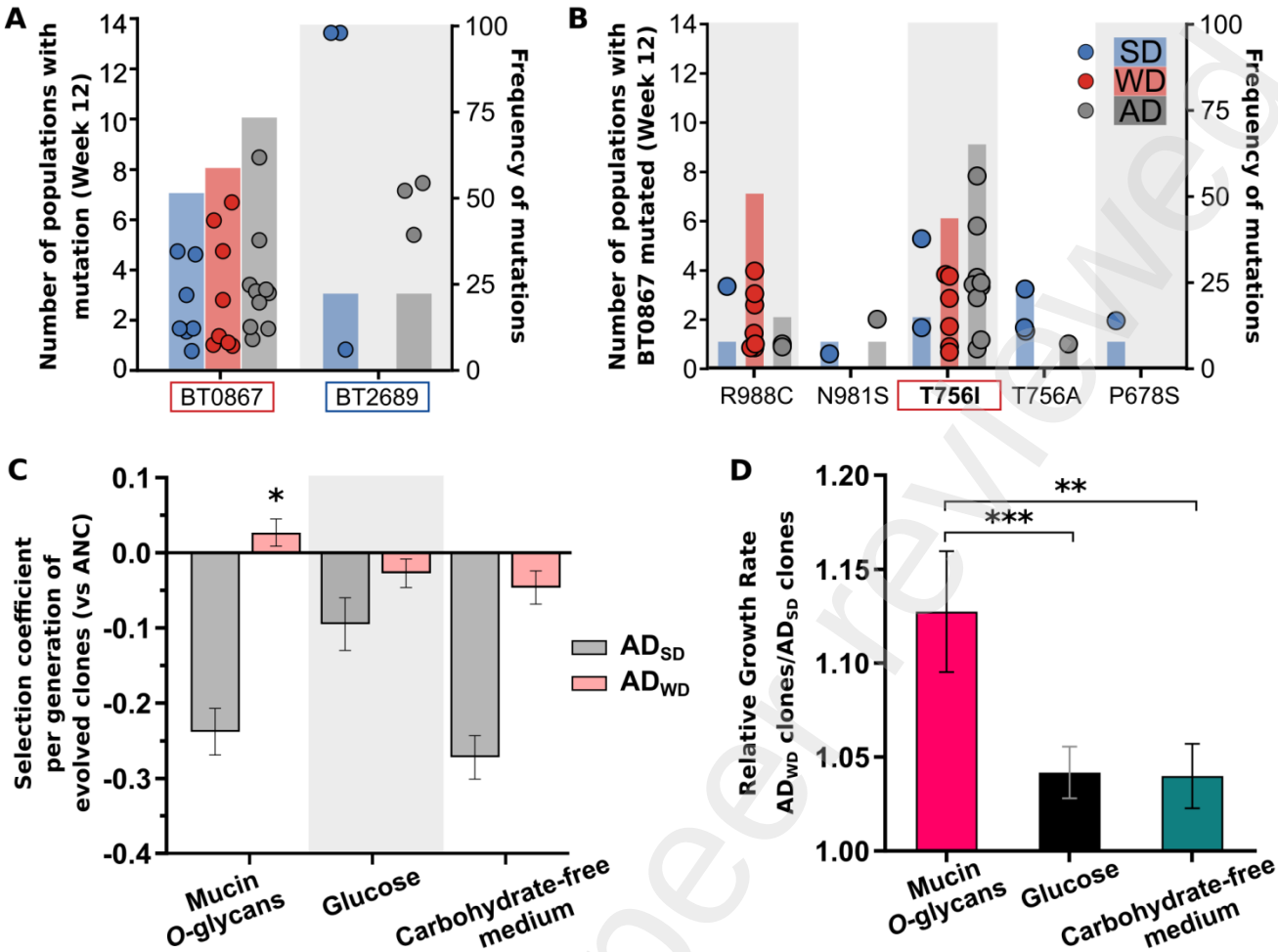

Figure S1

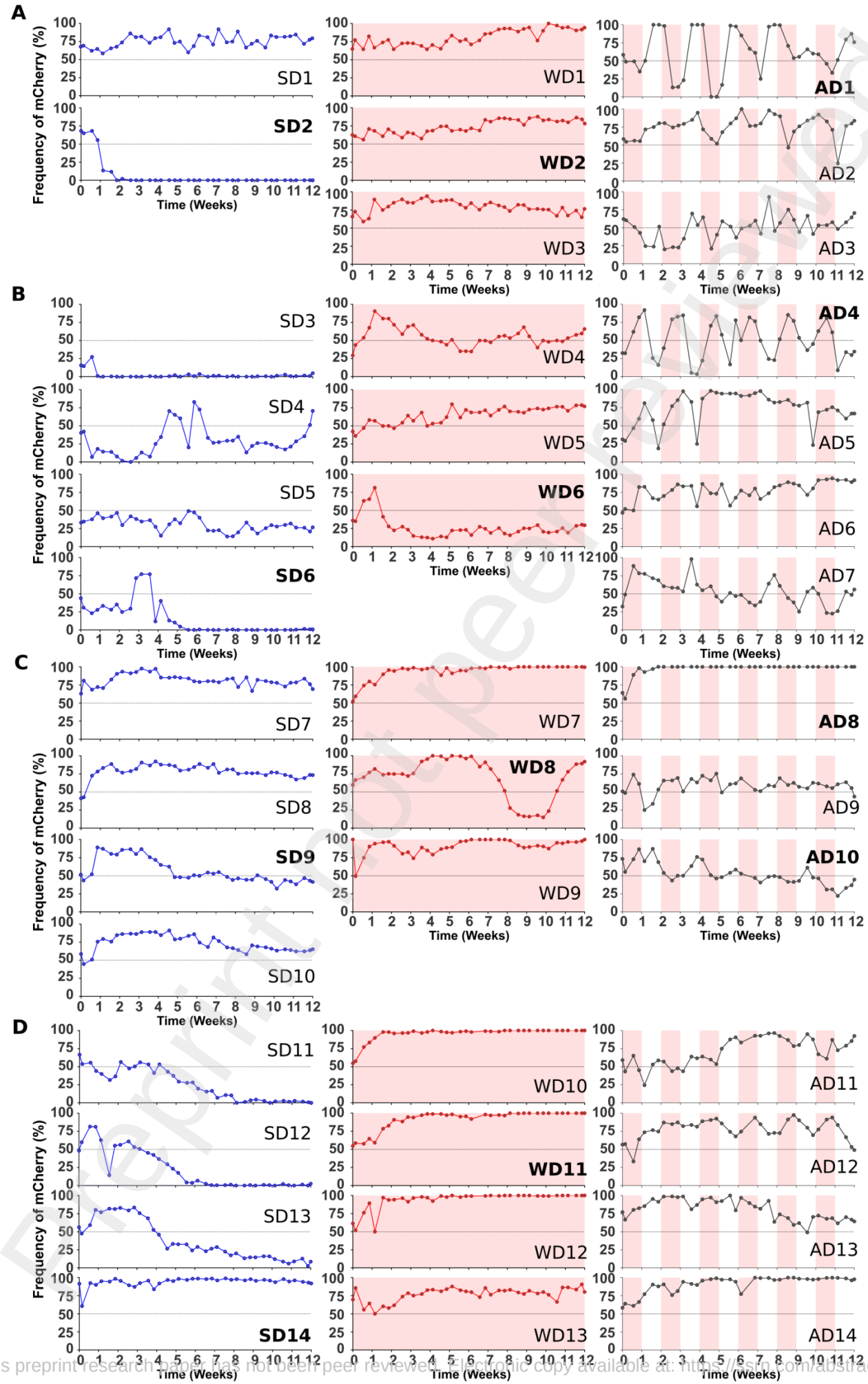

Figure S2

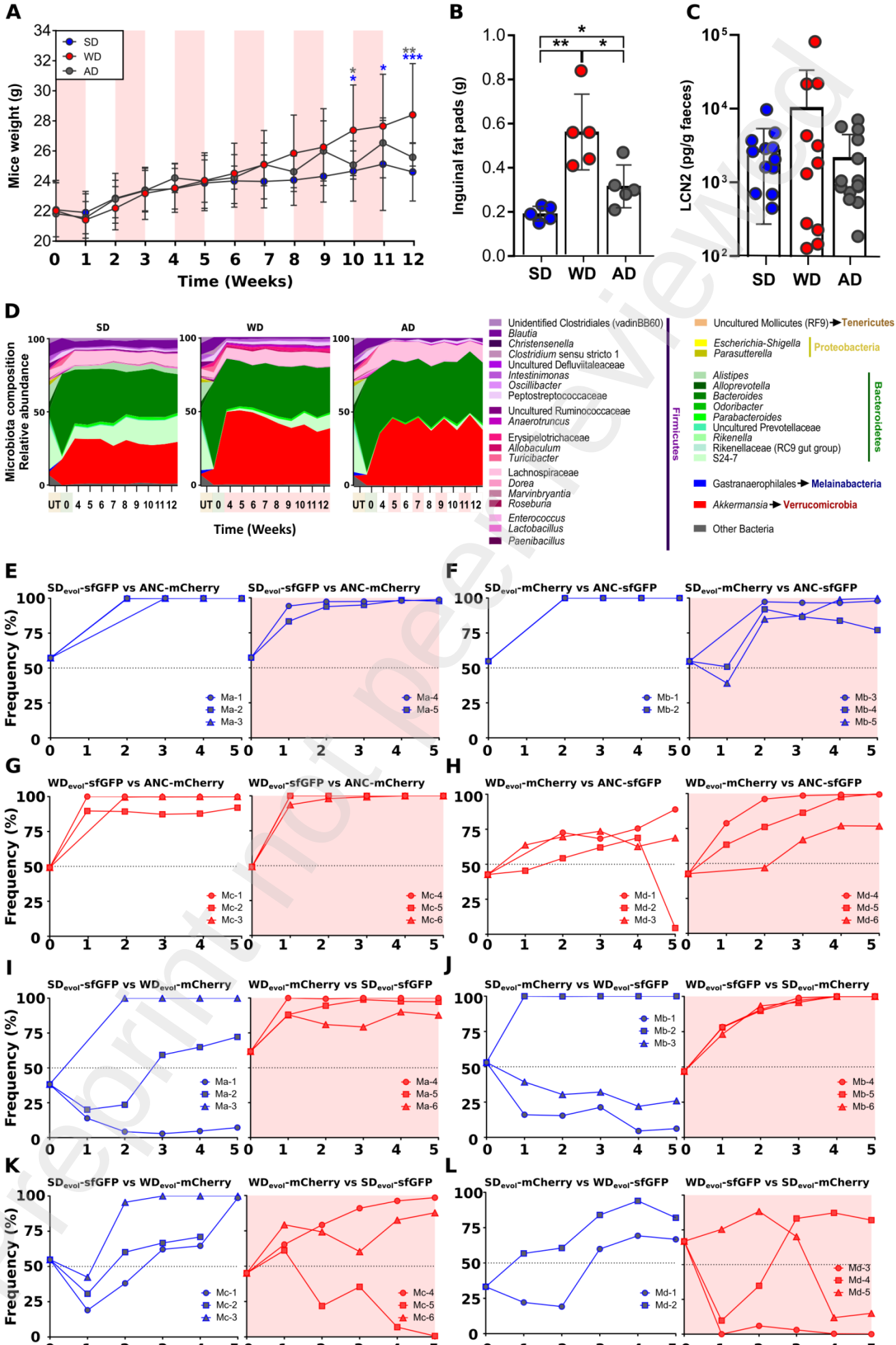

Figure S3

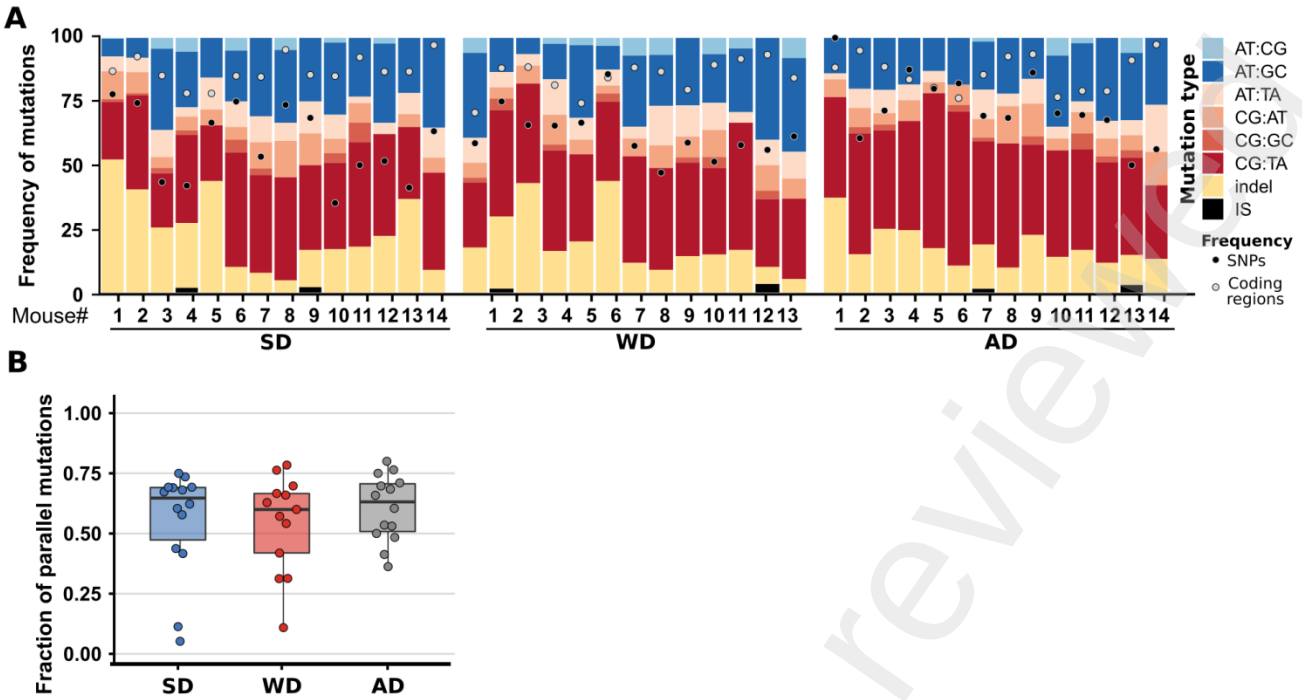

Figure S4

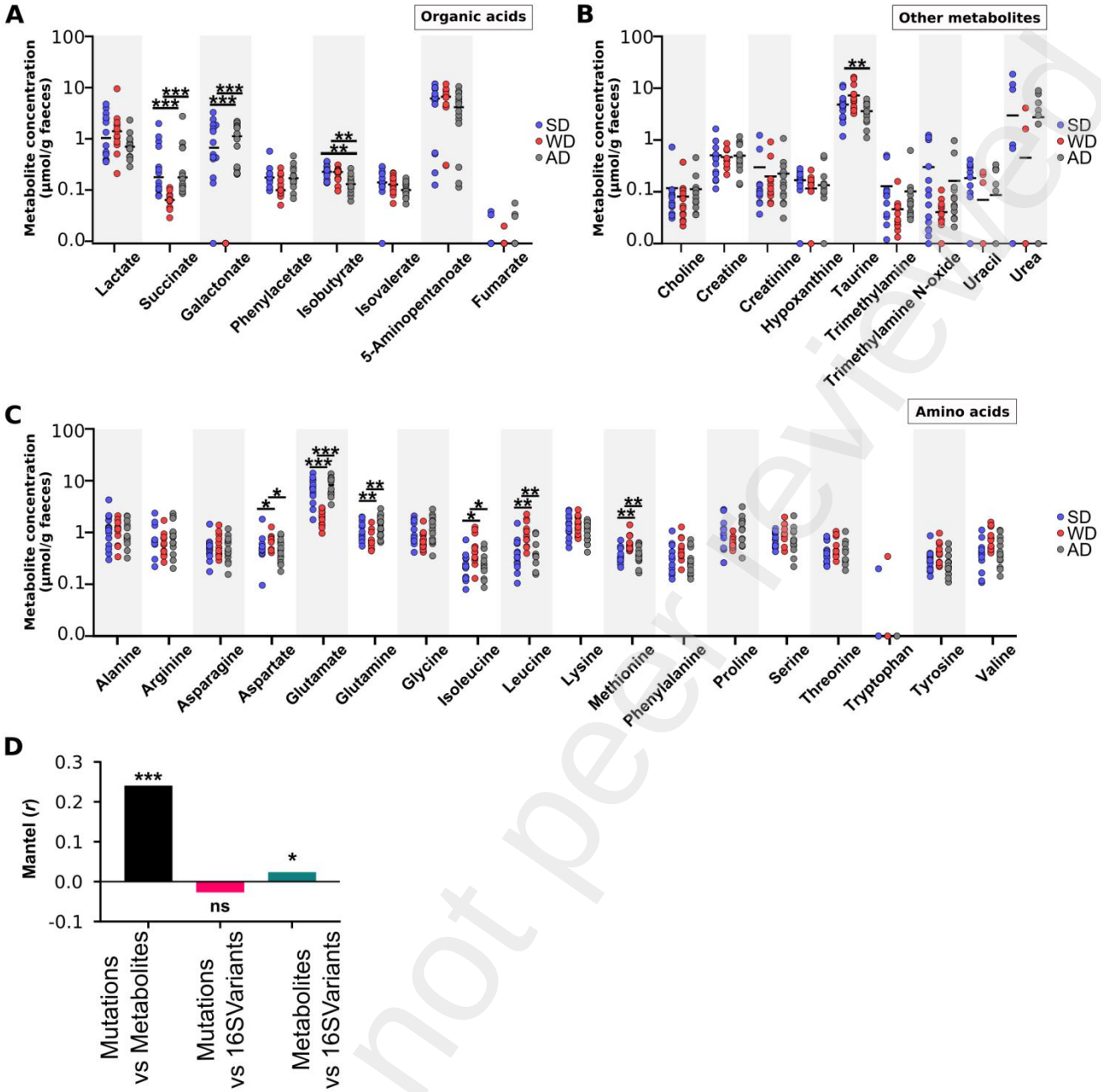

Figure S5

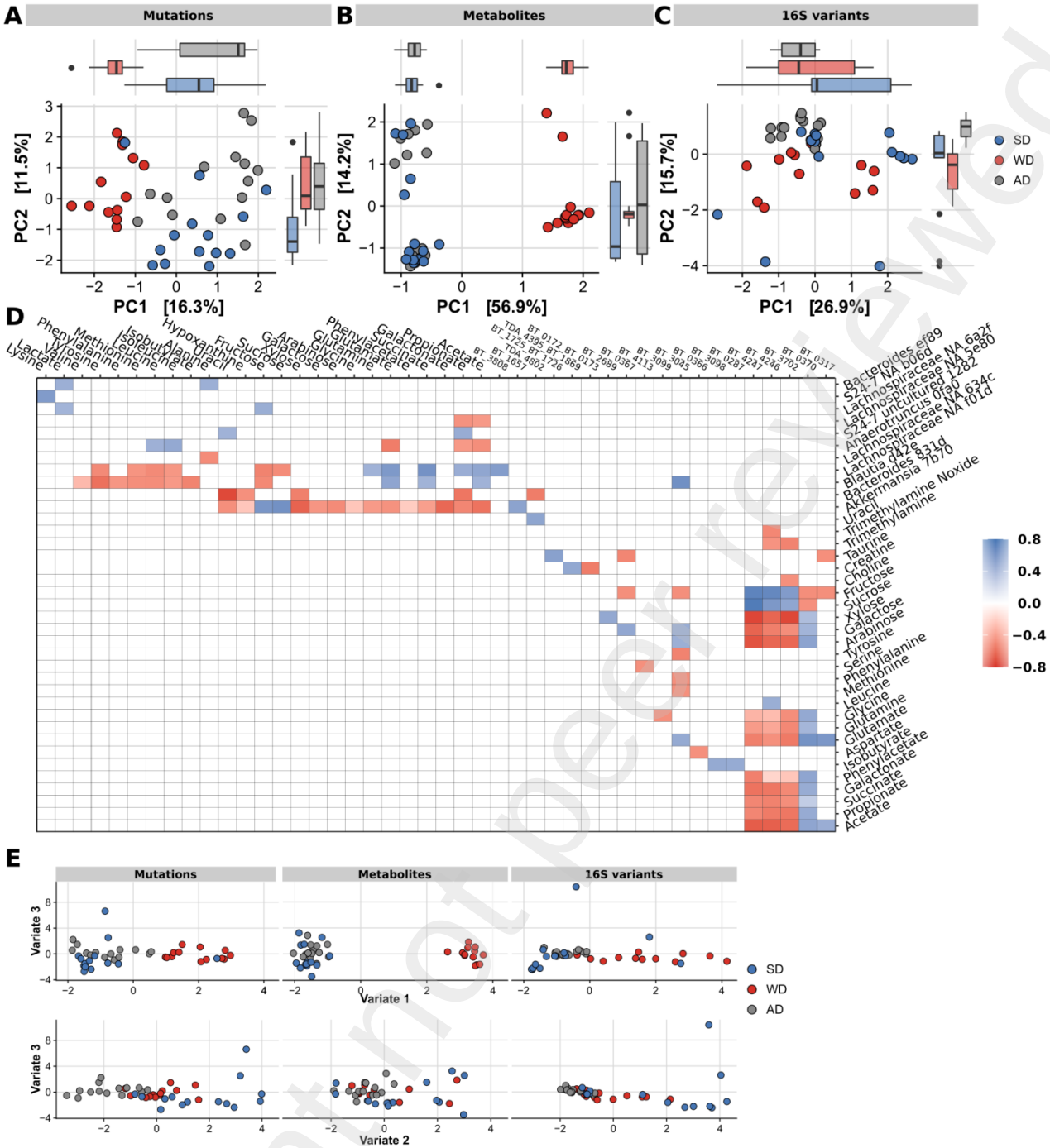

Figure S6

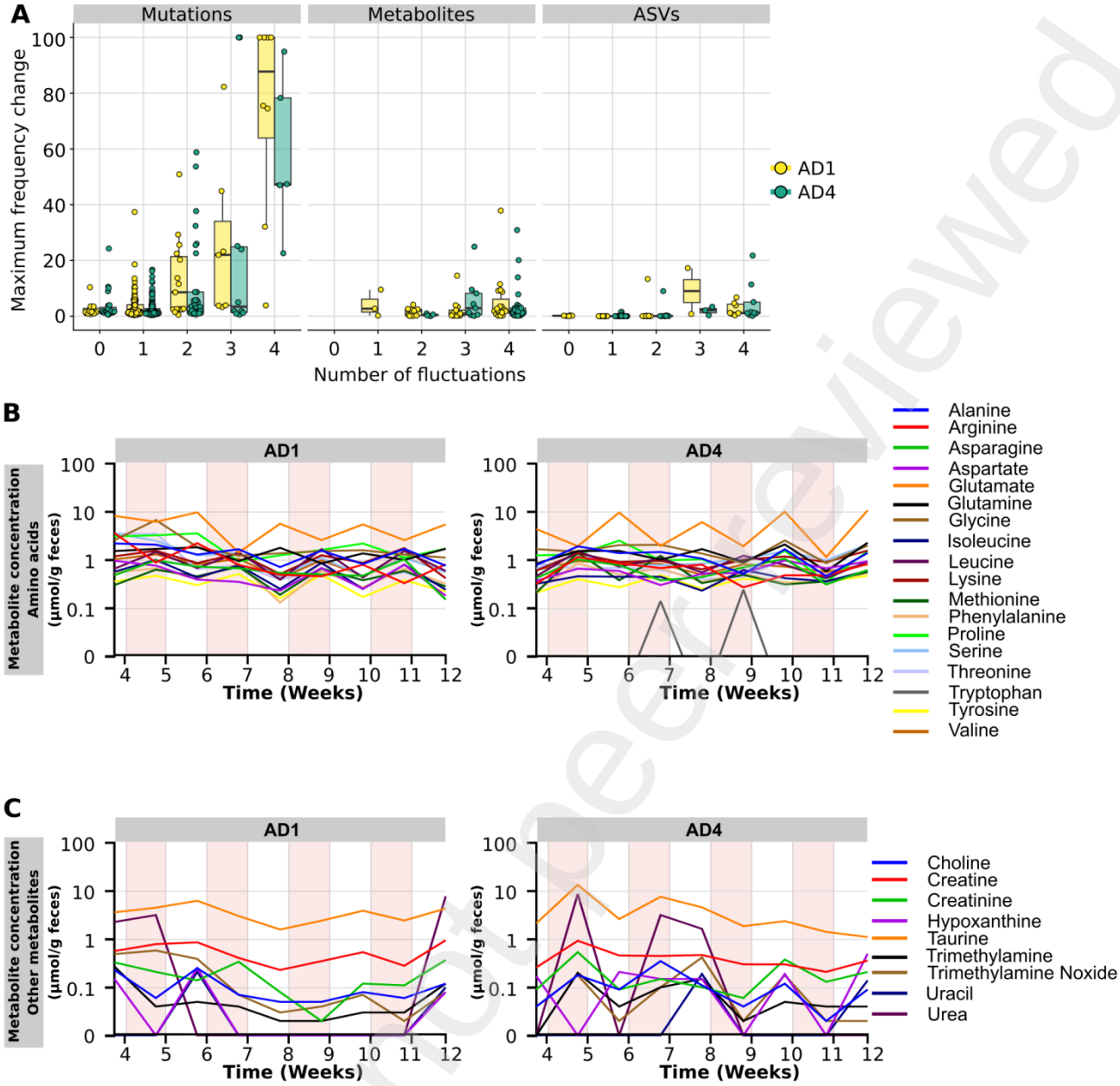

Figure S7