Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut

Kern Rei Chng[1]*, Tarini Shankar Ghosh[1,2]*, Yi Han Tan[3]*, Tannistha Nandi[1]*, Ivor Russel Lee[3], Amanda Hui Qi Ng[1], Chenhao Li[1], Aarthi Ravikrishnan[1], Kar Mun Lim[1], David Lye[4,5], Timothy Barkham[6], Karthik Raman[7], Swaine Chen[1], Louis Chai[3,8], Barnaby Young[4,5#], Yunn-Hwen Gan[3#], Niranjan Nagarajan[1,3#]

[1]Genome Institute of Singapore, Singapore 138672, Singapore

[2]APC Microbiome Ireland, University College Cork, Ireland, T12K8AF

[3]Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117596, Singapore

[4]National Centre for Infectious Disease, Tan Tock Seng Hospital, Singapore 308433, Singapore

[5]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore

[6]Department of Laboratory Medicine, Tan Tock Seng Hospital, Singapore 308433, Singapore

[7]Department of Biotechnology, Indian Institute of Technology (Madras), Chennai 600036, India

[8]Division of Infectious Diseases, University Medicine Cluster, National University Health System, Singapore 119228, Singapore

**Joint First Authors*

[#]Corresponding Authors

Lead Contact: nagarajann@gis.a-star.edu.sg

Abstract

Loss of diversity in gut microbiome can persist for extended periods after antibiotic treatment, impacting microbiome function, antimicrobial resistance and likelyhost health. Despite widespread antibiotic use, understanding of species and metabolic functions contributing to gut microbiome recovery is limited. Using data from 4 different discovery cohorts in 3 continents comprising >500 microbiome profiles from 117 subjects, we identified 21 bacterial species exhibiting robust association with ecological recovery post antibiotic therapy. Functional and growth-rate analysis showed that recovery is supported by enrichment in specific carbohydrate degradation and energy production pathways. Association rule mining on 782 microbiome profiles from MEDUSA database enabled reconstruction of the gut microbial 'food-web', identifying many recovery-associated bacteria (RABs) as keystone species, with ability to use host and diet-derived energy sources, and support the repopulation of other gut species. Experiments in a mouse model recapitulated the ability of RABs (*Bacteroides thetaiotamicron* and *Bifidobacterium adolescentis*) to promote recovery with synergistic effects, providing a two orders of magnitude boost to microbial abundance in early time-points and faster maturation of microbial diversity. Identification of specific species and metabolic functions promoting recovery opens up opportunities for rationally determining pre- and probiotic formulations that offer protection from long-term consequences of frequent antibiotic usage.

Introduction

The human gut microbiome harbors trillions of bacteria providing diverse metabolic capabilities and with essential roles in host health, particularly energy metabolism, immune homeostasis, and xenobiotic metabolism[1]. A stable consortium of commensal microbiota is also believed to play a key role in resisting colonization by pathogens, with reduced diversity being associated with increased risk for infections[2,3]. Several recent studies have further highlighted the importance of the gut microbiome for host health, particularly in infants and the elderly, with alterations and loss of diversity being associated with various metabolic, immunological and neurological diseases[4], and poorer response to cancer immunotherapy[5,6].

Among the many factors that are known to perturb the gut microbiome, antibiotics are the major cause of profound and long-term alterations[7-9]. Antibiotics are widely used in farming and healthcare, and global consumption is estimated to have increased by 65% from 2000 to 2015[10]. While the impact of antibiotics on host health through microbiome disruption is likely to be significant, it has not been fully quantified to date[11,12]. Antibiotic associated diarrhea and *Clostridium difficile* colitis are common early complications of microbiome disruption[13], while antibiotics also select for drug resistance genes and organisms, thus creating a reservoir for transmission of resistance cassettes[14,15]. In the medium to long term, recovery of the microbial community can be slow and variable[7-9], and is conditioned on the initial state[16]. Epidemiological and model organism studies suggest that long-term consequences of antibiotic usage include immunological diseases in children[17], metabolic diseases in adults[18], and an increased risk of infections[19].

Despite mounting evidence on the importance of gut microbiome function and how antibiotic usage can severely impact it, our understanding of the post-antibiotic recovery process is limited. Several studies have noted that high initial diversity in the gut microbiome may be associated with better recovery from antibiotic-induced perturbations[16,20-22]. In addition, the carriage of specific antibiotic resistance genes has been linked with the recovery process in some studies[22,23]. While it is expected that bacteria that are resistant to the antibiotic used will have an advantage in seeding the

repopulation of the gut, it is unclear if antibiotic resistance alone is sufficient or necessary to recover the ecological and functional richness of the gut microbiome. In particular, we do not know which specific groups of microbial taxa, and the functions they perform, accelerate or impede the process and explain the substantial variability in speed and extent of recovery that is seen across individuals[7-9,16]. For example, while commonly used probiotics can be generally beneficial to host health, their utility after antibiotic treatment remains unclear, with a recent study providing evidence that they may in fact delay microbiome recovery[24].

The interactions between species play a key role in the recovery of many ecosystems after severe perturbations[25,26]. Typically, reseeding by a few keystone species is essential to trigger a chain of food-web interactions that eventually lead to recovery of the overall ecosystem. Several important constituents of the healthy gut microbiome have been identified (e.g. *Bacteroides* species[27,28]) and correlations in their abundance have been used to postulate cross-feeding interactions[27,29,30]. However, the role of these species and their interactions in the context of post-antibiotic microbiome recovery has not been explored[31].

In this study, we employed a metagenome-wide association approach[32] to identify microbial species and functions that could contribute to robust recovery of the microbiome after antibiotic usage. We then show how *in vivo* human metagenomic data from multiple cohorts supports a mechanistic model where gut microbiome recovery is facilitated by carbohydrate degradation and microbial cross-feeding triggered by a subset of the identified species. Validation experiments in a mouse model demonstrate how recovery-associated bacterial species (RABs) can synergistically provide a >100-fold boost to absolute microbial abundance and higher diversity in the gut microbiome after antibiotic treatment. Systematic investigations using higher-order combinations of RABs can thus help us understand the interactions between them that likely contribute to the complex ecological processes underlying gut microbiome recovery.

RESULTS

Robust identification of microbial taxa associated with gut microbiome recovery

In order to identify microbial markers associated with gut microbiome recovery, we assembled and systematically analyzed longitudinal data from 4 cohorts (a total of 117 individuals with >500 samples; **Methods**). These cohorts represent individuals from 4 countries on 3 continents (Singapore, Canada[16], England[8], Sweden[8]), a range of age groups (21-81) and using different classes of antibiotics, allowing us to infer common factors associated with microbiome recovery (**Table 1**). Data from the Singaporean cohort was newly generated and analyzed (deep shotgun metagenomic sequencing of 74 samples; >80 million reads on average), involving mostly elderly subjects receiving inpatient antibiotic treatment (**Suppl. Data File 1**). Each cohort was analyzed independently to account for cohort-specific biases, and the results were aggregated using a cross-cohort validation approach to only identify microbial taxa that were independently associated with recovery in at least 2 cohorts (**Methods**).

To stratify individuals based on their recovery status, we noted that many individuals exhibited a U-shaped profile for gut microbial diversity, with a significant drop in diversity during antibiotic treatment, but with recovery of diversity in post-treatment timepoints ('recoverers', **Fig. 1a**). A subset of individuals, however, continued to have low gut microbial diversity even 3 months post antibiotics ('non-recoverers', **Fig. 1a**), contrasting with those at the other end of the diversity spectrum (**Suppl. Fig. S1a**). We therefore stratified subjects based on post-antibiotic microbial diversity as a readily defined reference-free metric for recovery across cohorts[23] (**Methods**). This metric correlated well with alternative definitions, for e.g. as expected, post-antibiotic microbiomes for recoverers were much more similar to control microbiomes overall, compared to non-recoverers (one-sided Wilcoxon p-value<0.001; **Fig. 1b, c**). This pattern was seen to be consistent across cohorts and using different diversity metrics (**Suppl. Fig. S1b**). Recoverers and non-recoverers also did not have significant differences in microbial diversity in the pre-antibiotic state (Wilcoxon p-value>0.05).

To determine microbial taxa with a role in microbiome recovery, a two-stage approach and cross-cohort validation strategy was used to increase sensitivity and

specificity of the association analysis across all timepoints (**Methods**; **Suppl. Data File 2**; 34 bacterial species in stage 1). In total, 21 microbial species were identified to be significantly associated with microbiome recovery in at least 2 cohorts (Recovery Associated Bacteria – RAB; **Table 2**), with 9 species identified in 3 cohorts and 1 in all 4 cohorts (*Bacteroides uniformis*; **Fig. 1d**, using data for all timepoints). Variability across cohorts may reflect differences in diet[33], environment[34] and antibiotics used, while genus-level consistencies (e.g. *Bacteroides* species; **Fig. 1d**; **Table 2**) may reflect functional redundancies in associated species. While some RABs are common gut bacteria (e.g. *Alistipes putredinis)*, are known to have host-beneficial functions (e.g. *Faecalibacterium prausnitzii*[35]) and have been observed to be depleted in disease states (e.g. *B. uniformis*[36]), others are more variably distributed, with limited understanding of their function in the gut microbiome, and their role in gut microbiome recovery after antibiotic treatment being unknown (**Table 2**). The distribution of most RABs across recoverers and non-recoverers suggests that their abundance, rather than their presence or absence, likely contributes to the recovery process. In addition, as no RAB segregates recoverers and non-recoverers on its own in any cohort, the combined influence of multiple RABs likely determines successful microbiome recovery.

RABs were initially identified across treatment stages (pre-, during and post-antibiotics; **Methods**) to capture species that may contribute to recovery at any stage. We then investigated abundance patterns of RABs across stages and noted that while some were 2-4× more abundant in recoverers before treatment (e.g. *B. uniformis*), others were enriched in later timepoints, indicating that they may play a secondary or synergistic role in recovery (**Suppl. Fig. S2**; e.g. *F. prausnitzii*), and that combinatorial effects across treatment stages may play a role in recovery. Interestingly, no RABs were depleted in the gut microbiomes of recoverers versus non-recoverers, indicating that they do not have specific inhibitory roles. Training of machine learning models across cohorts showed that post-antibiotic recovery status can be predicted to an extent using pre-antibiotic taxonomic abundances for an individual (70.4% accuracy; **Suppl. Note 1**).

We enlisted a fifth cohort of healthy young adults in Singapore taking antibiotics (NUH, **Table 1**), whose metagenomes were not sequenced at the point of initial association analysis with the original four cohorts, to study the consistency of RABs across cohorts. Overall, 12 out of 21 RAB species were significantly associated (one-sided Wilcoxon *p*-value <0.1) in the new cohort as well, similar to the overlap of the four original cohorts with RAB species (6-14 species, **Table 2**), confirming the robustness of associations despite differences in age, location and antibiotics used. In addition, incorporation of the fifth cohort in the cross-cohort association analysis only increase the list of RABs by 2, highlighting the consistency and reproducibility of this list.

Enrichment in carbohydrate degradation and energy metabolism pathways links RABs with microbial community growth and recovery

To study microbial functions that link RABs to microbiome recovery, we systematically identified all differentially abundant gene families and pathways in the pre- and during treatment metagenomes of recoverers and non-recoverers (CA and SG cohorts, **Methods**; FDR adjusted *p*-value<0.1 and LDA score >1.25; **Suppl. Data File 3**). This analysis highlighted a core set of growth-associated pathways pertaining to the biosynthesis of amino acids, nucleotides, co-factors and cell wall constituents (**Suppl. Fig. S3**). In addition, pathways involved in carbohydrate degradation and energy production were also significantly over-represented in the gut microbiomes of recoverers. Analysis of inferred pathway abundances from 16S rRNA profiles in the pre- and during treatment stages of the English and Swedish cohorts further confirmed these associations (carbohydrate and butanoate metabolism, Wilcoxon test *p*-value<0.05; **Suppl. Fig. S4**; **Suppl. Data File 3**). In comparison, analysis of resistomes of recoverers and non-recoverers in the pre- and during treatment stages did not show any significant enrichment for RAB species indicating that antibiotic resistance functions do not, in general, explain the taxonomic differences observed (**Methods**; **Suppl. Fig. S5**).

To further understand the role of carbohydrate processing functions in microbiome recovery, carbohydrate-active enzyme families were annotated in RABs and the gut metagenomes of recoverers and non-recoverers (based on CAZyme families, **Methods**). Overall, RABs exhibited a significant enrichment for CAZyme families

compared to non-RABs (two-sided Wilcoxon test *p*-value<0.001; **Fig. 2a**), though this does not seem to be a necessary or sufficient condition for identification as a RAB[37]. The enrichment of CAZyme families in RABs was also reflected at the community level where the metagenomes of recoverers at all timepoints were enriched in CAZyme families compared to non-recoverers (one-sided Wilcoxon test *p*-value<0.001 and <0.05 for CA and SG respectively; **Fig. 2b**; **Suppl. Data File 4**), consistent with enriched pathways in **Suppl. Fig. S3**.

Linking the two major classes of pathways enriched in recoverers versus non-recoverers, we hypothesized that in broad terms, higher carbohydrate metabolism capabilities in RABs could enable better nutritional harvest, thus enhancing biosynthesis and microbial growth (**Suppl. Fig. S3**), and subsequent recovery of gut microbial diversity and biomass (**Fig. 2**). Using *in silico* estimates of community growth rates (from DNA coverage skews in replicating cells) from metagenomic data[38] (**Suppl. Data File 5**), we observed that recoverers exhibited higher microbial community growth rate overall than non-recoverers across all stages of antibiotic treatment (one-sided Wilcoxon test *p*-value<0.001 and <0.05 for CA and SG cohorts, respectively; **Fig. 2c**). Additionally, we noted that the pre- and during treatment abundance of RABs had a significantly higher correlation with post-treatment community growth rate across individuals (one-sided Wilcoxon test *p*-value<0.001 for CA cohort; **Fig. 2d**). Finally, in both the CA and SG cohorts, community growth rate at all timepoints was positively correlated with the number of CAZyme families (for CA, r=0.729; for SG, r=0.556; *p*-value<0.001; **Fig. 2e**). Taken together, these analyses consistently link together enrichment in RABs, carbohydrate degradation potential, microbial community growth rate and microbiome recovery as successive steps in a plausible mechanism for how RABs promote recovery.

Specific carbohydrate degradation functions define the role of RABs in the gut microbial food-web

Carbohydrate active enzymes can be varied in their function and their differential and combinatorial usage by RABs could contribute to microbiome recovery. To study this, we clustered a set of 137 bacterial genomes annotated for their CAZyme repertoire[37]

based on their genome-wide profiles of substrate-specific enzyme copy numbers to obtain 5 distinct clusters (**Suppl. Fig. S6**). Interestingly, RABs were primarily observed to aggregate in 2 out of the 5 clusters, with significant enrichment in cluster 1 containing genomes abundant in host (mucins) as well as diet-derived (plant and animal) carbohydrate degrading enzymes (Fisher's exact test p-value<0.001). The ability to degrade mucins is key for bacterial colonization of the intestine[39], and may assist some RABs in seeding the recovery process. While a few RABs fall in cluster 2 that is characterized by diet-derived (plant and animal) carbohydrate degrading enzymes, clusters 3, 4 and 5 (Starch, Fungal carbohydrate and Peptidoglycan degradation, respectively) were sparsely represented, highlighting the importance of specific carbohydrate degradation processes in microbiome recovery.

The recovery of many natural ecosystems is driven by ecological interactions[25,26] and we hypothesized that a similar 'food-web' of cross-feeding between RABs and other constituents of the gut microbiome is important for microbiome recovery. As experimental information about the gut microbial food-web is sparse, we developed a data-driven approach based on association rule mining (782 microbiome profiles from the MEDUSA database[40]; **Methods**) to identify dependency relationships between bacteria in the gut microbiome (A $\rightarrow$ B), where the presence of species B appears conditional on the presence of species A (but not *vice versa*). The resulting network contains 1,166 directed edges linking 266 bacterial species, identified directly from gut microbiome data (**Suppl. Data File 6**), and recapitulating several known cross-feeding interactions. (e.g. *Bacteroides* species and group *C. coccoides* species[41]).

We noted in the bacterial food-web that a few species mostly have outgoing edges, indicating that they are essential for the presence of other species, while many species have mostly incoming edges highlighting their dependence on the presence of many other species. Based on this, we visualized the network by sorting species based on the difference in outgoing to incoming edges (bottom to top), revealing a pyramidal web structure (with RAB nodes highlighted, **Fig. 3a**). Interestingly, many RABs belonging to cluster 1, and correspondingly enriched in mucin degrading enzymes, were clustered in the bottom third of this network (denoted as primary species). No RABs were found in

the middle third of the network (secondary species), while RABs in the top third of the network belong to a diverse set of CAZyme clusters (tertiary species). These observations are in agreement with the ecological expectation that while some RABs should be keystone species that are essential to triggering the repopulation effect (primary species), others play a synergistic role in later stages or serve as indicator species for ecological recovery (tertiary species).

Overall, the carbohydrate degradation profiles of RABs and their organization in the food-web is consistent with a model (**Fig. 3b**) where: (i) primary RABs employ their mucin degrading capabilities to successfully colonize/recolonize the gut epithelium[39,42,43]; some of the primary RABs also serve as specialists in breaking down complex diet-derived carbohydrates[42] (e.g. *B. uniformis*), (ii) this helps initiate a chain of cross-feeding interactions that support the repopulation of other bacteria (secondary or tertiary species) that cannot degrade mucins and/or are dependent on the breakdown of complex carbohydrates into simple sugars[44], (iii) as the microbial community repopulates, some RABs (e.g. *F. prausnitizii* and *Roseburia* species) contribute to production of SCFAs that in turn provide energy for colonocytes[45,46], and (iv) the resulting increased production of mucin creates a positive feedback loop that drives faster recovery of microbial biomass[47,48]. The overall effect is the rebuilding of a food-web in the gut microbial ecosystem to support a diverse community concurrently and is distinct from the microbial succession processes that have been described in other contexts[49].

A mouse model of microbiome recovery recapitulates synergy between primary and tertiary RABs *in vivo*

To study synergistic interactions between RABs, genome scale metabolic models were used to evaluate the benefit of co-culture for various species (**Methods**). Overall, RABs were observed to derive greater metabolic support from each other than from other non-RAB species (Wilcoxon p-value<0.001). In particular, tertiary RABs such as *B. adolescentis*, *Ruminococcus bromii* and *Alistipes shahii* could derive metabolic benefits from several other species, including the primary RAB *B. thetaiotamicron* (**Suppl. Fig. S7**, **Suppl. Data File 7**). For investigating potential synergies *in vivo* and cause-effect

relationships, we used a physiologically relevant mouse model of microbiome recovery after antibiotic treatment. Specifically, conventional healthy mice (C57BL/6J, normal gut development, mucin production) were given antibiotics for 5 days before being randomly allocated to four different groups to study treatment effects in a case-control setting: oral gavage with (a) the primary RAB species *B. thetaiotamicron* (Bt), (b) tertiary RAB species *B. adolescentis* (Ba), (c) combination of *B. thetaiotamicron* and *B. adolescentis* (Bt+Ba), and (d) PBS media (Vehicle; **Methods**). Recovery was then monitored over a period of 22 days by collecting stool samples every three days and analyzing the microbiome with shotgun metagenomic sequencing (9 timepoints and 2-6 cages per group with 2 mice per cage, **Methods**; **Fig. 4a**).

As expected, all treatment groups exhibited a >3-log reduction in microbial biomass after antibiotic treatment (**Methods**; **Fig. 4b**). Starting from 1 day after gavage (day 7), and more noticeably at 4 days after gavage (day 10), the Bt and Bt+Ba groups exhibited significantly enhanced biomass recovery (>100×; excluding gavaged species) compared to the PBS and Ba groups (**Fig. 4b**; **Suppl. Fig. S8a**; qPCR verification in **Methods**). While the Bt and Bt+Ba groups converge to their microbial biomass at pre-antibiotic levels by day 10, the PBS and Ba groups continued to have lower biomass than pre-antibiotic levels at day 22. Enhanced recovery was also associated with successful colonization, confirmed based on comparisons with metagenomic data from a control gavage (*Bacillus spp*, **Suppl. Fig. S9**). Interestingly, the Bt+Ba group was distinct from other treatment groups in recovering higher microbiome diversity at day 19 and 22 (**Fig. 4c**). This was also accompanied by reconstruction of a community that was more similar to the pre-antibiotic microbiome at day 22 in the Bt+Ba group (**Suppl. Fig. S8b**). These results highlight that while Bt gavage and colonization was sufficient for biomass recovery and Ba gavage alone was not, the combination of Bt and Ba promotes biomass and diversity recovery in a synergistic fashion. As observed in the human cohorts, an enrichment of mucin as well as dietary carbohydrate degradation pathways (but not peptidoglycan degradation, as control) was associated with the recovery process in the Bt and Bt+Ba groups (**Fig. 4d**, **e**, **f**).

Discussion

Cross-cohort analysis is a powerful way to account for confounding effects within individual studies, enabling the identification of consistent associations with microbiome recovery despite variations in cohort characteristics such as antibiotics used and patient demographics. The bacterial species and functions identified in this study provide a data-driven view of how shared microbial factors contribute to gut microbiome recovery in diverse human cohorts around the world, highlighting the value of data-sharing and re-analysis. Our findings emphasize the central role of enabling energy harvest from diet, and the ability to colonize the host by degrading mucins in the keystone species that underpin ecological recovery (primary RABs), connecting recovery of key microbiome functions to ecological recovery of biomass and diversity. Additional factors such as antibiotic resistance likely contribute to this process in a time and context-dependent manner. As environmental factors strongly influence the gut microbiome[34], the specific keystone species that are important for an individual could further vary with host and dietary factors. The analytical approaches used here could uncover these in larger cohorts, helping to train antibiotic and environment-specific machine learning models to predict microbiome recovery. We anticipate that such models would have clinical utility, especially for at-risk elderly or cancer patients, to guide targeted intervention strategies mitigating the impact of antibiotics on the gut microbiome.

Consistent with the emerging understanding of how diet modulates the gut microbiome[33,34], an additional perspective that emerges from this study is the potential to promote RABs and microbiome recovery via prebiotic effects, especially since few RABs are available as probiotics. Many of the identified RABs are specialist carbohydrate fermenters (e.g. pectin) and a high fiber/low fat diet could aid in selecting and expanding them. For example, in a study on how gut microbiota differ in twins discordant for obesity, Ridaura *et al* identified 3 RABs (*B. uniformis*, *B. thetaiotaomicron* and *A. putredinis*) as being transplantable features of a “lean microbiome”, but transplantation was dependent on a high fiber diet[50]. Similarly, pectin supplementation can promote species from the Bacteroidetes phylum with associated improvement in gut barrier function[51], as well as more stable fecal microbiota transplantation[52]. Finally,

different oligosaccharides can promote the growth of several butyrate producing RABs[53,54] (**Fig. 3b**), serving as an avenue to contribute to microbiome recovery by reducing host inflammation and increasing mucin production[48].

In general, ecological theory has suggested that ecosystem recovery is a complex, multi-step process that is determined by interactions between many species[25,26]. Our observations in the human gut microbiome are in agreement with this model, with the identification of multiple recovery-associated species, the potential for synergistic interactions and microbial cross-feeding, and a conceptual model for how this promotes ecological recovery in the gut. Results from our mouse experiments demonstrate that individual RABs likely have distinct functions, but can work in a synergistic fashion to recover microbial biomass and diversity. As these observations were made in conventional mice with normal physiology (versus germ-free mice), and in a case-control setting where single species gavages (Bt and Ba groups) serve as ideal controls for the combination (Bt+Ba), they highlight the robust role that microbial functions play in the recovery process across species. While investigating all RAB combinations *in vivo* might be infeasible, systematic investigation of the top predicted metabolic interactions between RABs (e.g. between *F. prausnitzii* and *A. shahii*) through *in vitro* co-cultures[55] could be the next step to unravel the combinatorial interactions among RABs driving microbiome recovery *in vivo*. Metabolic modeling could, in particular, help further explore the contributions of different carbohydrate degradation genes and processes to microbiome recovery[56], especially for many anaerobic bacteria that are hard to culture or genetically modify[57]. Further clinical studies incorporating detailed dietary information or with a controlled diet are also needed to evaluate the role of diet and its interaction with RABs and CAZymes in microbiome recovery.

The microbial 'food-web' in this study as determined by data-mining techniques is conceptually a valuable resource for organizing our understanding of how microbes interact and assemble in the human gut. By using a large database of human gut microbiome profiles, we were able to determine microbial assemblages that are feasible and the dependency relationships that they suggest. These can then help interpret longitudinal studies of recovery and infer the interactions between species that play a

role. While our current work highlights that introduction of primary species such as *B. thetaiotamicron* is necessary for biomass recovery, in comparison to common probiotics such as *B. adolescentis*, synergistic combinations can be more beneficial for robust recovery of a diverse gut microbial ecosystem. Similar interactions could also play a critical role in recovery from other microbiome perturbations, and thus a broader understanding of the microbial food-web could set the stage for rational design of pre- and probiotic formulations that promote functional and ecological resilience in gut microbiota.

METHODS

Study Populations

(a) Singapore: The Singaporean cohort ('SG'; manuscript in preparation) is a natural history cohort consisting of individuals admitted to Tan Tock Seng Hospital (TTSH) in Singapore and prescribed antibiotics for 1-2 weeks (primarily Co-amoxiclav and Clarithromycin; **Table 1**). Stool samples were collected as soon as possible after admission (pre-/early: <3 days into treatment), during and up to 3 months after antibiotic usage. The study was approved by the Institutional Review Board at TTSH (DSRB 2013/00769).

(b) Canada: Shotgun metagenomic datasets for a Canadian cohort[16] ('CA') were obtained from the European Nucleotide Archive database (Study Accession Number: PRJEB8094; **Table 1**). The study analyzed fecal samples from healthy individuals who were administered antibiotics (Cefprozil; three timepoints: pre-antibiotic day 0, during treatment day 7 and post treatment day 90).

(c) England and Sweden: 16S rRNA sequencing datasets for an English and a Swedish cohort[8] ('EN', 'SW') were obtained from the NCBI short read archive (Project ID: SRP057504; **Table 1**). In both cohorts, healthy volunteers were given antibiotics (EN: Amoxicillin, SW: Clindamycin/Ciprofloxacin) and fecal samples analyzed for day 0 (pre-antibiotic), day 7 (during treatment) and for one and two month follow-ups (post treatment).

(d) National university hospital (*NUH)*: A prospective cohort of young Chinese adults was recruited to study the impact of antibiotics on the gut microbiome at the National University Hospital (NUH; 5-day course of Co-amoxiclav; manuscript in preparation). Stool samples were collected before (day 0), during (day 1-5) and after antibiotic cessation (day 8 and day 28). The study was approved by the Institutional Review Board at NUH (DSRB 2012/00776).

For the CA, EN and SW cohorts, all antibiotic treated subjects with data from the 3 treatment stages were further analyzed to identify recovery associated bacterial taxa and functions.

DNA extraction and sequencing for SG and NUH cohorts

Extraction of DNA from stool samples was carried out using PowerSoil DNA Isolation Kit (MoBio Laboratories, California, USA) with minor modifications to the manufacturer's protocol (volume of solutions C2, C3 and C4 were doubled and centrifugation time was extended to twice the original duration). Purified DNA was eluted in 80µl of Solution C6. DNA libraries were prepared by using 20ng of extracted DNA re-suspended in a volume of 50µl and subjected to shearing using Adaptive Focused Acoustics™ (Covaris, Massachusetts, USA) with the following parameters; Duty Factor: 30%, Peak Incident Power (PIP): 450, 200 cycles per burst, Treatment Time: 240s. Sheared DNA was cleaned up with 1.5× Agencourt AMPure XP beads (A63882, Beckman Coulter, California, USA). End-repair, A-addition and adapter ligation was carried out using the Gene Read DNA Library I Core Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Custom barcode adapters (**Suppl. Table 1**) were used in place of GeneRead Adapter I Set for adapter ligation. DNA libraries were cleaned up twice using 1.5× Agencourt AMPure XP beads (A63882, Beckman Coulter, California, USA) before enrichment of libraries using the protocol adapted from Multiplexing Sample Preparation Oligonucleotide kit (Illumina, California, USA). Enrichment PCR was carried out with PE 1.0 and custom index-primers (**Suppl. Table 1**) for 14 cycles. Libraries were quantified using Agilent Bioanalyzer and prepared with Agilent DNA1000 Kit (Agilent Technologies, California, USA), pooled in equimolar concentrations. Sequencing of the samples was performed using the Illumina HiSeq 2500 (Illumina, California, USA) sequencing instrument to generate >80 million 2×101 bp reads on average.

Taxonomic and functional profiling for all cohorts

For metagenomic sequencing datasets (CA, SG and NUH cohorts) raw reads were quality filtered and trimmed using default options in famas (https://github.com/andreas-wilm/famas). Reads that are potentially from human DNA were removed by mapping to the hg19 reference using BWA-MEM[58] (default parameters; coverage >80% of read). The remaining reads were used for taxonomic profiling using MetaPhlAn with default parameters[59,60] (**Suppl. Data File 1**). Functional profiles for the metagenomes were obtained using the HUMAnN2[61] program (**Suppl. Data File 3**).

For the 16S rRNA sequencing datasets (EN and SW cohorts) taxonomic classification was done by mapping reads to the SILVA database[62] (v123) using blastn. For each read, the species corresponding to the best hit (with identity > 97% and query coverage > 95%) was obtained and was taken as the source species of the read. In the case of multiple hits, the source taxon was computed as the Lowest Common Ancestor of the hit species. Reads assigned to each taxon were aggregated to obtain a relative abundance profile for each sample (**Suppl. Data File 1**). PICRUSt[63] was used to infer KEGG pathway abundances from the corresponding taxonomic profiles (**Suppl. Data File 3**).

Identification of recovery associated bacterial taxa and functions

Individuals were classified as 'recoverers' and 'non-recoverers' in each cohort to enable cohort-specific association analysis and identification of recovery associated bacterial taxa and functions. As post-antibiotic microbiomes may not necessarily resemble the pre-antibiotic state for an individual (e.g. due to enterotype switching[64]), we used the post-treatment gut microbial diversity (species-level; Simpson) to define recoverers and stratify subjects into balanced groups (median threshold). Samples within a 10% window of the interquartile range from the median were marked as having indeterminate status and excluded from further analysis. A two-stage approach was used to combine results from all cohorts to sensitively identify recovery associated taxa and a cross-cohort validation strategy was used to identify taxa that are significant in at least 2 out of 4 cohorts. In stage 1, a non-parametric test was used within each cohort to identify candidate taxa (one-sided Wilcoxon test). The resulting *p*-values were merged across cohorts to compute a combined *p*-value using Fisher's method and filtered with a FDR adjusted *p*-value threshold of 0.01 (Benjamini-Hochberg method). Next, in stage 2, cohort-specific FDR adjusted *p*-values (Benjamini-Hochberg method) were re-computed for this subset of taxa and only taxa with consistent (in terms of direction of change) significant associations (FDR<0.05) in at least 2 cohorts were retained. This analysis was done within each treatment stage (pre-, during and post- antibiotics) as well as jointly to increase sensitivity in identifying recovery associated taxa regardless of treatment stage.

Functional profiles computed with HUMAnN2 were compared between recoverers and non-recoverers in the SG and CA cohorts using the linear discriminant analysis approach in LEfSe[65] (version 1.1.0) to identify differentially abundant pathways.

Microbial community growth rate analysis

An *in silico* approach, originally proposed by Korem et al[38], was used to compute the skew of DNA copy number starting from around the origin of replication to the termination region (peak-to-trough ration or PTR), as an estimate of growth rates for individual species in the microbiome from shotgun metagenomic data (PTRC1.1: https://genie.weizmann.ac.il/software/bac_growth.html, default parameters). The community growth rate (CGR) for each sample was then computed from the common species in the community (PTR values in >50% of samples) as the median PTR value (PTR set to lower-bound of 1 when not available; **Suppl. Data File 5**).

Profiling of carbohydrate active enzymes (CAZymes)

An in-house nucleotide gene database for CAZymes was created by downloading sequences from NCBI corresponding to Accession IDs for different CAZyme families annotated in dbCAN[66] (http://csbl.bmb.uga.edu/dbCAN/). Metagenomic reads were mapped to this database for each sample with BWA-MEM[58] (default parameters) to compute the fraction of reads mapping to the CAZyme gene per kbp per million reads in the metagenome (RPKM). Results were aggregated for each CAZyme family based on values for individual CAZyme genes belonging to a family.

Analysis of antibiotic resistance genes within gut microbiomes

Resistome profiling within a microbiome was performed similarly by mapping metagenomic reads using BWA-MEM (default parameters) to the ARG-ANNOT database[67], and calculating the fraction of reads mapping to a resistance gene per kbp per million reads of the metagenome (RPKM). Kraken[68] was used with default parameters to obtain the taxonomic classification of reads and thus obtain the relative representation of different taxonomic groups within the resistome.

Clustering of species based on their carbohydrate degradation profiles

The substrate-specificities of different Glycoside hydrolase (GH) and Polysaccharide lyase (PL) families were obtained from previous studies[43,69]. These included substrates such as plant cell wall carbohydrates, animal carbohydrates, peptidoglycans, fungal carbohydrates, sucrose/fructose, dextran, starch/glycogen and mucins. Copy number annotations for each GH and PL family in 137 bacterial species were obtained from a previous genome scale analysis of CAZymes in species belonging to the human gut microbiome[37]. Copy numbers of GH/PL genes within each of the 8 substrate specificities were aggregated and normalized to obtain an overall carbohydrate degradation profile for each bacterial species. Degradation profiles were then clustered using hierarchical clustering ('hclust' function in R with Euclidean distance and complete linkage clustering) to group species based on their enzyme repertoire for different categories of carbohydrates. Association of the identified recovery associated bacteria to one or more of these clusters was then evaluated using Fisher's exact test.

Construction of microbial food-web using association rule mining

To identify directed associations between bacterial species where the presence of one is important for the presence of another (but not *vice versa*), a data-mining technique called 'association rule mining'[70] was applied to a large public collection of gut microbiome profiles in the MEDUSA database[40] (782 gut microbiome profiles from USA, China and Europe). To convert relative abundance profiles from MEDUSA into presence-absence profiles (1 if a species is present and 0 otherwise), relative abundances $< \min_j a_{ij} + 0.01 \times \left(\mathrm{Q95}_j\, a_{ij} - \min_j a_{ij} \right)$, i.e. within 1% of the minimum relative abundance values a_{ij} for species i across subjects j ($\mathrm{Q95}$ or 95% percentile was used instead of $\max$ to improve robustness to outliers), were assumed to be due to technical noise. Note that overall results were confirmed to be robust (in terms RAB placement) to a range of threshold values ($\pm$50% of original values; **Suppl. Fig. S10).** Binary association rules between species were then inferred using the apriori algorithm implemented in the Python package 'efficient_apriori' (using Confidence threshold of 0.95 and Support threshold of 0.05). After removal of transitive edges and symmetric

relationships, a total of 1166 directed association edges remained across 266 species (**Suppl. Data File 6**). Association edges and corresponding nodes for species were plotted using the hierarchical layout of Cytoscape[71], where the hierarchical level of a species is influenced by the difference between the number of outgoing and incoming edges.

Metabolic interaction analysis

Genome-scale metabolic models (GSMMs) for RABs and control species were downloaded from the AGORA database[72] (v1.03). Metabolic interactions were quantified by computing the Metabolic Support Index[73] (MSI) which quantifies the percentage of metabolic reactions in an organism that become feasible in the presence of another organism. All simulations were conducted under anoxic conditions with high-fiber diet, and mucin and bile acid derived metabolite supplementation. Species pairs with high MSI values (top 10%) were visualized using Cytoscape[71] (v3.7.2).

Promoting microbiome recovery in a mouse model

Ethics statement: Mouse experimental protocols were reviewed, approved and carried out in strict accordance to the recommendations by the Institutional Animal Care and Use Committee (IACUC) in the animal facility at Comparative Medicine, National University of Singapore (NUS). The care and use of animals for research and teaching in NUS is bound by the Singapore Animals and Birds Act, Animals and Birds (Care and Use of Animals for Scientific Purposes) Rules 2004, and is carried out in accordance with the National Advisory Committee for Laboratory Animal Research (NACLAR) Guidelines. NUS is an AAALAC-accredited institution. For this study, animals were used under Protocol R15-0135 as approved by the NUS IACUC.

*Bacterial strains and culture conditions***:** Lyophilized probiotic strains (ATCC 29148 *Bacteroides thetaiotaomicron,* DSM 20083 *Bifidobacterium adolescentis*) were revived in TSB media supplemented with 5% defibrinated sheep blood under anaerobic conditions at 37°C. Upon revival, *B. thetaiotaomicron* was subcultured and maintained in TYG media, whereas *B. adolescentis* and an environmental *Bacillus* isolate were subcultured and maintained in BHI media.

Antibiotic administration and inoculation with test strains: Eight-week-old C57BL/6J male mice from a single breeding colony were purchased from InVivos Singapore. The mice were gavaged individually with 2.5 mg ampicillin sodium salt (Sigma Aldrich) prepared in 1XPBS per day for 5 days using flexible sterile plastic feeding tubes (Instech Labs) under specific pathogen-free conditions. Upon cessation of antibiotic treatment, mice were allowed to recover for 24 hours, before the cages of mice (two mice per cage; two experimental batches) were each orally inoculated with: A) 5×10^7 CFUs *B. thetaiotaomicron,* B) 5×10^7 CFUs *Bacillus spp.,* C) 5×10^7 CFUs *B. adolescentis,* D) 5×10^7 CFUs *B. thetaiotaomicron* + 5×10^7 CFUs *B. adolescentis,* E) 5×10^7 CFUs *Bacillus spp.* + 5×10^7 CFUs *B. adolescentis,* or F) phosphate-buffered saline (PBS). Mice were kept on a 12h light/dark cycle, and water and autoclaved standard chow diet were provided ad libitum. Mice were caged in pairs in transparent plastic cages with corn cob bedding that had been pre-sterilised by autoclaving. Only mice in Bt/Bt+Ba cages where gavage was successful to result in detection in fecal samples were used for further analyses. Strains were transported from anaerobic chamber to animal facility via anaerobic "balch-type" culture tubes with aluminum seals (Chemglass Life Sciences, New Jersey, USA).

Fecal sample collection and DNA extraction: Fecal pellets were freshly collected as a cage unit (two mice per cage) over multiple times points: before antibiotic treatment (Day 0), mid-point of antibiotic treatment (Day 3), end-point of antibiotic treatment (Day 6), 1-day post-gavage (Day 7), 4-days post-gavage (Day 10), 7-days post-gavage (Day 13), 10-days post-gavage (Day 16), 13-days post-gavage (Day 19) and 16-days post-gavage (Day 22). Total bacterial DNA was extracted from fecal samples using the PowerSoil DNA isolation kit (MoBio Laboratories) according to the manufacturer's instructions.

Library preparation and deep sequencing: DNA libraries were prepared and sequenced with the same kits and workflow as used for the SG and NUH cohorts, except that the input DNA amount was 50ng.

Taxonomic profiling: For obtaining the taxonomic profiles of the mouse gut metagenomes, reads were mapped to the NR database using DIAMOND[74]. The

taxonomic classification of each sequence was then obtained by using the LCA-based approach in MEGAN[75] (default parameters, minimum score of 50).

Calculation of microbial biomass: Bacterial biomass (up to a constant factor) was estimated by taking all reads classified to bacterial taxa and normalizing by non-microbial reads. Specifically, plant or host-derived reads were used, respectively, based on the assumption that the absolute amounts of their DNA would remain roughly constant in the analyzed mouse fecal samples. Similar trends were observed for both forms of normalization (default=plant normalized), normalization based abundances were found to correlate with qPCR estimates (plant normalized, r=0.73, *p*-value=10^{-4}; host normalized, r=0.82, *p*-value=3.5×10^{-6}), and the observed differences between Bt and Bt+Ba groups versus other groups were also validated using qPCR (day 10, fold-change=94-170×). Note that sequencing based biomass estimates have the advantage that they allow us to subtract reads belonging to the gavaged species and are also not affected due to variations in 16S rRNA copy number across taxa. This approach was also further validated based on spike-in of isolate DNA into mouse stool samples showing that (i) qPCR based measurement of 16S rRNA DNA copies correlates highly with microbial CFUs (slope=0.98, R^2=1.0; **Suppl. Fig. S11a**), (ii) Metagenomic sequencing based calculation of host-normalized microbial reads accurately quantitated varying microbial CFUs (**Suppl. Fig. S11b**).

qPCR Analysis: Absolute quantification of the 16S rRNA gene was done by quantitative PCR (qPCR). A pair of universal 16S bacterial primers[76] were used to amplify DNA extracted from the six different treatment groups on days 0, 3, 10 and 13 (**Suppl. Table 1**). Reactions were prepared on a 384-well plate, in triplicates, using 5 µL of PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, Massachusetts, USA), 0.5 µL of 5µM primers and 1 µL of 10× diluted DNA, in a total volume of 10 µL for each reaction. The ViiA 7 Real-Time PCR System (Thermo Fisher Scientific, Massachusetts, USA) was used for qPCR with the following amplification parameters: 1 cycle of 95˚C for 2 min, 40 cycles of 95˚C for 15 s, 60˚C for 15 s, and 72˚C for 1 min. A standard curve was created using serial dilution of synthesized double-stranded DNA oligomers (gBLOCK, Integrated DNA Technologies, Inc., Iowa, USA; **Suppl. Table 1**) to convert CT values to

copy numbers. Copy numbers from day 0 were used to scale bacterial abundances to the same starting baseline.

Data Availability

Illumina sequencing data for this study (mouse models) has been deposited to the Sequence Read Archive under project ID SRP142225. Samples are labelled in SRA as shorthand (e.g. PBS6D22) where “PBS” represents gavage condition, “6” represents cage number and “D22” represents day of sampling.

Code Availability

Analysis scripts are available at https://github.com/CSB5/Recovery_Determinants_Study.

Acknowledgements

This work was supported by funding from the National Healthcare Group (NHG-CSCS/12008), the National Medical Research Council, the National Research Foundation and A*STAR, Singapore.

Author Contributions

N.N., Y.G., B.Y. and S.C. planned and designed the project. B.Y., L.C., T.B. and D.L. contributed the clinical cohorts. Y.H.T and I.R.L. performed the mice experiments, analysed by T.N. under Y.G.’s and N.N.’s guidance. A.H.Q.N, and K.M.L. conducted the wet-lab experiments with K.R.C.’s and N.N.’s guidance. T.S.G., K.R.C., A.R., C.L. and T.N. coordinated computational analysis with K.R.’s and N.N.’s supervision. N.N. wrote the manuscript with inputs from all authors.

Competing Interests Statement

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1: **Gut microbiome recovery profiles and key associated taxa.** (a) Density plots showing the two different recovery profiles for microbial diversity (Simpson) that were observed in the antibiotic treatment cohorts (CA, SG). (b) Principal Component Analysis (PCA) plot showing the distribution of post-antibiotic gut microbiome profiles for recoverers and non-recoverers in relation to healthy control gut microbiome profiles (CA; n=8 for recoverers, n=7 for non-recoverers and n=18 for controls). (c) Boxplots showing the distribution of Bray Curtis Distances for post-antibiotic gut microbiomes for recoverers and non-recoverers in relation to healthy controls (median value; CA; n=8 for recoverers and n=7 for non-recoverers). '***' represents *p*-value (one-sided Wilcoxon test) less than 0.001. (d) Relative abundance boxplots for 6 of the RABs that were identified in at least 3/4 cohorts (**Table 2**) based on all timepoints. Note that '*', '**' and ''***' denote cohort-specific FDR adjusted *p*-values (one-sided Wilcoxon test; n=24(CA); 32(SW); 16(EN); 41(SG) samples for recoverers and n=21(CA); 24(SW); 24(EN); 22(SG) samples for non-recoverers) less than 0.05, 0.01 and 0.001 respectively. For all subfigures, boxplots are represented with center line: median; box limits: upper and lower quartiles; whiskers: 1.5× interquartile range; outlier points not included in visualization.

Figure 2: **Mechanistic model linking microbial functions with recovery.** Subfigures provide evidence for a model of microbiome recovery based on RABs being enriched for carbohydrate degradation capabilities (CAZyme), which in turn promote faster community growth (CGR), and ultimately microbiome recovery (associations shown in each subfigure are highlighted in blue). (a) Empirical distributions for the number of CAZyme families in RABs and non-RABs showing that RABs are strongly enriched for CAZymes (two-sided Wilcoxon test;). (b) Bean plots showing the variation in the number of CAZyme families (empirical distributions) detected in the gut microbiomes of recoverers and non-recoverers in the CA and SG cohorts (all timepoints; n=24(CA), 41(SG) recoverers and n=21(CA), 22(SG) non-recoverers). In both cohorts, recoverers have more CAZyme families represented in their metagenomes (one–sided Wilcoxon test). (c) Bean plots showing variation in the gut microbial community growth rate

(empirical distributions) of recoverers and non-recoverers in the CA and SG cohorts (all timepoints; n=18(CA), n=40(SG) recoverers and n=21(CA), n=18(SG) non-recoverers). In both cohorts, recoverers have higher community growth rates (one-sided Wilcoxon test). (d) Bean plots showing that the abundance of RABs in the pre- and during phase of antibiotic treatment was better correlated (Spearman) to the post-treatment community growth rate of individuals in the CA cohort compared to non-RAB species (empirical distributions; one-sided Wilcoxon test; n=21(CA) RABs, n=89(CA) non-RABs ; *p*-value>0.1 for SG cohort). (e) Correlation between the number of CAZyme families detected and the overall community growth rate across all gut microbiomes constituting the CA and SG cohorts (all timepoints). In both cohorts, community growth rates were consistently correlated with CAZyme diversity. Note that '*', '**' and ''***' denote *p*-values less than 0.05, 0.01 and 0.001 respectively for all subfigures. For all subfigures, bean plots are represented with beanline: median.

Figure 3: **Role of RABs in ecological recovery via the microbial food web.** (a) Graph showing network structure of microbial dependencies inferred using an association rule mining approach, where an edge from species A to species B indicates that A's presence is required to have B in the community. Nodes are ordered from the bottom to the top such that species at the bottom have more outgoing edges than incoming edges ('Primary Species'), while species at the top have more incoming edges than outgoing edges ('Tertiary Species'). RABs (highlighted in different colors based on the genus they belong to) were observed either at the bottom or top of the graph. Many RABs at the bottom of the graph were from cluster 1 (degradation profile; **Suppl. Fig. S6**), defined by mucin degrading CAZymes. Clusters based on abundance profile over time (**Suppl. Fig. S2**) are indicated using numbers and do not seem to be biased in different regions of the graph. (b) Schematic representation of the gut showing a model for microbiome recovery based on these observations. RABs from cluster 1 (**Suppl. Fig. S6**) colonize the epithelial mucosa better because of their mucin degrading capabilities (step 1)[39,42,43], and since they can also break down dietary plant and animal derived carbohydrates[42] (step 2), they act as primary species that facilitate the growth of other species[44] (step 3). Some of the tertiary RABs and other species can produce short chain fatty acids (SCFAs), which are then utilized by colonocytes for their growth

leading to increased mucin production[45,46] (step 4). This positive feedback loop may enable faster ecological recovery in terms of diversity and biomass.

Figure 4: Promoting microbiome recovery in a mouse model using RABs. (a) Schematic depicting the design of a mouse model experiment to study the impact of RABs in promoting microbiome recovery. Mice were given antibiotics for 5 days, followed by a rest day and gavage of different RABs and controls (Vehicle: n=5, Ba: n=6, Bt: n=2, and Bt+Ba: n=2, where n represents cage units). Shotgun metagenomics was then used to monitor microbiome changes every 3 days. (b) Microbial biomass (median ± 1 MAD) in different groups of mice across time (excluding gavaged species). (c) Microbiome diversity (Simpson) (median ± 1 MAD) in different groups of mice across time. Stars ('*') indicate timepoints where Bt+Ba group differs from other groups (one-sided Wilcoxon test *p*-value<0.1). (d, e, f) Reads per million (RPM) mapping to CAZymes (median ± 1 MAD) associated with plant/animal cell wall, mucin and peptidoglycan degradation, respectively, across different experimental groups and timepoints. Stars in all subfigures ('**') indicate timepoints where the Bt and Bt+Ba groups were significantly different from other groups (one-sided Wilcoxon test *p*-value<0.01).

TABLES

Table 1: Details of the different cohorts used in this study.

Cohort	No. of Subjects/ Samples	Sequencing	Age Range	Antibiotics Used
Singapore (SG)	27/129	Shotgun Metagenomic	32-81	Primarily Co-amoxiclav and Clarithromycin
Canada (CA)	24/72	Shotgun Metagenomic	21-35	Cefprozil
England (EN)	37/219	16S rRNA	24-26	Amoxicillin
Sweden (SW)	29/173	16S rRNA	22-30	Clindamycin/Ciprofloxacin
NUH	24/72	Shotgun Metagenomic	23-40	Co-amoxiclav

Table 2: List of recovery associated bacterial taxa (RABs). RABs are ordered by the number of cohorts in which they are significantly associated (bold *p*-values).

Species	Cohort-specific FDR adjusted *p*-value				Known functions or associations in gut microbiome	NUH *p*-value
	Canada	England	Sweden	S'pore		
Bacteroides uniformis	**0.009**	**0.003**	**0.005**	**0.019**	Negatively associated with obesity[41]	0.354
Alistipes putredinis	**0.002**	0.737	**0.011**	**<0.001**	Associated with weight loss in obese individuals[50]	**0.011**
Alistipes shahii	**0.009**	**0.018**	0.113	**<0.001**		**0.026**
Bacteroides thetaiotaomicron	**0.002**	0.953	**0.011**	**0.002**	Diverse carbohydrate degrading enzymes[77]	**0.007**
Parabacteroides distasonis	**0.004**	0.927	**0.005**	**<0.001**	Carbohydrate degrading[78]	0.218
Coprococcus catus	**0.034**	**0.003**	**0.022**	0.492		**0.096**
Bifidobacterium adolescentis	**0.003**	**0.014**	0.342	**0.006**	Known probiotic[79]	**0.008**
Ruminococcus bromii	**0.023**	**0.014**	0.477	**0.046**		0.138
Subdoligranulum variabile	**0.002**	**0.039**	**0.039**	0.401	Produces butyrate[80]	0.197
Bacteroides stercoris	0.351	**0.013**	**0.011**	**0.050**		0.977
Bacteroides eggerthii	0.087	0.570	**0.016**	**0.022**		**0.039**
Bacteroides coprocola	0.075	**0.003**	0.933	**0.015**		**0.030**
Bifidobacterium bifidum	**0.049**	0.737	0.239	**0.013**		0.327
Roseburia inulinivorans	0.133	**0.024**	**0.022**	0.775	Produces butyrate[80,81]	0.308
Bacteroides caccae	**0.001**	0.737	0.156	**<0.001**	Negatively associated with obesity[50]	**0.003**
Faecalibacterium prausnitzii	**0.001**	**0.013**	0.150	0.504	Butyrate producing with anti-inflammatory properties[54]	**0.081**
Ruminococcus torques	0.775	**0.013**	0.662	**0.015**	Degrades mucin[43]	**0.003**
Bifidobacterium longum	**0.033**	0.737	0.150	**0.021**	Known probiotic[82]	0.378
Bacteroides intestinalis	**0.002**	0.737	0.574	**<0.001**	Carbohydrate degrading; Negatively associated with obesity[83]	0.377
Desulfovibrio piger	0.223	0.149	**0.011**	**0.023**	Sulfate-reducing bacteria	**0.055**
Parabacteroides johnsonii	**0.005**	0.439	0.933	**0.012**		**0.030**

REFERENCES

1 Marchesi, J. R. *et al.* The gut microbiota and host health: a new clinical frontier. *Gut* **65**, 330-339, doi:10.1136/gutjnl-2015-309990 (2016).

2 Bäumler, A. J. & Sperandio, V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85-93 (2016).

3 Kampmann, C., Dicksved, J., Engstrand, L. & Rautelin, H. Composition of human faecal microbiota in resistance to Campylobacter infection. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **22**, 61 e61-68, doi:10.1016/j.cmi.2015.09.004 (2016).

4 Gilbert, J. A. *et al.* Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94-103, doi:10.1038/nature18850 (2016).

5 Routy, B. *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91-97, doi:10.1126/science.aan3706 (2018).

6 Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359**, 97-103, doi:10.1126/science.aan4236 (2018).

7 Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS biology* **6**, e280, doi:10.1371/journal.pbio.0060280 (2008).

8 Zaura, E. *et al.* Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. *mBio* **6**, e01693-01615, doi:10.1128/mBio.01693-15 (2015).

9 Perez-Cobas, A. E. *et al.* Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* **62**, 1591-1601, doi:10.1136/gutjnl-2012-303184 (2013).

10 Klein, E. Y. *et al.* Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proceedings of the National Academy of Sciences* **115**, E3463-E3470, doi:10.1073/pnas.1717295115 (2018).

11 Blaser, M. J. & Falkow, S. What are the consequences of the disappearing human microbiota? *Nature Reviews Microbiology* **7**, 887, doi:10.1038/nrmicro2245 (2009).

12 Blaser, M. J. Antibiotic use and its consequences for the normal microbiome. *Science* **352**, 544-545, doi:10.1126/science.aad9358 (2016).

13 Stevens, V., Dumyati, G., Fine, L. S., Fisher, S. G. & van Wijngaarden, E. Cumulative antibiotic exposures over time and the risk of Clostridium difficile infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **53**, 42-48, doi:10.1093/cid/cir301 (2011).

14 Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241-244, doi:10.1038/nature10571 (2011).

15 Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219-222, doi:10.1038/nature12212 (2013).

16 Raymond, F. *et al.* The initial state of the human gut microbiome determines its reshaping by antibiotics. *The ISME journal* **10**, 707-720, doi:10.1038/ismej.2015.148 (2016).

17 Livanos, A. E. *et al.* Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nature microbiology* **1**, 16140, doi:10.1038/nmicrobiol.2016.140 (2016).
18 Cox, L. M. & Blaser, M. J. Antibiotics in early life and obesity. *Nature reviews. Endocrinology* **11**, 182-190, doi:10.1038/nrendo.2014.210 (2015).
19 Langdon, A., Crook, N. & Dantas, G. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome medicine* **8**, 39, doi:10.1186/s13073-016-0294-z (2016).
20 Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4554-4561, doi:10.1073/pnas.1000087107 (2011).
21 Jakobsson, H. E. *et al.* Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS One* **5**, e9836, doi:10.1371/journal.pone.0009836 (2010).
22 Raymond, F., Deraspe, M., Boissinot, M., Bergeron, M. G. & Corbeil, J. Partial recovery of microbiomes after antibiotic treatment. *Gut Microbes* **7**, 428-434, doi:10.1080/19490976.2016.1216747 (2016).
23 Palleja, A. *et al.* Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature microbiology* **3**, 1255-1265, doi:10.1038/s41564-018-0257-9 (2018).
24 Suez, J. *et al.* Post-Antibiotic Gut Mucosal Microbiome Reconstitution Is Impaired by Probiotics and Improved by Autologous FMT. *Cell* **174**, 1406-1423 e1416, doi:10.1016/j.cell.2018.08.047 (2018).
25 Harvey, E., Gounand, I., Ward, C. L., Altermatt, F. & Cadotte, M. Bridging ecology and conservation: from ecological networks to ecosystem function. *Journal of Applied Ecology* **54**, 371-379, doi:doi:10.1111/1365-2664.12769 (2017).
26 Bascompte, J. & Stouffer, D. B. The assembly and disassembly of ecological networks. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**, 1781-1787, doi:10.1098/rstb.2008.0226 (2009).
27 Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214, doi:10.1038/nature11234 (2012).
28 Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**, 593-621, doi:10.1128/CMR.00008-07 (2007).
29 Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat Rev Microbiol* **10**, 538-550, doi:10.1038/nrmicro2832 (2012).
30 Solden, L. M. *et al.* Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nature microbiology* **3**, 1274-1284, doi:10.1038/s41564-018-0225-4 (2018).
31 Adamowicz, E. M., Flynn, J., Hunter, R. C. & Harcombe, W. R. Cross-feeding modulates antibiotic tolerance in bacterial communities. *The ISME journal* **12**, 2723-2735, doi:10.1038/s41396-018-0212-z (2018).
32 Wang, J. & Jia, H. Metagenome-wide association studies: fine-mining the microbiome. *Nat Rev Microbiol* **14**, 508-522, doi:10.1038/nrmicro.2016.83 (2016).
33 David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559-563, doi:10.1038/nature12820 (2014).

34 Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210-215, doi:10.1038/nature25973 (2018).

35 Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**, 16731-16736, doi:10.1073/pnas.0804812105 (2008).

36 Takahashi, K. *et al.* Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease. *Digestion* **93**, 59-65, doi:10.1159/000441768 (2016).

37 El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **11**, 497-504, doi:10.1038/nrmicro3050 (2013).

38 Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101-1106, doi:10.1126/science.aac4812 (2015).

39 Sicard, J. F., Le Bihan, G., Vogeleer, P., Jacques, M. & Harel, J. Interactions of Intestinal Bacteria with Components of the Intestinal Mucus. *Frontiers in cellular and infection microbiology* **7**, 387, doi:10.3389/fcimb.2017.00387 (2017).

40 Karlsson, F. H., Nookaew, I. & Nielsen, J. Metagenomic data utilization and analysis (MEDUSA) and construction of a global gut microbial gene catalogue. *PLoS computational biology* **10**, e1003706, doi:10.1371/journal.pcbi.1003706 (2014).

41 Gauffin Cano, P., Santacruz, A., Moya, A. & Sanz, Y. Bacteroides uniformis CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. *PLoS One* **7**, e41079, doi:10.1371/journal.pone.0041079 (2012).

42 Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**, 289-306, doi:10.4161/gmic.19897 (2012).

43 Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Frontiers in genetics* **6**, 81, doi:10.3389/fgene.2015.00081 (2015).

44 Arike, L. & Hansson, G. C. The Densely O-Glycosylated MUC2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria. *J Mol Biol* **428**, 3221-3229, doi:10.1016/j.jmb.2016.02.010 (2016).

45 Finnie, I. A., Dwarakanath, A. D., Taylor, B. A. & Rhodes, J. M. Colonic mucin synthesis is increased by sodium butyrate. *Gut* **36**, 93-99, doi:10.1136/gut.36.1.93 (1995).

46 Willemsen, L. E., Koetsier, M. A., van Deventer, S. J. & van Tol, E. A. Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E(1) and E(2) production by intestinal myofibroblasts. *Gut* **52**, 1442-1447, doi:10.1136/gut.52.10.1442 (2003).

47 Cornick, S., Tawiah, A. & Chadee, K. Roles and regulation of the mucus barrier in the gut. *Tissue barriers* **3**, e982426, doi:10.4161/21688370.2014.982426 (2015).

48 Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Backhed, F. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **165**, 1332-1345, doi:10.1016/j.cell.2016.05.041 (2016).

49 Wampach, L. *et al.* Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. *Frontiers in microbiology* **8**, doi:10.3389/fmicb.2017.00738 (2017).

50 Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214, doi:10.1126/science.1241214 (2013).

51 Jiang, T. *et al.* Apple-Derived Pectin Modulates Gut Microbiota, Improves Gut Barrier Function, and Attenuates Metabolic Endotoxemia in Rats with Diet-Induced Obesity. *Nutrients* **8**, 126, doi:10.3390/nu8030126 (2016).

52 Wei, Y. *et al.* Pectin enhances the effect of fecal microbiota transplantation in ulcerative colitis by delaying the loss of diversity of gut flora. *BMC microbiology* **16**, 255, doi:10.1186/s12866-016-0869-2 (2016).

53 Onrust, L. *et al.* Steering Endogenous Butyrate Production in the Intestinal Tract of Broilers as a Tool to Improve Gut Health. *Frontiers in veterinary science* **2**, 75, doi:10.3389/fvets.2015.00075 (2015).

54 Scott, K. P., Martin, J. C., Duncan, S. H. & Flint, H. J. Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro. *FEMS microbiology ecology* **87**, 30-40, doi:10.1111/1574-6941.12186 (2014).

55 Van den Abbeele, P. *et al.* Microbial community development in a dynamic gut model is reproducible, colon region specific, and selective for Bacteroidetes and Clostridium cluster IX. *Applied and environmental microbiology* **76**, 5237-5246, doi:10.1128/AEM.00759-10 (2010).

56 Sung, J. *et al.* Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nature communications* **8**, 15393, doi:10.1038/ncomms15393 (2017).

57 Mimee, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell systems* **2**, 214, doi:10.1016/j.cels.2016.03.007 (2016).

58 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

59 Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS computational biology* **8**, e1002358, doi:10.1371/journal.pcbi.1002358 (2012).

60 Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature methods* **9**, 811-814, doi:10.1038/nmeth.2066 (2012).

61 Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nature methods* **15**, 962-968, doi:10.1038/s41592-018-0176-y (2018).

62 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).

63 Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* **31**, 814-821, doi:10.1038/nbt.2676 (2013).

64 Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-108, doi:10.1126/science.1208344 (2011).
65 Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome biology* **12**, R60, doi:10.1186/gb-2011-12-6-r60 (2011).
66 Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **40**, W445-451, doi:10.1093/nar/gks479 (2012).
67 Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial agents and chemotherapy* **58**, 212-220, doi:10.1128/AAC.01310-13 (2014).
68 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology* **15**, R46, doi:10.1186/gb-2014-15-3-r46 (2014).
69 Cantarel, B. L., Lombard, V. & Henrissat, B. Complex carbohydrate utilization by the healthy human microbiome. *PLoS One* **7**, e28742, doi:10.1371/journal.pone.0028742 (2012).
70 Hipp, J., Güntzer, U. & Nakhaeizadeh, G. Algorithms for association rule mining --- a general survey and comparison. *ACM SIGKDD Explorations Newsletter* **2**, 58-64, doi:10.1145/360402.360421 (2000).
71 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
72 Magnusdottir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature biotechnology* **35**, 81-89, doi:10.1038/nbt.3703 (2017).
73 Ravikrishnan, A., Blank, L. M., Srivastava, S. & Raman, K. Investigating metabolic interactions in a microbial co-culture through integrated modelling and experiments. *Computational and Structural Biotechnology Journal*, doi:https://doi.org/10.1016/j.csbj.2020.03.019 (2020).
74 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).
75 Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome research* **21**, 1552-1560, doi:10.1101/gr.120618.111 (2011).
76 Chua, M. C. *et al.* Effect of Synbiotic on the Gut Microbiota of Caesarean Delivered Infants: A Randomized, Double-Blind, Multicenter Study. *Journal of pediatric gastroenterology and nutrition*, doi:10.1097/MPG.0000000000001623 (2017).
77 Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**, 2074-2076, doi:10.1126/science.1080029 (2003).
78 Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut bacteroidetes: the food connection. *Frontiers in microbiology* **2**, 93, doi:10.3389/fmicb.2011.00093 (2011).
79 Fernandez-Duarte, K. P., Olaya-Galan, N. N., Salas-Cardenas, S. P., Lopez-Rozo, J. & Gutierrez-Fernandez, M. F. Bifidobacterium adolescentis (DSM 20083) and Lactobacillus casei (Lafti L26-DSL): Probiotics Able to Block the In Vitro Adherence of Rotavirus in

MA104 Cells. *Probiotics and antimicrobial proteins*, doi:10.1007/s12602-017-9277-7 (2017).

80 Thomas, L. V., Ockhuizen, T. & Suzuki, K. Exploring the influence of the gut microbiota and probiotics on health: a symposium report. *The British journal of nutrition* **112 Suppl 1**, S1-18, doi:10.1017/S0007114514001275 (2014).

81 Riviere, A., Selak, M., Lantin, D., Leroy, F. & De Vuyst, L. Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. *Frontiers in microbiology* **7**, 979, doi:10.3389/fmicb.2016.00979 (2016).

82 Lee, D. K. *et al.* Probiotic bacteria, B. longum and L. acidophilus inhibit infection by rotavirus in vitro and decrease the duration of diarrhea in pediatric patients. *Clinics and research in hepatology and gastroenterology* **39**, 237-244, doi:10.1016/j.clinre.2014.09.006 (2015).

83 Dewulf, E. M. *et al.* Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women. *Gut* **62**, 1112-1121, doi:10.1136/gutjnl-2012-303304 (2013).

Figure 1

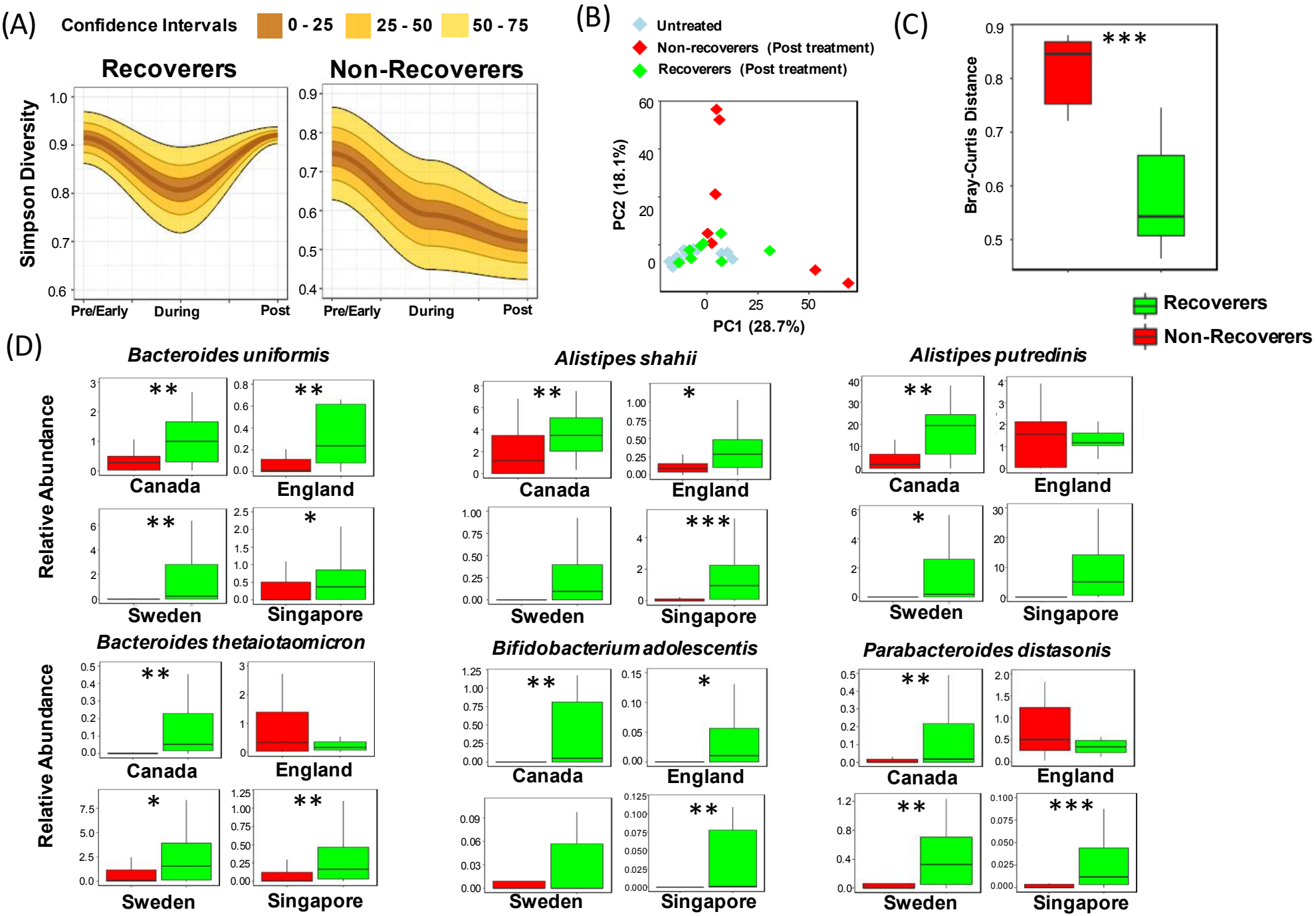

(A)
Confidence Intervals
0 - 25
25 - 50
50 - 75
Recoverers
Non-Recoverers
Simpson Diversity
Pre/Early
During
Post
(B)
Untreated
Non-recoverers (Post treatment)
Recoverers (Post treatment)
PC2 (18.1%)
PC1 (28.7%)
(C)
Bray-Curtis Distance
***
Recoverers
Non-Recoverers
(D)
Bacteroides uniformis
Alistipes shahii
Alistipes putredinis
Bacteroides thetaiotaomicron
Bifidobacterium adolescentis
Parabacteroides distasonis
Relative Abundance
Canada
England
Sweden
Singapore

Figure 2

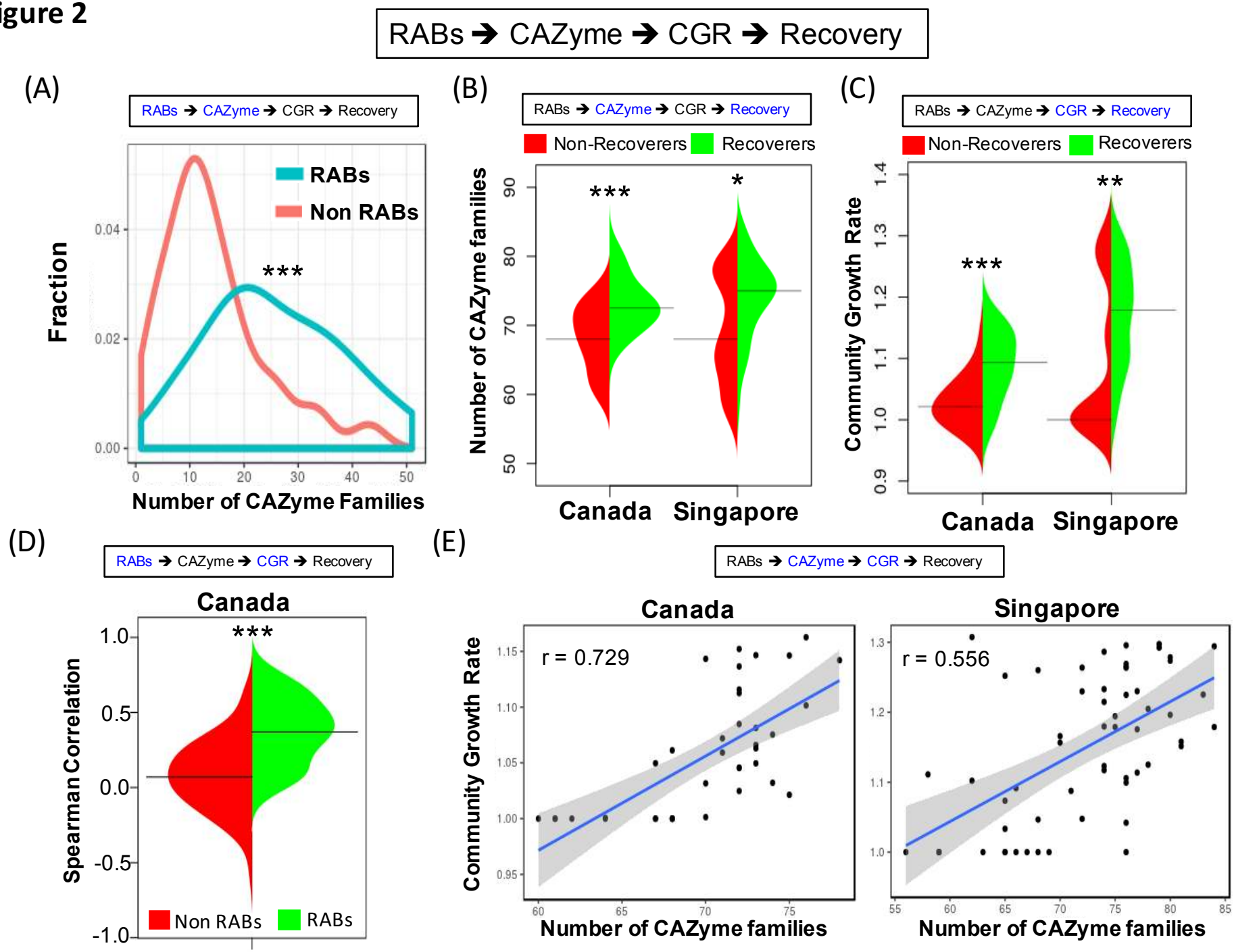

RABs ➔ CAZyme ➔ CGR ➔ Recovery
(A)
RABs ➔ CAZyme ➔ CGR ➔ Recovery
RABs
Non RABs
***
Fraction
0.04
0.02
0.00
0
10
20
30
40
50
Number of CAZyme Families
(B)
RABs ➔ CAZyme ➔ CGR ➔ Recovery
Non-Recoverers
Recoverers
***
*
Number of CAZyme families
90
80
70
60
50
Canada
Singapore
(C)
RABs ➔ CAZyme ➔ CGR ➔ Recovery
Non-Recoverers
Recoverers
***
**
Community Growth Rate
1.4
1.3
1.2
1.1
1.0
0.9
Canada
Singapore
(D)
RABs ➔ CAZyme ➔ CGR ➔ Recovery
Canada
***
Spearman Correlation
1.0
0.5
0.0
-0.5
-1.0
Non RABs
RABs
(E)
RABs ➔ CAZyme ➔ CGR ➔ Recovery
Canada
r = 0.729
Community Growth Rate
1.15
1.10
1.05
1.00
0.95
60
65
70
75
Number of CAZyme families
Singapore
r = 0.556
1.3
1.2
1.1
1.0
55
60
65
70
75
80
85
Number of CAZyme families

Figure 3A

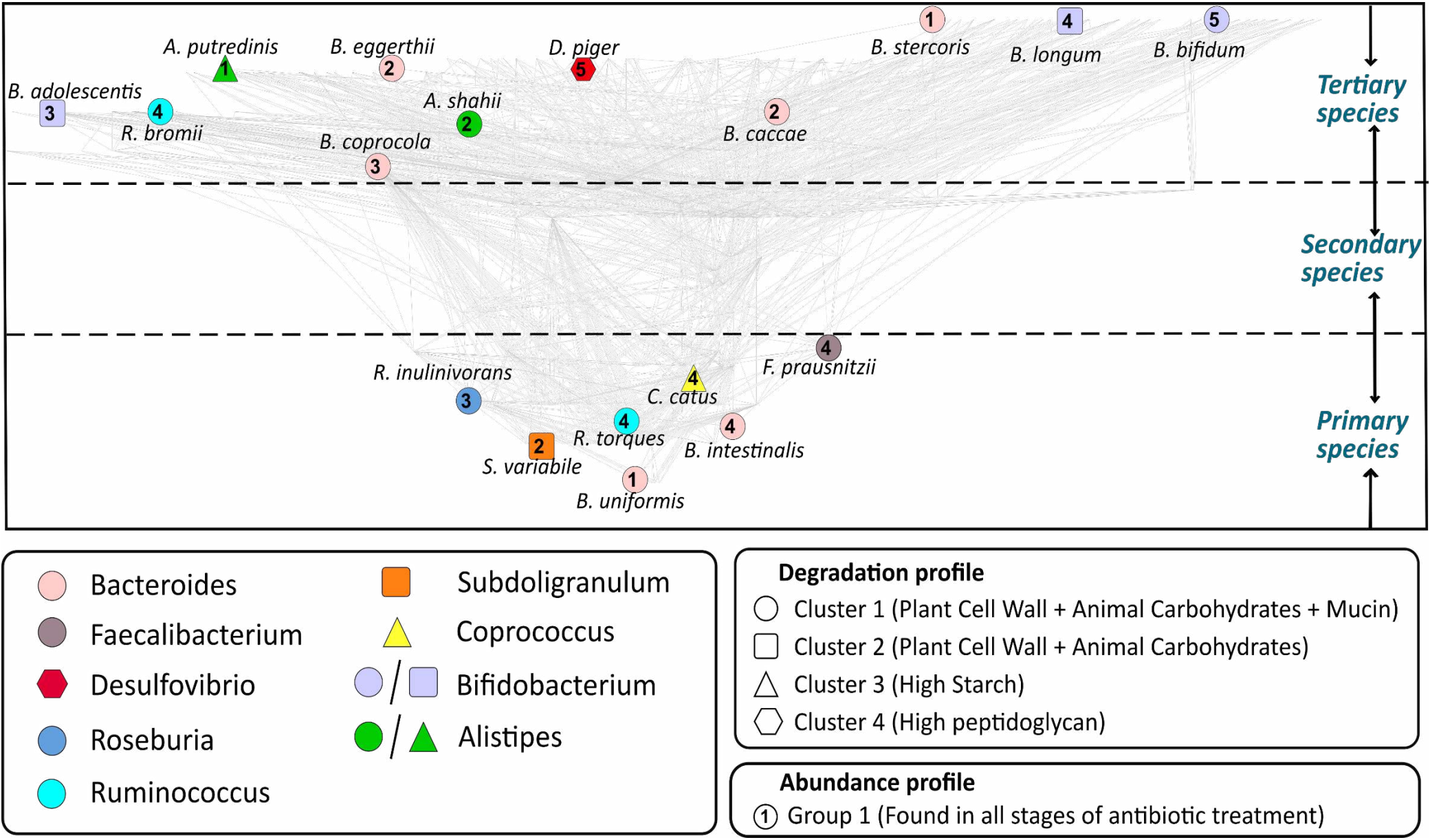

A. putredinis
B. eggerthii
D. piger
B. stercoris
B. longum
B. bifidum
B. adolescentis
R. bromii
A. shahii
B. coprocola
B. caccae
Tertiary species
Secondary species
R. inulinivorans
C. catus
F. prausnitzii
R. torques
B. intestinalis
S. variabile
B. uniformis
Primary species
Bacteroides
Faecalibacterium
Desulfovibrio
Roseburia
Ruminococcus
Subdoligranulum
Coprococcus
Bifidobacterium
Alistipes
Degradation profile
Cluster 1 (Plant Cell Wall + Animal Carbohydrates + Mucin)
Cluster 2 (Plant Cell Wall + Animal Carbohydrates)
Cluster 3 (High Starch)
Cluster 4 (High peptidoglycan)
Abundance profile
Group 1 (Found in all stages of antibiotic treatment)

Figure 3B

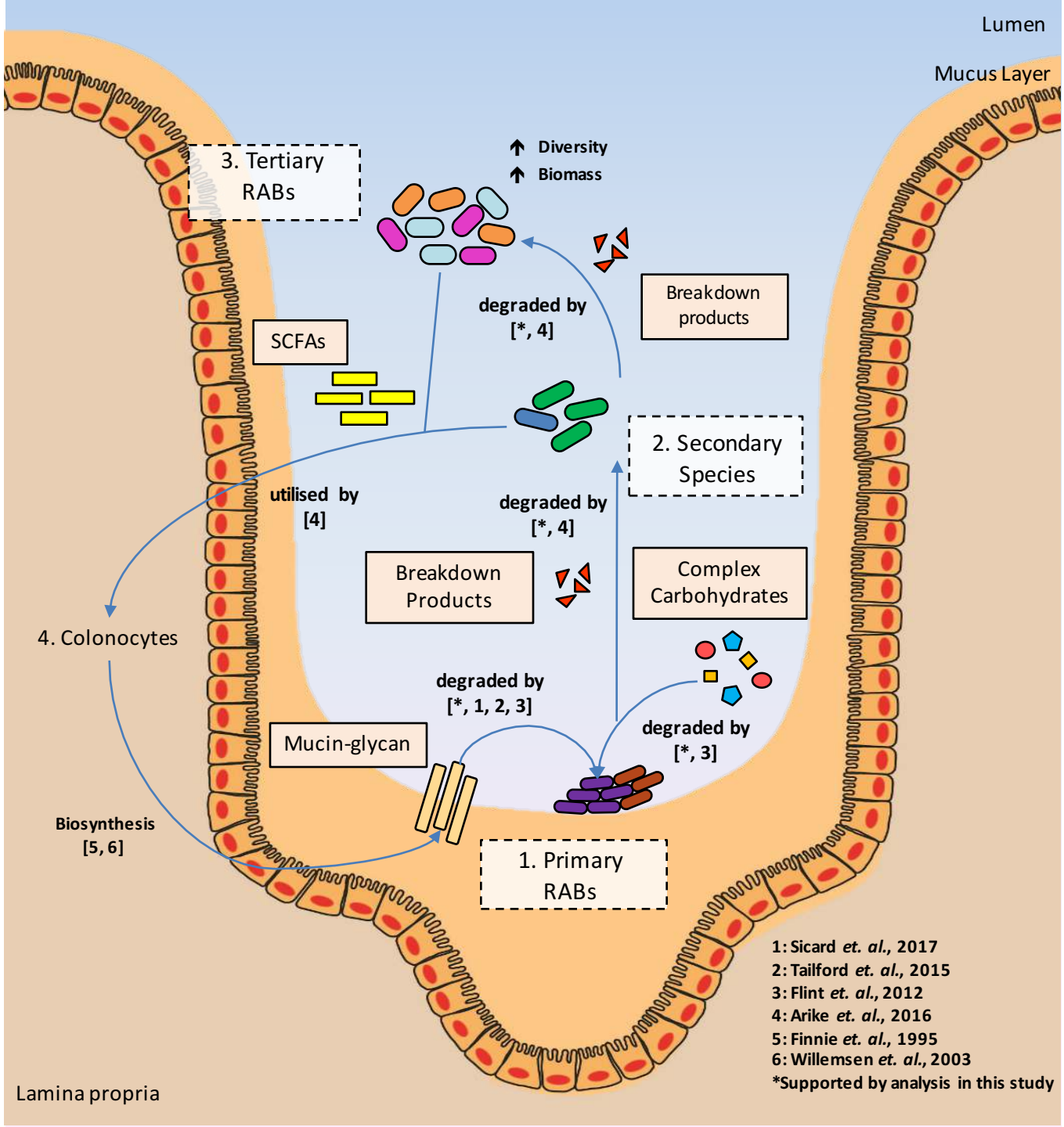
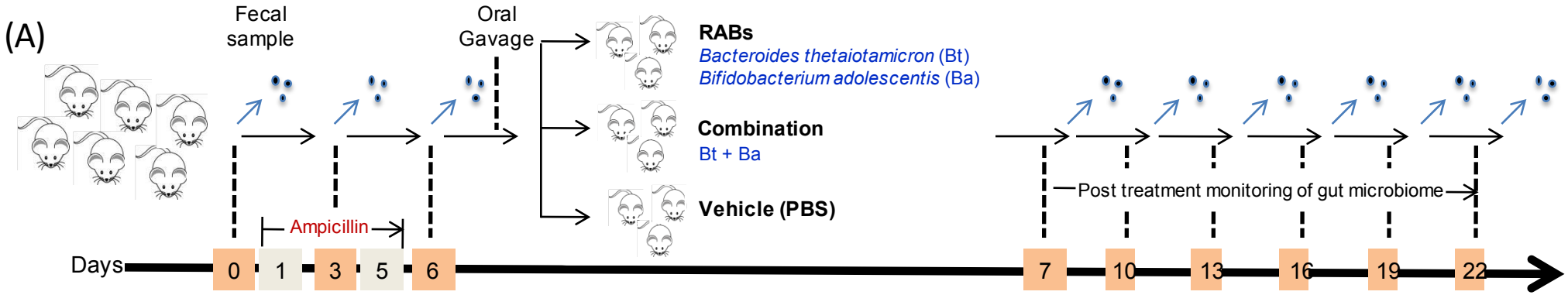
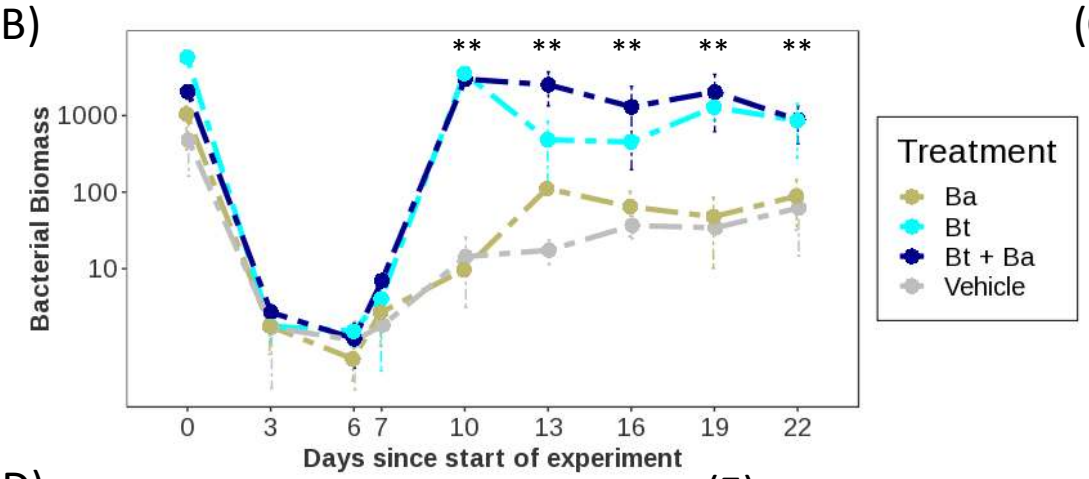
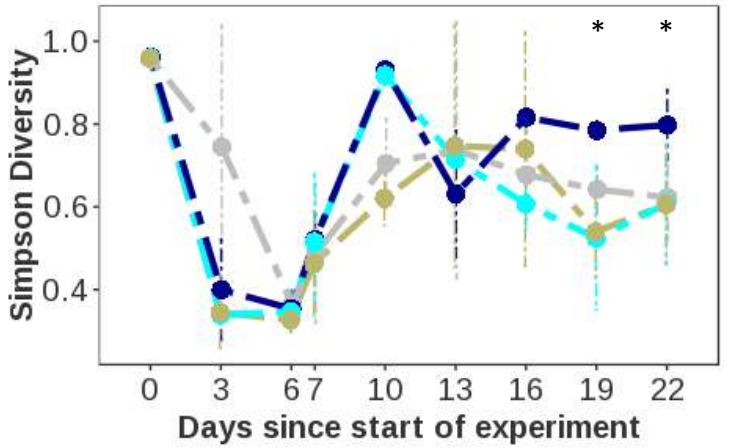
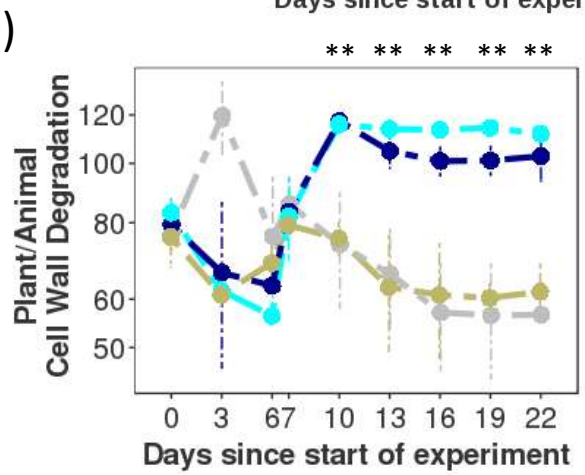

Lumen
Mucus Layer
3. Tertiary RABs
Diversity
Biomass
degraded by [*, 4]
Breakdown products
SCFAs
2. Secondary Species
utilised by [4]
degraded by [*, 4]
Breakdown Products
Complex Carbohydrates
4. Colonocytes
degraded by [*, 1, 2, 3]
degraded by [*, 3]
Mucin-glycan
Biosynthesis [5, 6]
1. Primary RABs
Lamina propria
1: Sicard et. al., 2017
2: Tailford et. al., 2015
3: Flint et. al., 2012
4: Arike et. al., 2016
5: Finnie et. al., 1995
6: Willemsen et. al., 2003
*Supported by analysis in this study

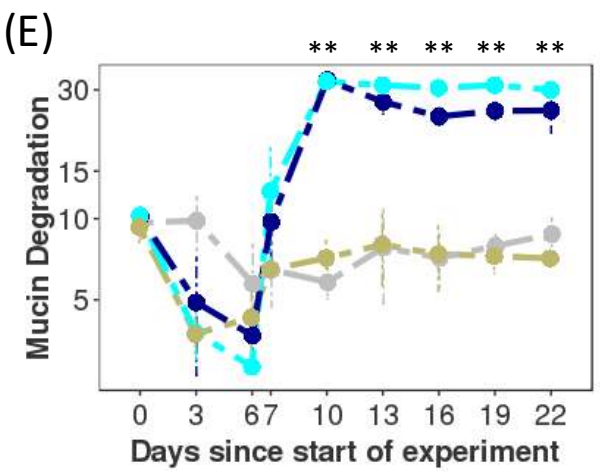
Figure 4

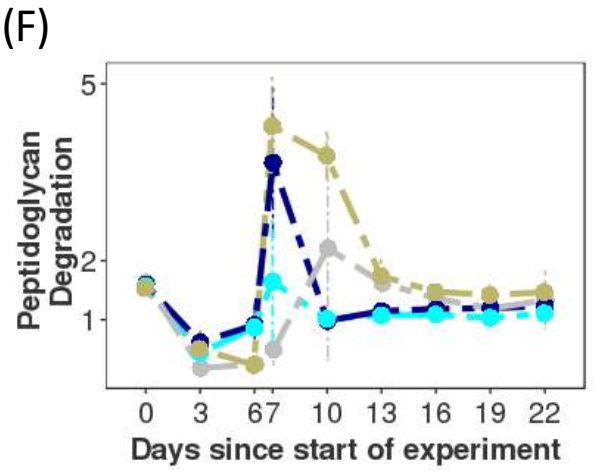

(A)
Fecal sample
Oral Gavage
RABs
Bacteroides thetaiotamicron (Bt)
Bifidobacterium adolescentis (Ba)
Combination
Bt + Ba
Vehicle (PBS)
Ampicillin
Post treatment monitoring of gut microbiome
Days
0
1
3
5
6
7
10
13
16
19
22

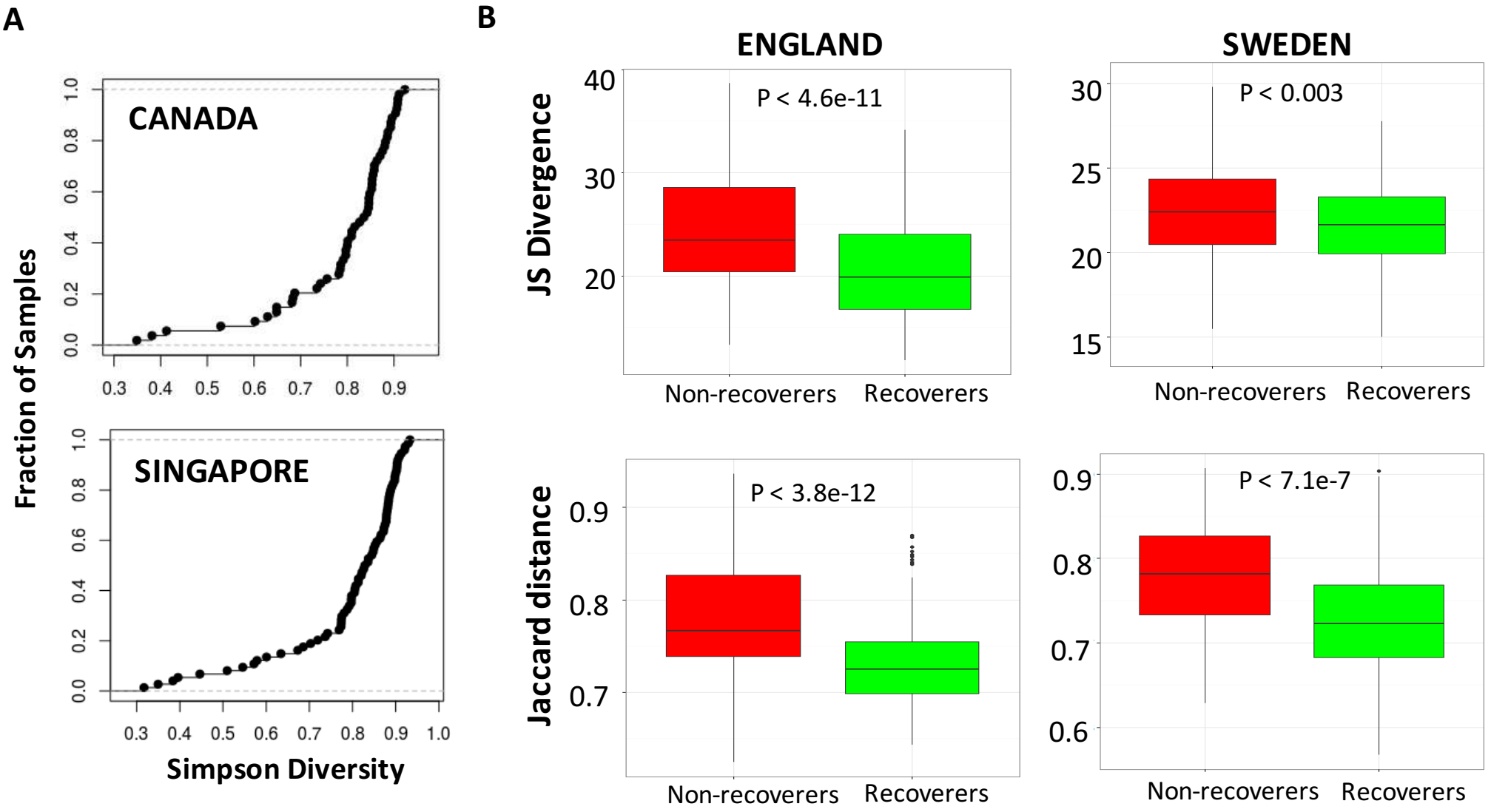

(B)
Bacterial Biomass
1000
100
10
**
Treatment
Ba
Bt
Bt + Ba
Vehicle
0 3 6 7 10 13 16 19 22
Days since start of experiment


(C)
Simpson Diversity
1.0
0.8
0.6
0.4
*
0 3 6 7 10 13 16 19 22
Days since start of experiment

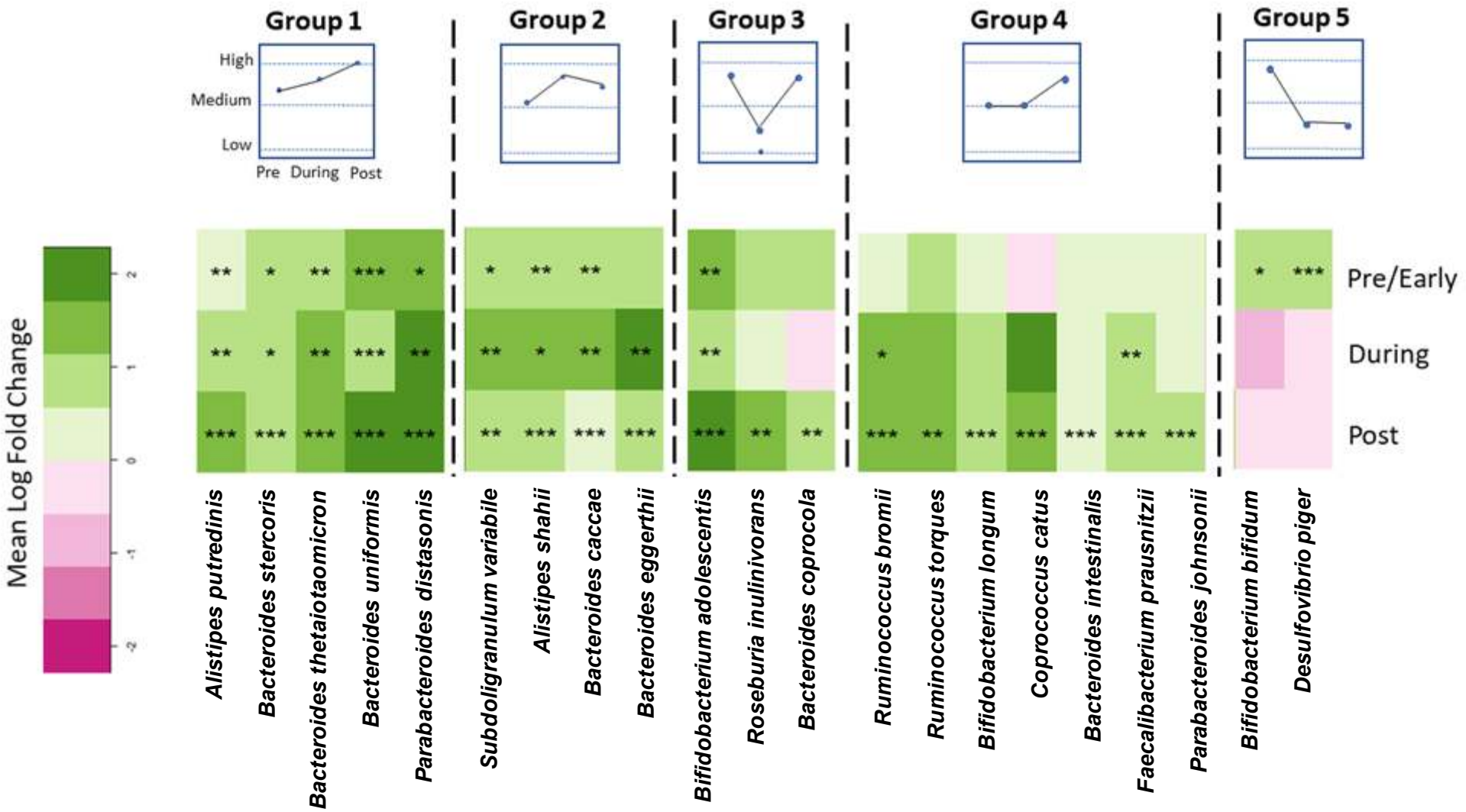

(D)
Plant/Animal Cell Wall Degradation
120
100
80
60
50
**
0 3 67 10 13 16 19 22
Days since start of experiment


(E)
Mucin Degradation
30
15
10
5
**
0 3 67 10 13 16 19 22
Days since start of experiment

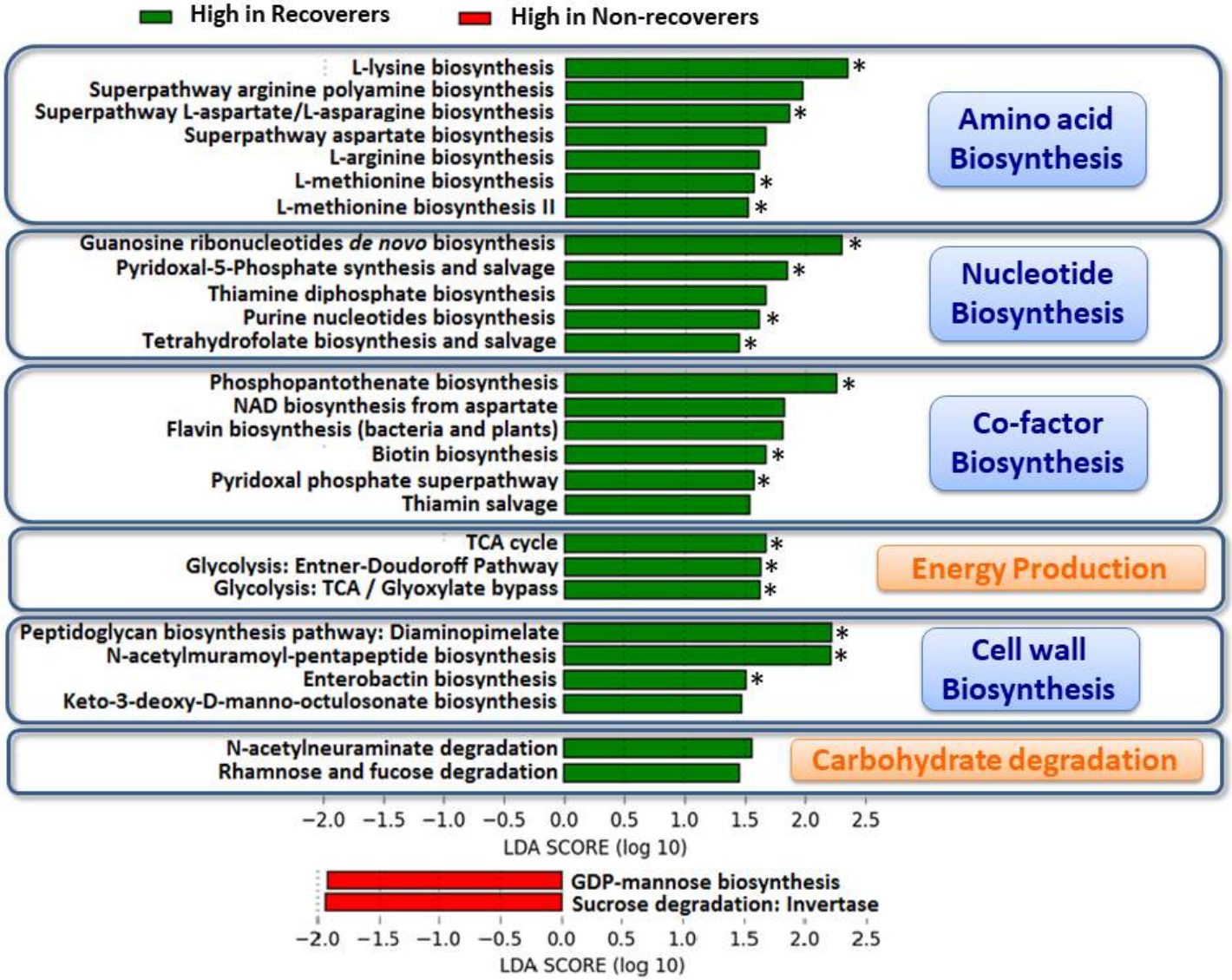

(F)
Peptidoglycan Degradation
5
2
1
0 3 67 10 13 16 19 22
Days since start of experiment



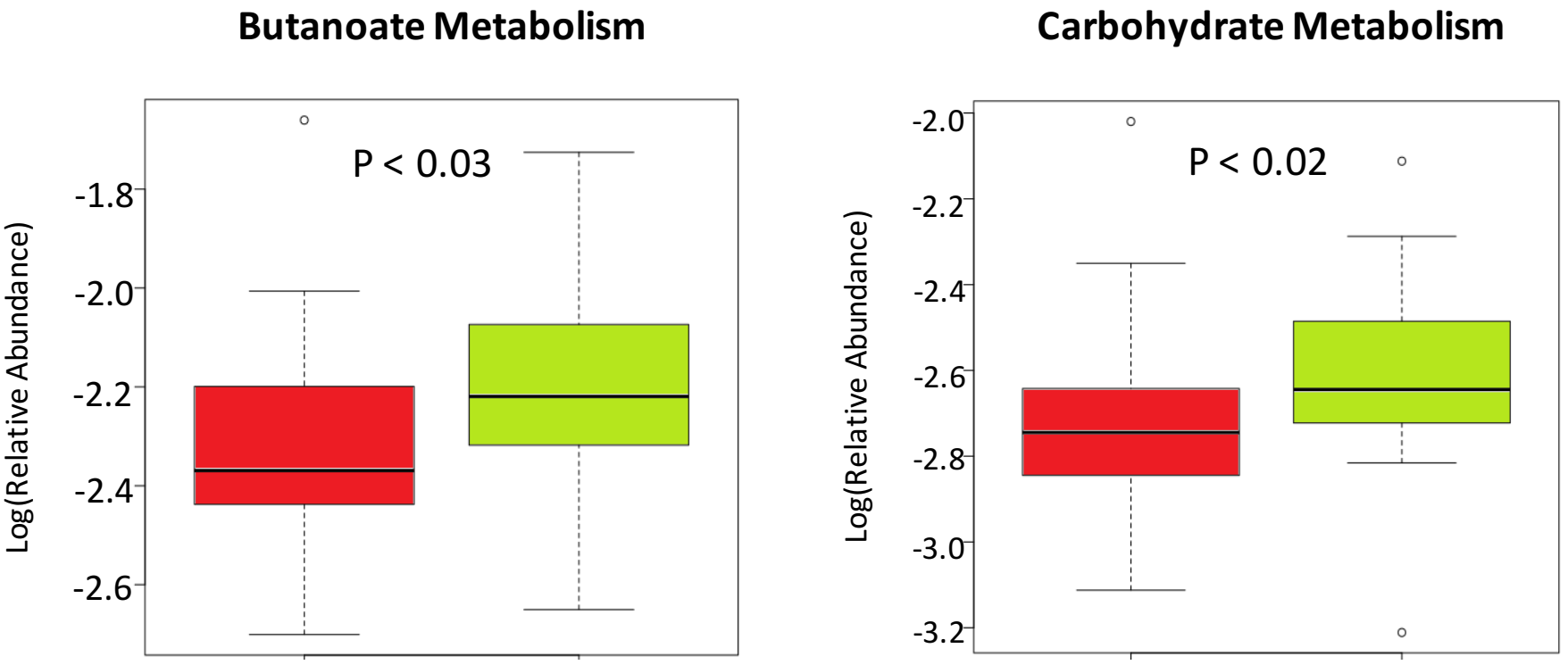
Supplementary Figure S1: Properties of microbiome recovery across cohorts. (A) Cumulative density function for Simpson diversity in the CA and SG cohorts, highlighting the large number of low diversity samples. (B) Microbiomes of recoverers are more similar to control microbiomes than for non-recoverers. Jensen-Shannon (JS) divergence and Jaccard distances for each sample were computed in comparison to the untreated ("control") microbiomes in each cohort. The figures show the median values for each sample in the form of a boxplot. Boxplot whiskers represent 1.5× interquartile range or the maximum/minimum data point within the range.



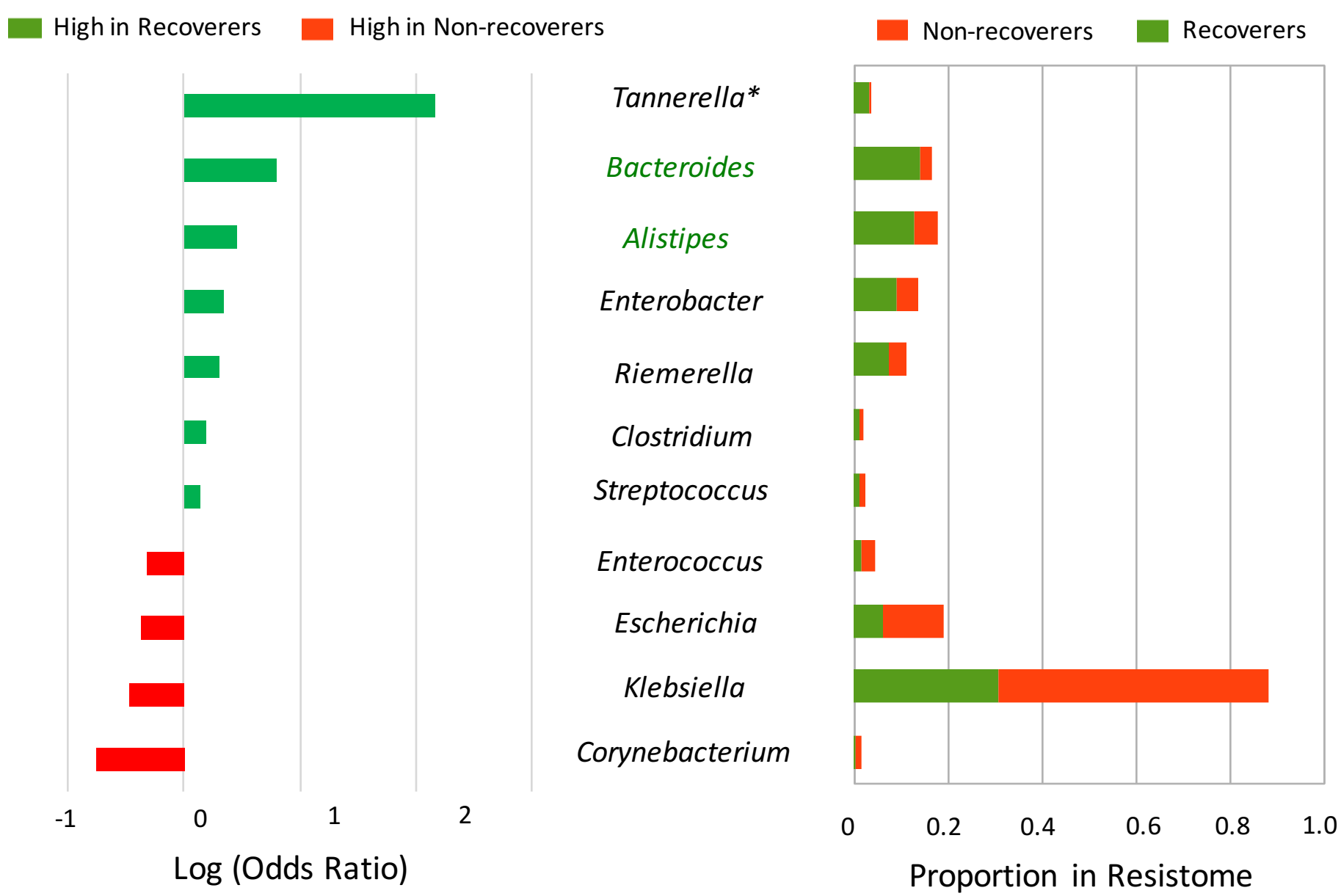
Supplementary Figure S2: Enrichment of RABs during different stages of antibiotic treatment. Fold change was computed for median abundance in recovered vs non-recovered subjects per cohort and averaged across all 4 cohorts. Groups were determined manually (due to limited dimensionality) based on approximate trends and taxonomic similarity. The symbols "*", "**" and "***" indicate p-values <0.1, <0.05 and <0.01, respectively based on Wilcoxon test comparison between recoverers and non-recoverers.



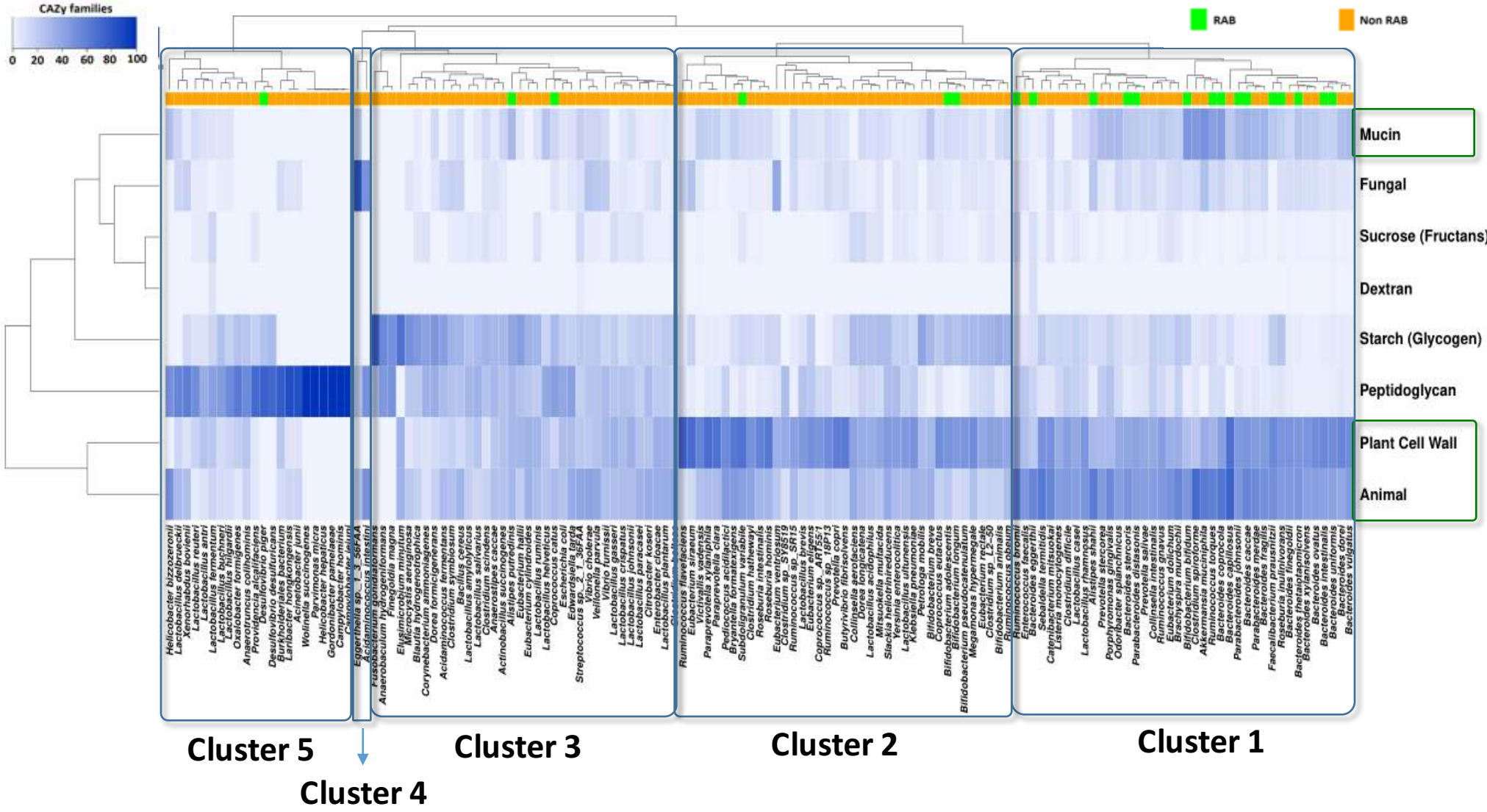
Supplementary Figure S3: **Differentially abundant metagenomic functions in post-antibiotic recovery.** Functional pathways enriched in the gut microbiomes of recoverers or non-recoverers (of the SG cohort) in the 'Pre/Early' and 'During' stages of antibiotic treatment. Note that a star ('*') indicates those pathways for which significant differences were also obtained in the CA cohort. Pathways were grouped into those important for energy production (in orange) and those involved in biosynthesis (in blue), highlighting the role of these two processes in microbiome recovery.



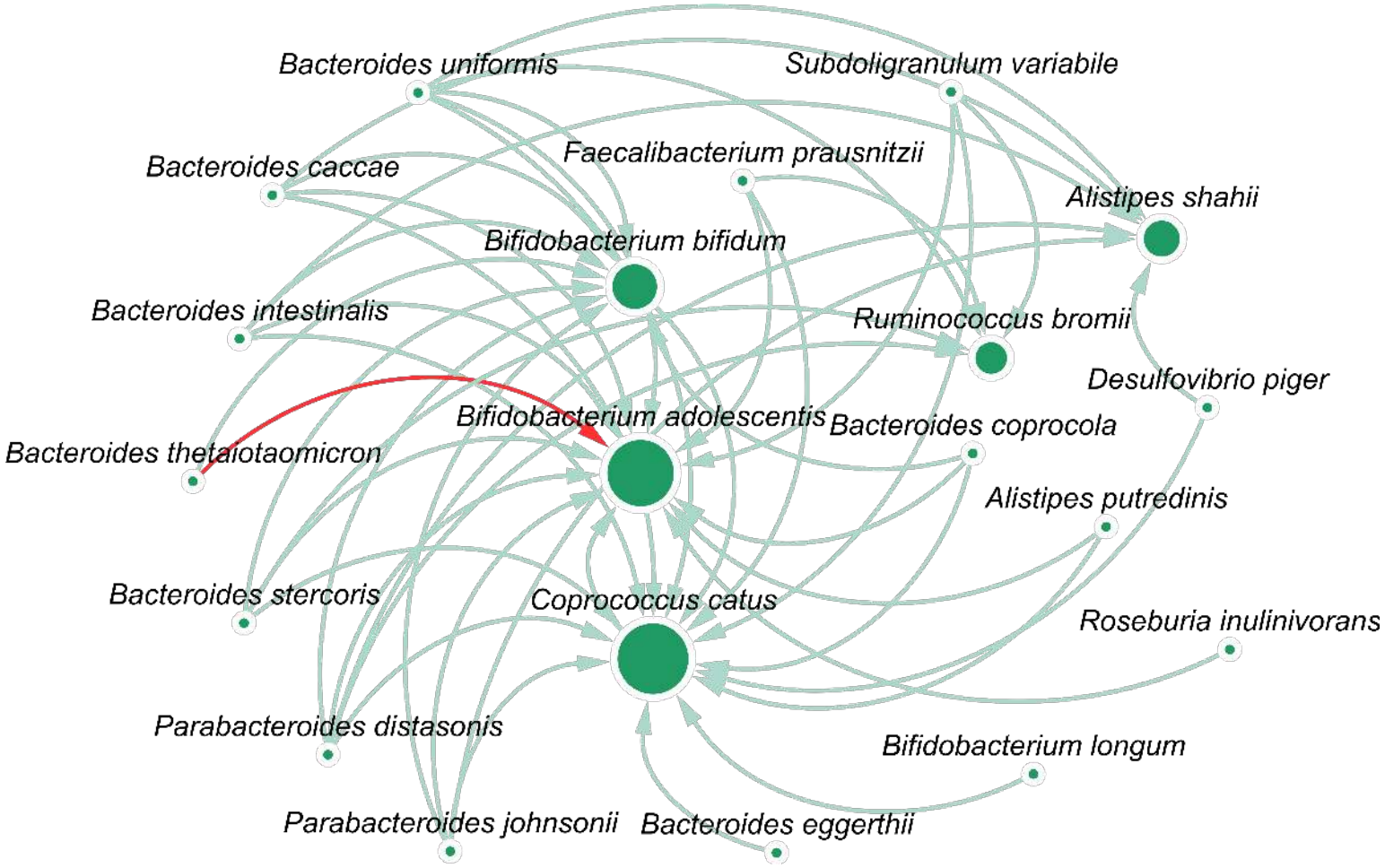
Supplementary Figure S4: Enrichment of Carbohydrate Metabolism and Butanoate Metabolism pathways in the gut microbiomes of recoverers in the EN and SW cohorts: Abundances of the various pathways in the Pre/Early and During stages of treatment were inferred using PICRUSt and then compared among the recoverers and non-recoverers in these cohorts.



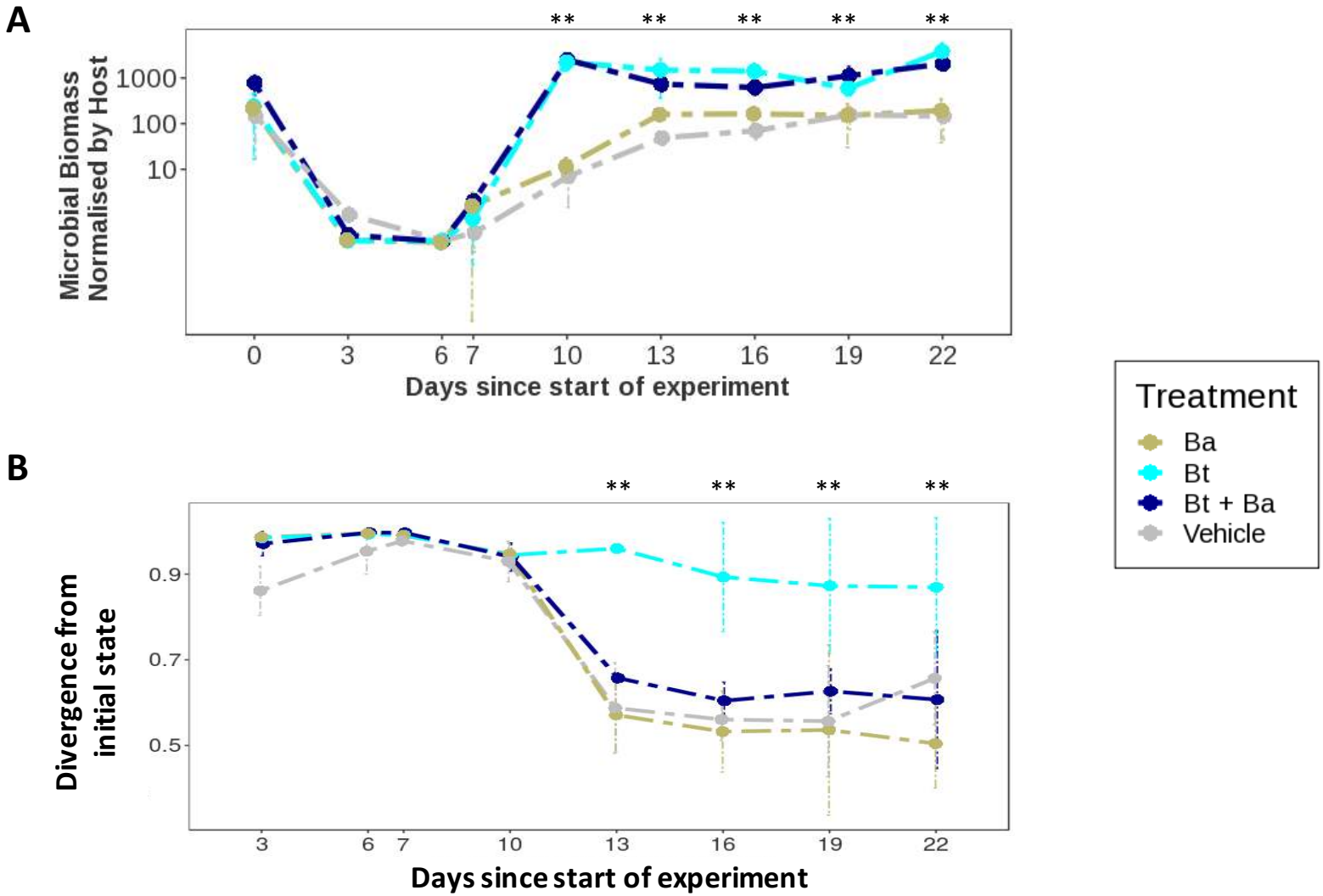
Supplementary Figure S5: Enrichment of Bacterial Genera in the Resistome. Reads belonging to the resistome were assigned to bacterial genera using Kraken (right panel) and odds ratio between groups computed to identify enriched genera (left panel; * = χ^2 test p-value < 0.05). Genera with RAB species are highlighted in green.



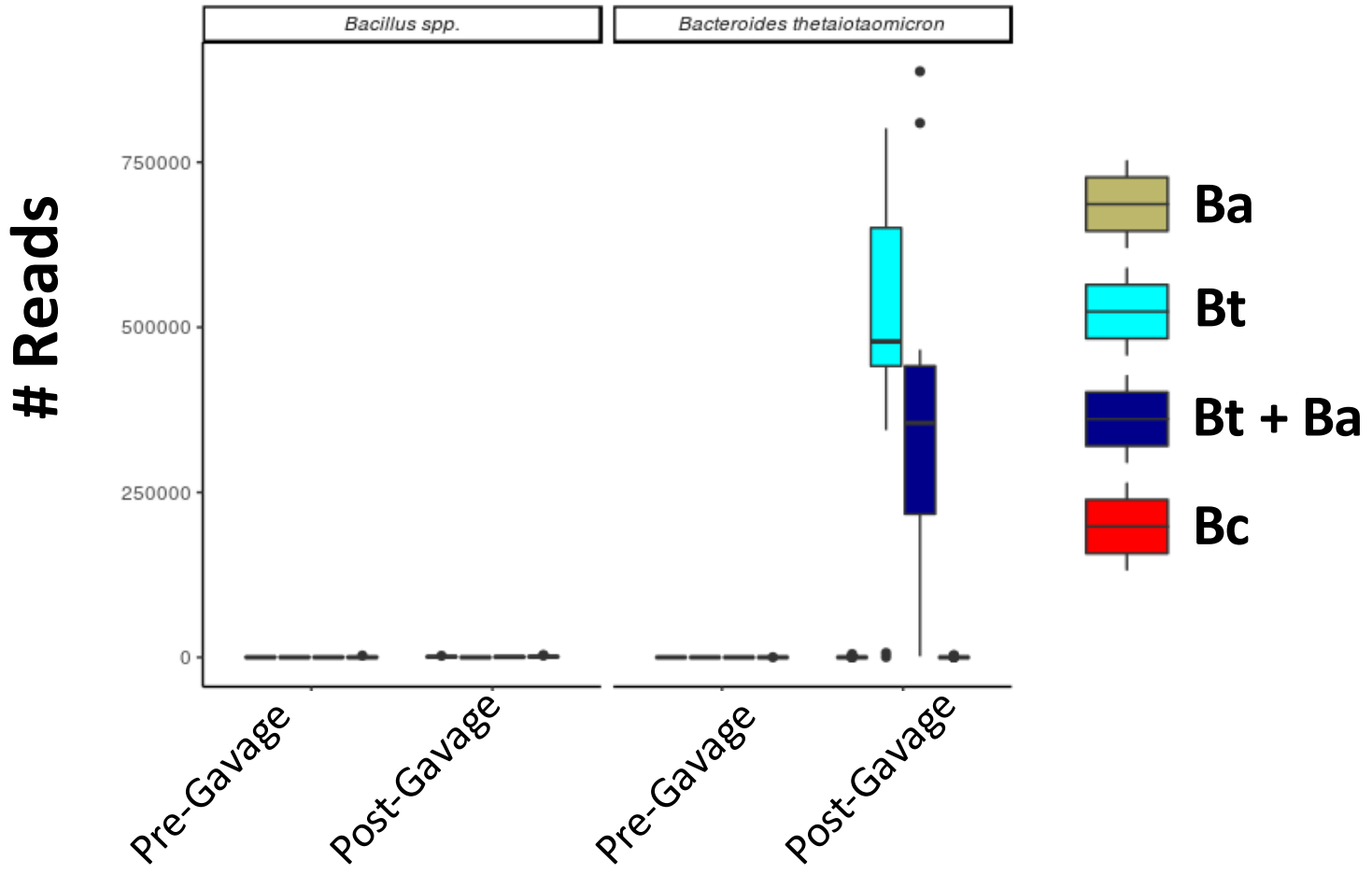
Supplementary Figure S6: RABs have distinct preferences for carbohydrate active enzyme families. Copy numbers of CAZymes having specificities for different categories of carbohydrates were counted for various bacterial species and then range normalized across species. Four distinct large clusters could be observed among the species based on their carbohydrate degradation specificities. RABs were observed to be significantly enriched in cluster 1 (Fisher's exact test) that is defined by an abundance of enzymes that are specific to plant/animal carbohydrates as well as mucin.



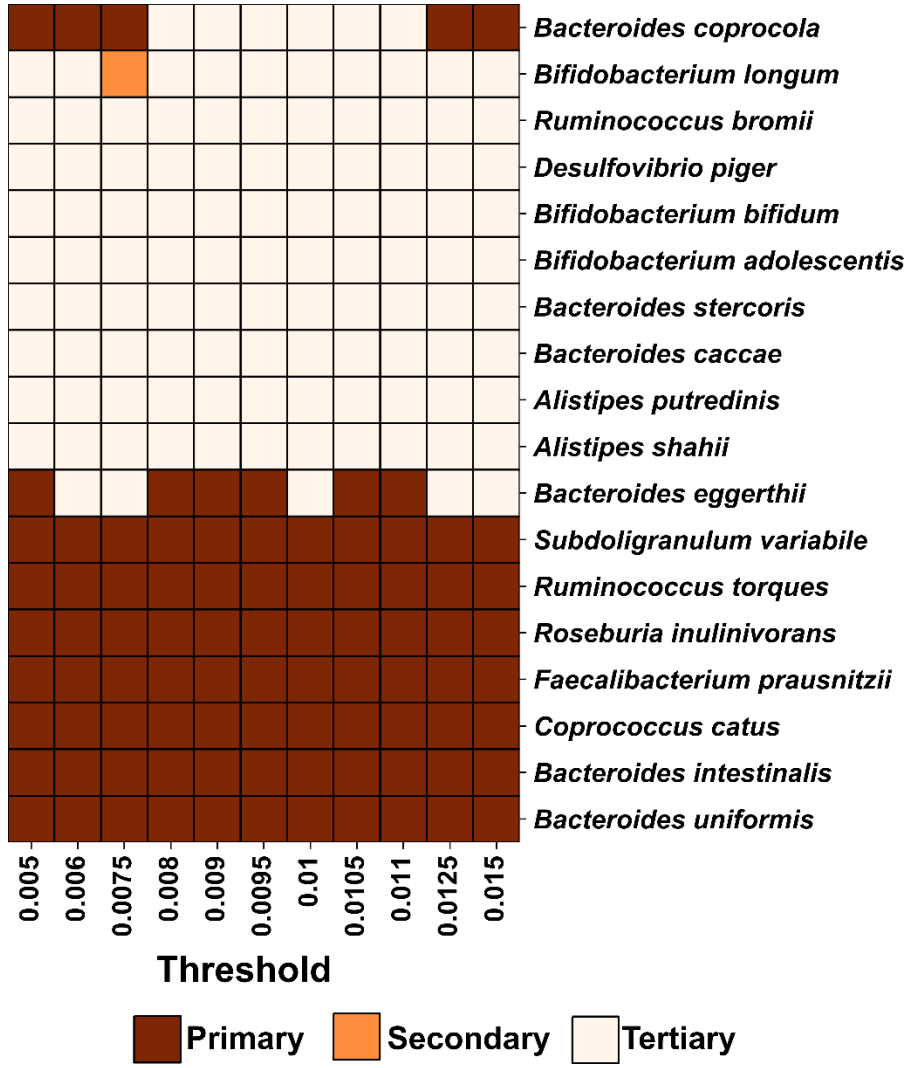
Supplementary Figure S7: Key metabolic interactions between RABs. Directed lines indicate RAB species with high metabolic support to other RAB species (top 10% of MSI values). Node sizes reflect the number of incoming edges and the red edge marks the interaction between *B. thetaiotamicron* and *B. adolescentis* which was evaluated further in an *in vivo* model for microbiome recovery.



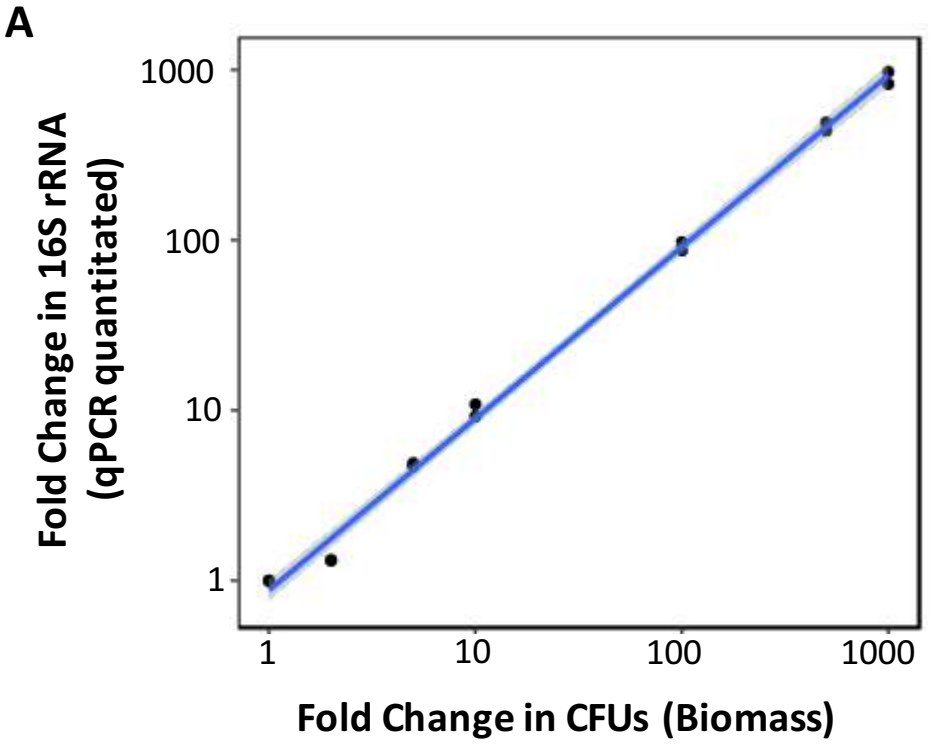
Supplementary Figure S8: Microbiome recovery profiles across treatment groups. (A) Microbial biomass values obtained after normalizing by host reads reveal similar trajectories as plant normalized values (**Figure 5B**). Stars ('**') indicate timepoints where the Bt and Bt+Ba groups were significantly different from other groups (one-sided Wilcoxon test p-value < 0.01). (B) Median Bray-Curtis distance of species level taxonomic profiles compared to day 0 profiles, in different treatment groups and across time (median ± 1 MAD). Stars ('**') indicate timepoints where the Bt group was significantly different from other groups (one-sided Wilcoxon test p-value < 0.01).



Supplementary Figure S9: Successful colonization of *B. thetaiotaomicron* in the mouse gut microbiome post gavage. Boxplots showing high number of *B. thetaiotaomicron* metagenomic reads from mouse stool after Bt gavage, but not *Bacillus spp.* reads after *Bacillus* gavage (Bc group), indicating successful colonization specific to Bt. Whiskers represent 1.5× interquartile range.



Supplementary Figure S10: Placement of RABs in the food web at different thresholds. Heatmap showing that at different thresholds (±50% from the threshold of 0.01 used for results in **Figure 3A**), the position of RABs as primary, secondary and tertiary species in the food-web is retained.



Supplementary Figure S11: Establishing validity of microbial biomass estimation using host normalised microbial read counts. (A) 16S rRNA qPCR demonstrates that the fold change in 16S rRNA copies is directly proportional to fold change in microbial biomass (CFUs). (B) Metagenomic analysis demonstrate that the fold change in host-normalised microbial reads is directly proportional to fold change in microbial biomass (CFUs). *Corynebacterium tuberculostearicum* and *Klebsiella pneumonia* DNA were mixed in equal CFU ratio, and spiked at different biomass amounts into multiple mouse stool DNA samples (Day 0) to achieve the varying fold difference. Data shown for two mouse stool samples (biological replicates).

Barcode adapter, double stranded	1st strand: 5'P-GATCGGAAGAGCACACGTCT 2nd strand: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCT
PE 1.0	5’AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATC* T
Index Primer	5’CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTG GAGTTCAGACGTGTGCTCTTCCGATC*T
16S Forward	5’ACTCCTACGGGAGGCAGC
16S Reverse	5’TTACCGCGGCTGCTGGCAC
gBLOCK:	5’GGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTC GGCAATGGACGGAAGTCTGACCGAGCAACGCCGCGTGAG TGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTAAGAGAA GAACGAGTGTGAGAGTGGAAAGTTCACACTGTGACGGTAT CTTACCAGAAAGGGACGGCTAACTACGTGCCAGCAGCCGC GGTAATACGTAGGTCCCGAG

Supplementary Table S1: Primers and adapter sequences used in this study.