Aus dem Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie
Institut der Ludwig-Maximilians-Universität München
Lehrstuhl für Medizinische Mikrobiologie und Krankenhaushygiene
Vorstand: Prof. Dr. med. Sebastian Suerbaum

Computational methods for tracking the evolution of complex bacterial communities

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Philipp Christian Münch

aus

Coburg

Jahr

2022

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Betreuerin:	Prof. Dr. rer. nat Barbara Stecher-Letsch
Zweitgutachter(in):	Prof. Dr. Alexander Sczyrba
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	04. Oktober 2022

Münch, Philipp Christian

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Titel *Computational methods for tracking the evolution of complex bacterial communities* selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, den 18 November 2022 Philipp Christian Münch

Contents

Publication list vii
1. Publications for the cumulative thesis vii
2. Other publications not part of the cumulative thesis vii

Abstract xi

Einleitung xiii

Contributions to publications xv
1. Summary and own contributions to Publication I xv
2. Summary and own contributions to Publication II xvii
3. Summary and own contributions to Publication III xix

Introduction xxi
1. Scaling laws highlight the need for bioinformatics xxi
2. Computational methods for quantifying evolution xxii
3. Metagenomics provides a more unbiased view for studying evolutionary propertiesxxiii
4. Virus–bacteria coevolution can be quantified by d_N/d_S statistics xxiv
5. CRISPR–Cas immunity is an example of adaptive evolution xxvi
6. Antibiotics and "Red Queen" dynamics xxvii
7. The gastrointestinal system is a melting pot of coevolutionary interactions . . xxix

List of Figures xxxi

Bibliography xxxiii

Appendix A: Unpublished Manuscript III xxxix

Acknowledgement xliii

Curriculum vitae xlv

Publication list

1. Publications for the cumulative thesis

- P. C. Münch, B. Stecher and A. C. McHardy. EDEN: evolutionary dynamics within environments. Bioinformatics (**Manuscript I**)
- P. C. Münch, E. A. Franzosa, B. Stecher, A. C. McHardy and C. Huttenhower. *Identification of Natural CRISPR Systems and Targets in the Human Microbiome*. Cell Host & Microbe (**Manuscript II**)
- P. C. Münch, C. Eberl, D. Ring, S. Wölfel, A. Fritz, S. Herp, R. Geffers, E. A. Franzosa, C. Huttenhower, A. C. McHardy and B. Stecher. *Pulsed antibiotic treatments of gnotobiotic mice manifests in complex community dynamics and resilience effects*. (**Unpublished Manuscript III**)

2. Other publications not part of the cumulative thesis

- A. R. Ghazi, P. C. Münch, D. Chen, J. Jensen and C. Huttenhower, *Strain identification and quantitative analysis in microbial communities*. **Journal of Molecular Biology (in review)**
- Z-L. Deng, P. C. Münch, R. Mreches, A. C. McHardy, *Rapid and accurate identification of ribosomal RNA sequences via deep learning*. **Nucleic Acids Research (in review)**
- H. A. Gündüz, R. Mreches, P. C. Münch, A. C. McHardy, B. Bischl, M. Binder and M. Rezaei, *Self-GenomeNet: Self-supervised Learning with Reverse-Complement Context Prediction for Nucleotide-level Genomics Data* **ICLR 2022 Conference (in review)**
- B. Stecher, A. Weiss, A. D. Raj, A. Burrichter, A. von Strempel, P. C. Münch, C. Meng, K. Kleigrewe, Wolfgang Eisenreich, C. Huber, L. Roessler, K. Jung, Stephanie Goeing, Lara Jochum, A. Sanchez, C. Lincetto, J. Hübner, G. Marinos, J. Zimmermann and C. Kaleta, *In vitro interaction network of a synthetic gut bacterial community*. **The ISME Journal**
- F. Goeser, P. C. Münch, T. R. Lesker, P. L. Lutz, B. Krämer, D. J. Kaczmarek, C. Finnemann, H. D. Nischalke, R. Geffers, M. Parcina, A. C. McHardy, C. Strassburg, A. Hoerauf, J. Nattermann, I. Bekeredjian-Ding and U. Spengler, *Neither black nor white: do altered intestinal microbiota reflect chronic liver disease severity?*. **Gut (PostScript Letter)**

- S. Spriewald, E. Stadler, B. A. Hense, P. C. Münch, A. C. McHardy, A. S. Weiss, N. Obeng, J. Müller, and B. Stecher, *Evolutionary Stabilization of Cooperative Toxin Production through a Bacterium-Plasmid-Phage Interplay*. **Mbio**

- C. Eberl, D. Ring, P. C. Münch, M. Beutler, M. Basic, E. C. Slack, M. Schwarzer, D. Srutkova, A. Lange, J. S. Frick, A. Bleich and B. Stecher, *Reproducible colonization of germ-free mice with the oligo-mouse-microbiota in different animal facilities*. **Frontiers in Microbiology**

- Z-L. Deng, A. Dhingra, A. Fritz, J. Götting, P. C. Münch, L. Steinbrück, T. F. Schulz, T: Ganzenmüller and A. C. McHardy, *Evaluating assembly and variant calling software for strain-resolved analysis of large DNA viruses*. **Briefings in Bioinformatics**

- E. Asgari, P. C. Münch, T. R. Lesker, A. C. McHardy and M. R. K. Mofrad, *DiTaxa: nucleotide-pair encoding of 16S rRNA for host phenotype and biomarker detection*. **Bioinformatics**

- T. Vatanen, D. R. Plichta, J. Somani, P. C. Münch, T. D Arthur, A. B. Hall, S. Rudolf, E. J. Oakeley, X. Ke, R. A. Young, H. J. Haiser, R. Kolde, M. Yassour, K. Luopajärvi, H: Siljander, S. M. Virtanen, J: Ilonen, R. Uibo, V: Tillmann, S- Mokurov, N. Dorshakova, J. A. Porter, A. C. McHardy, H. Lähdesmäki, H. Vlamakis, C. Huttenhower, M. Knip and R. J. Xavier, *Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life*. **Nature Microbiology**

- S. Duggimpudi, A. Kloetgen, S. K. Maney, P. C. Münch, K. Hezaveh, H. Shaykhalishahi, W. Hoyer, A. C. McHardy, P. A. Lang, A. Borkhardt and J. I. Hoell, *Transcriptome-wide analysis uncovers the targets of the RNA-binding protein MSI2 and effects of MSI2's RNA-binding activity on IL-6 signaling*. **Journal of Biological Chemistry**

- S. Krause, A. Bremges, P. C. Münch, A. C. McHardy and J. Gescher, *Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms*. **Scientific Reports**

- S. Brugiroux, M. Beutler, C. Pfann, D. Garzetti, H.-J. Ruscheweyh, D. Ring, M. Diehl, S. Herp, Y. Lötscher, S. Hussain, B. Bunk, R. Pukall, D. H. Huson, P. C. Münch, A. C. McHardy, K. D. McCoy, A. J. Macpherson, A. Loy, T. Clavel, D. Berry and B. Stecher, *Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium*. **Nature Microbiology**

- S. Hacquard, B. Kracher, K. Hiruma, P. C. Münch, R. Garrido-Oter, M. R. Thon, A. Weimann, U. Damm, J.-F. Dallery, M. Hainaut, B. Henrissat, O. Lespinet, Soledad Sacristán, E. V. L. Van Themaat, E. Kemen, A. C. McHardy, P. Schulze-Lefert and R. J. O'connell, *Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi*. **Nature communications**

- Y. Bai, D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, P. C. Münch, S. Spaepen, M. Remus-Emsermann, B. Hüttel, A. C. McHardy, J. A. Vorholt

and P. Schulze-Lefert, *Functional overlap of the Arabidopsis leaf and root microbiota*. **Nature**

- D. Bulgarelli, R. Garrido-Oter, P. C. Münch, A. Weiman, J. Dröge, Y. Pan, A. C. McHardy, P. Schulze-Lefert *Structure and function of the bacterial root microbiota in wild and domesticated barley*. **Cell Host & Microbe**

- A. Kloetgen, P. C. Münch, A. Borkhardt, J. I. Hoell and A. C. McHardy, *Biochemical and bioinformatic methods for elucidating the role of RNA–protein interactions in posttranscriptional regulation*. **Briefings in functional genomics**

Abstract

The focus of my PhD thesis was to study evolutionary aspects of host-associated microbial communities. In order to better understand these effects, I developed and applied computational methods to search for protein families that are under selection within metagenomes (**Publication I**) and applied them to various environments (see the articles "*Structure and function of the bacterial root microbiota in wild and domesticated barley*" (Bulgarelli et al. 2015), and "*Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi*" (Hacquard et al. 2016). One consistent finding of these studies was the high selective pressure acting on gene families associated with the bacterial defense system (the so-called CRISPR–Cas system) and families annotated as being related to bacteriophages. To study this CRISPR–phage relationship more closely, I systematically analysed CRISPR cassettes and CRISPR-related genes in samples from the Human Microbiome project (HMP) (**Publication II**). This resulted in one of the most comprehensive CRISPR collections to date. Further, we found novel sequence characteristics in the CRISPR loci and described the differences in the composition of CRISPR-associated genes in different body habitats and a potential relationship between the CRISPR defence system and the restriction modification system of bacteria. Furthermore, I performed a similar but less extensive search on metagenomic samples from infants: "*Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life*" (Vatanen et al. 2019). Next, I turned my focus to study microbiome evolution in a gnotobiotic mouse model, since this provided the opportunity to study bacteria evolution on an intermediate scale of complexity. I contributed to the development of the mouse model described in the Manuscript "*Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium*" (Brugiroux et al. 2016) and a study comparing the stability of the community between animal facilities "*Reproducible colonization of germ-free mice with the Oligo-Mouse-Microbiota in different animal facilities*" (Eberl, Ring, et al. 2019) and to a study focusing on the interaction network of this community (Weiss et al. 2021). However, my main work with the OMM^{12} model has been to study community effects and evolution during repeated rounds of AB exposure (unpublished **Publication III**).

Einleitung

Der Schwerpunkt meiner Doktorarbeit lag auf der Untersuchung evolutionärer Aspekte von wirtsassoziierten mikrobiellen Gemeinschaften. Um diese Effekte besser zu verstehen, habe ich computergestützte Methoden entwickelt und angewandt, um nach Proteinfamilien zu suchen, die in Metagenomen (**Publikation I**) der Selektion unterliegen, und sie auf verschiedene Umgebungen angewandt (siehe die Artikel "*Structure and function of the bacterial root microbiota in wild and domesticated barley*" (Bulgarelli et al. 2015) und "*Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi*" (Hacquard et al. 2016). Ein durchgängiges Ergebnis dieser Studien war der hohe Selektionsdruck, der auf Genfamilien wirkt, die mit dem bakteriellen Abwehrsystem (dem sogenannten CRISPR-Cas-System) und Familien, die als mit Bakteriophagen verwandt beschrieben werden, verbunden sind. Um diese CRISPR-Phagen-Beziehung genauer zu untersuchen, analysierte ich systematisch CRISPR-Kassetten und CRISPR-verwandte Gene in Proben aus dem Human Microbiome Project (HMP) (**Publikation II**). Dies führte zu einer der bisher umfassendsten CRISPR-Sammlungen. Darüber hinaus fanden wir neuartige Sequenzmerkmale in den CRISPR-Loci und beschrieben die Unterschiede in der Zusammensetzung von CRISPR-assoziierten Genen in verschiedenen Körperregionen sowie eine mögliche Beziehung zwischen dem CRISPR-Abwehrsystem und dem Restriktionsmodifikationssystem von Bakterien. Außerdem habe ich eine ähnliche, aber weniger umfangreiche Suche an metagenomischen Proben von Säuglingen durchgeführt: "*Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life*" (Vatanen et al. 2019). Als Nächstes konzentrierte ich mich auf die Untersuchung der Mikrobiomevolution in einem gnotobiotischen Mausmodell, da dies die Möglichkeit bot, die Evolution von Bakterien auf einer mittleren Komplexitätsebene zu untersuchen. Ich war an der Entwicklung des Mausmodells beteiligt, das im Manuskript "*Genom-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium*" (Brugiroux et al. 2016) und eine Studie zum Vergleich der Stabilität der OMM^{12} Gemeinschaft zwischen Tierhaltungsanlagen "*Reproducible colonization of germ-free mice with the Oligo-Mouse-Microbiota in different animal facilities*" (Eberl, Ring, et al. 2019) sowie eine Studie, die sich auf das Interaktionsnetzwerk dieser Gemeinschaft konzentriert (Weiss et al. 2021). Meine Hauptarbeit mit dem OMM^{12}-Modell bestand jedoch darin, die Auswirkungen und die Entwicklung der Gemeinschaft während wiederholter AB-Expositionen zu untersuchen (unveröffentlichtes **Manuscript III**).

Contributions to publications

1. Summary and own contributions to Publication I

In the manuscript "**EDEN: evolutionary dynamics within environments**", published in Bioinformatics (Authors: P. C. Münch, B. Stecher and A. C. McHardy), I described a method that makes use of the d_N/d_S statistic to detect protein families that are under positive and negative selection in metagenomic samples and applied this method to a collection of such samples. This method overcomes the current limitations of phylogenetic methods for studying adaptation and evolutionary dynamics that are not able to cope with Gb-sized metagenome data. To my knowledge, this is the first software package designed for the detection of protein families and the regions thereof that are under positive selection within metagenomic datasets, as well as their associated biological processes. In a further publication (not part of this cumulative thesis), EDEN generated evidence for a higher degree of selection acting on protein families within the root-associated microbiota than on those found in bulk soil, especially protein families linked to pathogenesis, bacteria–phage interactions, secretion and nutrient mobilization (Bulgarelli et al. 2015). In a further publication (not part of this cumulative thesis), EDEN was used to compare the selection patterns of strains of *Colletotrichum* with different lifestyles (Hacquard et al. 2016). We furthermore analysed human metagenome samples from different body sites of healthy individuals, as well as individuals with different body mass indexes, thus providing an initial quantitative analysis of the traces of directional selection for different types of human-associated metagenomic samples (Münch, Stecher, and McHardy 2017). EDEN comes in a Docker installation, allowing its use under all common operating systems. Via a web-browser interface, the user can upload samples and gene family definitions in the form of a set of profile Hidden Markov models (pHMM). These samples are then analysed by fast methods for searching pHMM hits in the sample(s), multiple sequence alignment, tree inference and substitution counting. Finally, the tool performs statistical tests to rank these gene families according to the estimated evolutionary pressure and generates publication reads figures showing (i) the distribution of d_N/d_S rates among and between the samples, and (ii) the distribution of d_N/d_S over the alignment of each protein family, e.g. to identify epitopes under evolutionary pressure.

In this publication, I conceived and designed the experiments together with BS and ACM; I implemented, tested and benchmarked the software; I generated all the figures and applied the method to the dataset of the Human Microbiome Project (HMP) and other datasets and analysed the results; ACM and I wrote the manuscript with comments from BS. Further, this work contributed to other publications (Bulgarelli et al. 2015; Hacquard et al. 2016) (not part of this cumulative thesis).

2. Summary and own contributions to Publication II

In the manuscript "**Identification of natural CRISPR systems and targets in the human microbiome**" published in Cell Host & Microbe (Authors: P. C. Münch, E. A. Franzosa, B. Stecher, A. C. McHardy and C. Huttenhower), I carried out a comprehensive taxonomic and functional characterisation of natural CRISPR systems in over 2,000 metagenomes from the expanded human microbiome project (HMP1-II) and identified over 2.9 million novel CRISPR spacer sequences within the metagenomes native to oral, skin, vaginal and gut body sites. For comparison, this is over 20 times greater than the number of CRISPR systems cataloged from complete microbial genomes in CRISPRCasdb (Pourcel et al. 2020). Furthermore, I performed a metagenome-based quantification of *cas* genes, which, together with the CRISPR detection, provides information on both the potential functional roles of CRISPR–Cas systems and their targets during human health, and the evolutionary properties and principles of bacteria–virus relationships. Based on this data resource, I estimated the distribution of CRISPR sequence variants in different body habitats, identifying oral plaque and the tongue as the sites with significantly higher CRISPR densities than gut and urogenital sites. I characterised the nucleotide sequence properties of spacer and repeat regions, identifying a hitherto-unknown palindromic nucleotide distribution pattern and an association between GC content and spacer length. This appears to be newly detected in the human microbiome, in contrast to earlier studies on much smaller populations or on environmental microbial communities. We quantified the functional potential of these human-associated CRISPR spacers, including their ability to target bacteriophages. In addition to CRISPR spacers that appear to target phages, we found potential CRISPR influence on bacterial genes involved in methylation activity, suggesting a yet unknown connection between the restriction modification and CRISPR defence systems. The spacer and repeat set described in this study is, to our knowledge, the largest and most comprehensive comparison of CRISPR carriage and ecology in the human microbiome. As a resource, it can aid in the identification of viral sequences associated with the human microbiome, which remains technically challenging, and for the development of a novel CRISPR detection methodology. I believe that this work advances our understanding of the CRISPR system, and that the results may be further applied for the optimization of plasmid and phage-based therapies.

In this publication, I conceived and planned the experiments together with EAF, ACM and CH; I performed the CRISPR cassette and *cas* gene search and analysed the data; I generated all the figures and tables and wrote the manuscript with input from EAF, ACM, CH and BS.

3. Summary and own contributions to Publication III

In the unpublished manuscript "**Pulsed antibiotic treatments of gnotobiotic mice manifests in complex community dynamics and resistance effects**" prepared for Nature Microbiology, (Authors: P. C. Münch, C. Eberl, D. Ring, S. Wölfel, A. Fritz, S. Herp, R. Geffers, E. A. Franzosa, C. Huttenhower, A. C. McHardy and B. Stecher), I performed an evolutionary analysis of a defined bacterial consortium consisting of 12 cultured strains, the Oligo-Mouse Microbiota (OMM^{12}) (Brugiroux et al. 2016) in mice, during repeated antibiotic exposures to three different clinically important antibiotic compounds. Despite the known impact of antibiotics on the human microbiome, only a few studies have been performed to understand the short- and long-term impact of antibiotic treatment on the human microbiome, and this work aimed to address one of the biggest problems facing the public health sector: understanding the evolutionary effects on bacterial environments during antibiotic (AB) stress, and evolution towards AB tolerance or resistance. In complex microbial populations such as the human gut, understanding the temporal emergence and dynamics of antibiotic-resistant mutants are difficult to obtain because of the complexity of the microorganisms present. Although pure *in vitro* studies on cultured sequenced isolates are valuable, some evolutionary drivers such as mutation load and fitness costs differ between *in vitro* and *in vivo* studies, which makes it necessary to study these systems as a whole. In this work, I analysed how the community responds to repeated exposure to antibiotics in the short and long term, and studied the community dynamics during repeated AB application over 80 days (in five treatment phases interspaced with resilience phases). Interestingly, some species, such as *E. faecalis* in the ciprofloxacin group, increased tolerance to AB within the duration of the study, whereas other community members, such as *A. muciniphila*, exhibited a more complex response pattern to AB. To identify these observations with genomic alterations, I quantified the differences between samples taken throughout the study on the single nucleotide (SNP) level and accounted for a range of OMM^{12}-specific characteristics that could lead to false positive results in the variant calling, such as the high intra-species similarity of OMM^{12} members or duplication events. Furthermore, I confirmed that many of the remaining SNPs were indeed located in genes with a known relationship to antibiotic resistance, such as 30S ribosomal protein S11 and efflux pumps. Since the response pattern of some bacteria remains complex and could not be explained by the SNP counts, I analysed other putative evolutionary drivers such as bacterial growth. Here, I implemented existing methods to estimate growth via the Peak-to-Trough (PT2) ratio based on the genomic coverage by accounting for OMM^{12}-specific characteristics such as local variations in the coverage pattern. I observed that the growth rate changed during the study time and was often linked to treatment regimes. Finally, I analysed prophage activation by identifying regions with high similarity to prophage genes and loci with an abnormally high read coverage on the reference genome, since these are likely to be the products of prophage activation. I analysed these regions at all time points and uncovered a treatment-dependent relationship. Again, these regions contain genes that are potentially associated with AB resistance. Finally, I developed a R package that applies non-negative matrix factorization to the problem of genotype identification. Taken together, this work shows that mice stably colonized with synthetic bacterial communities are a valuable tool for studying community dynamics in response to AB stress. The responses

to AB treatment are multifaceted, and their interaction patterns could be investigated further by the research community with an R package that we aim to publish alongside the dataset.

In this study, I conceived and designed the experiments together with BS and ACM; I wrote the grant applications with input from BS and ACM; I performed the mouse experiments together with DR, CE and SH and BS; I analysed and interpreted the datasets, and wrote the R libraries and functions for P2T estimation and SNP decomposition, and generated all the figures and tables; I wrote the manuscript with comments from BS, ACM, EAF and CH.

Introduction

1. Scaling laws highlight the need for bioinformatics

Moore accurately predicted that the number of transistors would double every two years (Brock and Moore 2006), driven by a series of exponential growth modes driven in turn by single interventions that overcome current bottlenecks. Similar advances are also observable if we compare the improvements in the runtime of bioinformatics software over the past few decades, such as sequence alignment tools (Muir et al. 2016), probably caused by the introduction of parallel architectures. A similar law can be observed in genomics, where the costs of DNA sequencing have dropped significantly over the last few decades, quantified as dollars per base sequence (ibid.). Technical breakthroughs have resulted in a reduction in the cost per megabase from $\sim$ 10,000 USD with Sanger technology (1986) to $\sim$10 USD with 454 technology (2007) to around 10 cents with modern Illumina sequencers (today), reducing costs by a factor of 5–10 per year on average and outpacing Moore's law. This translates to reductions in the growth rate of databases that collect genetic information, such as GenBank, which has a doubling time of approximately 18 months, since this is not solely bound by the sequencing costs but also by data preparation. Comparing these different scaling laws hints that there could be a "biological overhang", where the amount of biological data generated will presumably continue to outpace processing speeds, which would set constraints on runtime complexity. These factors underpin the need for further improvements in computational methods that can be applied to process large amounts of biological data, such as the development of more computationally efficient bioinformatic tools. However, this also highlights that we are in an era with a large amount of unprocessed sequence information that could be harvested to study the functions of humans, animals and microbial communities, and how organisms evolve over time. The presence of this data overhang was my main motivation to study bioinformatics and to focus my Ph.D. research on developing and applying methods and tools that are suitable for large collections of genomic datasets. Recently, algorithms were developed (based on deep neural networks) that seem to exploit the existing computational power far more efficiently than before (Jumper et al. 2021; Brown et al. 2020; Kaplan et al. 2020), which leads to algorithmic efficiency outpacing the gains from hardware efficiency, which motivated me to focus my further research on finding and applying deep learning architectures within the field of genomics (not part of this cumulative thesis).

2. Computational methods for quantifying evolution

According to the projections of currently available studies, the earth should contain $\sim 10^{10}$ – 10^{12} microbial species. Of these, $\sim 10^7$ have been already catalogued; of these, $\sim 10^5$ species are represented in reference genomes and $\sim 10^4$ have been cultured, leaving over $\sim$ 99.99% of microbial taxa to be discovered (Locey and Lennon 2016; Lennon and Locey 2020). At the smallest scale, these microorganisms have evolved progressively through genetic changes at the single nucleotide level, called SNPs (**Fig. 1**). These are the results of mutations, a source of biological variation that can vary in bacteria from 10^{-8} to 10^{-10} per base per generation (Lynch et al. 2016) and gives bacterial populations the ability for long-term survival and to adapt to extreme environmental conditions (Pérez et al. 2017; Merino et al. 2019). Although mutation rates, in combination with Darwinian selection, seem to be one of the main evolutionary drivers, the involvement of further mechanisms such as lateral gene transfer mechanisms, homologous recombination and interactions between hosts, and viruses draw a more complicated portrait of evolutionary forces in the observed genomic content of microbial populations.

Some of these substitutions may be neutral or silent, meaning that a change in a particular base pair in an open reading frame does not translate to a change in the amino acid, and may not cause an erroneous (or evolutionarily fitter) or shortened protein. However, these substitutions should not be considered as completely neutral, since these still change the underlying codons and may result in the formation of codon patterns, since DNA replication may prefer certain codons over others, even if they encode for the same protein. These codon patterns are the result of *translational selection*, which describes how codon usage reflects evolutionary forces acting within an organism (Grantham et al. 1980; Gouy and Gautier 1982). This results in selection for mRNA expressivity, where highly expressed genes are encoded by chemically more optimal codons with respect to their translational accuracy. However, methods for systematically estimating translational selection have been rare until a unified framework was presented (Reis, Savva, and Wernisch 2004). This showed that genome size is a major contributor to this evolutionary force, since it is tightly linked to tRNA copy numbers, and that this force can be observed in organisms when there is a redundancy in the tRNA genes that the evolutionary force can act on. On the other hand, we can reasonably assume that genome size is itself a target for selection processes, since it imposes a higher metabolic burden with reduced fitness on the cell and is associated with the mutation rate (Ochman, Lawrence, and Groisman 2000; Ranea et al. 2005; Marais, Batut, and Daubin 2020).

Other substitutions on open reading frames might cause a change at the amino acid level; these substitutions are known as non-neutral substitutions. Different measures have been developed to quantify the evolutionary pressure of protein-coding sequences, which mainly can be classified into tools that detect the signatures of selection in a population and those that do so in comparative settings. One of the most widely used and reliable measures applied to protein-coding regions is the d_N/d_S ratio, which measures the excess of nucleotide-altering substitutions in a set of genes, which is a general indicator of the mode and strength of selection (Nielsen 2005; Buschiazzo et al. 2012). Substitutions can be either synonymous without causing a change in the coding amino acid (also called silent or neutral, as discussed above) or non-synonymous, causing a replacement of the amino acid. In the absence of

selection pressure, synonymous and non-synonymous mutations should occur at the same rate ($d_N/d_S = 1$), although evidence for positive or negative selection appears when the ratio is > 1 or < 1, respectively. Synonymous substitutions are often quantified as substitutions per synonymous sites d_S (where the number of synonymous sites is a measure of the likelihood that substitution at a given position in the codon is synonymous, typically occurring in the third position of codons), are usually regarded as neutral and are not affected by natural selection. Conversely, non-synonymous changes per non-synonymous site (again, sites indicate the likelihood that a mutation will cause a change in the amino acid, typically in the first and second position of a codon), denoted d_N, are seen as influencing the bacterial fitness and will be removed from the population via purifying or negative selection (DeSalle 2000). A comparison of the relative frequency of both d_N and d_S can be used as a measure of the selective forces that act on a given region, and can be performed at different levels such as sites, codons, genomes and populations (Rocha et al. 2006). As part of my dissertation I have extended and applied this d_N/d_S measure to metagenomic samples for quantifying selection pressures acting on protein families and the putative epitopes thereof (**Publication 1**).

3. Metagenomics provides a more unbiased view for studying evolutionary properties

Albeit valuable, previous types of studies on microbial evolution used either single bacterial cultures *in vitro*, such as the long-term evolution experiments by Lenski and Travisano (1994) or were of a comparative nature, analysing a specific set of genomes (Petersen et al. 2007) or all genomes present in public databases (Endo, Ikeo, and Gojobori 1996; Ciufo et al. 2018; Jain et al. 2018). These methods are arguably limited. Although the first type of study misses out potential microbe–host interactions, the latter suffers from cultivation and observation bias and has led to a wide range of controversies; for example, if this cultivation bias should be accounted for when interpreting species boundaries (Murray, Gao, and Wu 2021; Rodriguez-R et al. 2021). However, with metagenomic techniques, one can obtain near-complete genome sequences from environmental samples (West et al. 2007), giving access to Gb-sized datasets of microbial communities under natural conditions without any cultivation bias. With metagenomics, the sequenced nucleotide content typically originates from different cells in the sample and does not show any strong biases regarding the part of the genome or the strain they represent, thus providing random representations of the naturally occurring diversity (Amann, Ludwig, and Schleifer 1995). These methods overcome the limitations of the cultivation process (the cultivation bottleneck) of single isolates. These metagenomic techniques helped to clarify the aforementioned debate on the presence of species boundaries by applying distance measures to genomes reconstructed from metagenomes instead of genomes isolated from different populations and habitats, giving evidence of the presence of a species boundary (Caro-Quintero and Konstantinidis 2012).

Several technical challenges and limitations must be faced when handling metagenomic shotgun sequencing datasets. One of them is correct selection of the sequencing depth, defined by the quantity of reads obtained from a sample. The chosen depth needs to be high enough

to assemble the genomes of less abundant community members of interest (Kunin et al. 2008). This is especially a problem in highly complex environments that are dominated by a few single-bacteria phyla, such as the human gut. Furthermore, downstream analysis of metagenomic samples often requires the assembly of reads into larger fragments. Here, the current methods have several limitations; for example, they perform poorly on repetitive regions such as CRISPR elements (Skennerton, Imelfort, and Tyson 2013).

To understand the evolutionary drivers guiding genomic variation in microbial populations, metagenomics could give researchers a scalable method for analysing within-community variation on a populational scale. This would enable fine-grained analyses of the functions of community members, and studies of their association with phenotypes and environments, as well as studies of microevolution and adaptation to changing environmental conditions. Using metagenomic datasets to study molecular evolution would broaden the focus from individual species (Lenski and Travisano 1994) to whole communities, highlighting the importance of the reciprocal evolution of interacting species and opening up a universe of previously unknown interactions to be studied. In all publications presented in this thesis, I analysed such metagenomic samples and applied existing and novel methods to the raw or assembled genetic content of these communities to determine the genes, protein families and organisms under evolutionary selection.

4. Virus–bacteria coevolution can be quantified by d_N/d_S statistics

Environmental and host-associated microbial communities harbour trillions of microbes including viruses, among which bacteriophages (phages) that prey on bacteria are the most abundant. Phages are generally classified as virulent (strictly lytic) or temperate. The latter enables the phage to integrate its genome into the bacterial chromosome and become a prophage that is vertically transmitted in the bacterial population. Clearly, lysogenic phages play an important role in managing bacterial populations and also in bacterial evolution. Since the majority of bacteria harbour prophage sequences, one of the main evolutionary drivers is the reciprocal interaction between temperate bacteriophages (phages) and the bacterial host (Canchaya et al. 2003). During this integration process via lysogenization, phage genes can alter the phenotype of the bacterial host and, once integrated, phages can multiply via concomitant lysis of the host (ibid.). However, viruses often have a relatively error-prone replication machinery, which, combined with a fast infection cycle, results in rapid coevolution with their hosts and in highly diverse sequences (Pedulla et al. 2003). It has been shown that phages can contain genes associated with virulence such as toxins, effector proteins or other fitness factors (Harald, Carlos, and Wolf-Dietrich 2004), highlighting that the coevolution of phages and bacteria is mutualistic rather than predatory (Boyd and Brüssow 2002). On the other hand, the presence of prophages alone can be seen as a further evolutionary factor, since these phages build anchor points for genome rearrangements and drive genome inversions and translocations (Harald, Carlos, and Wolf-Dietrich 2004). These prophages constitute, besides other mobile genetic elements, a substantial part of the laterally acquired DNA (Canchaya et al. 2003), and several databases have collected the genomic content of publicly available sequenced phages, such as

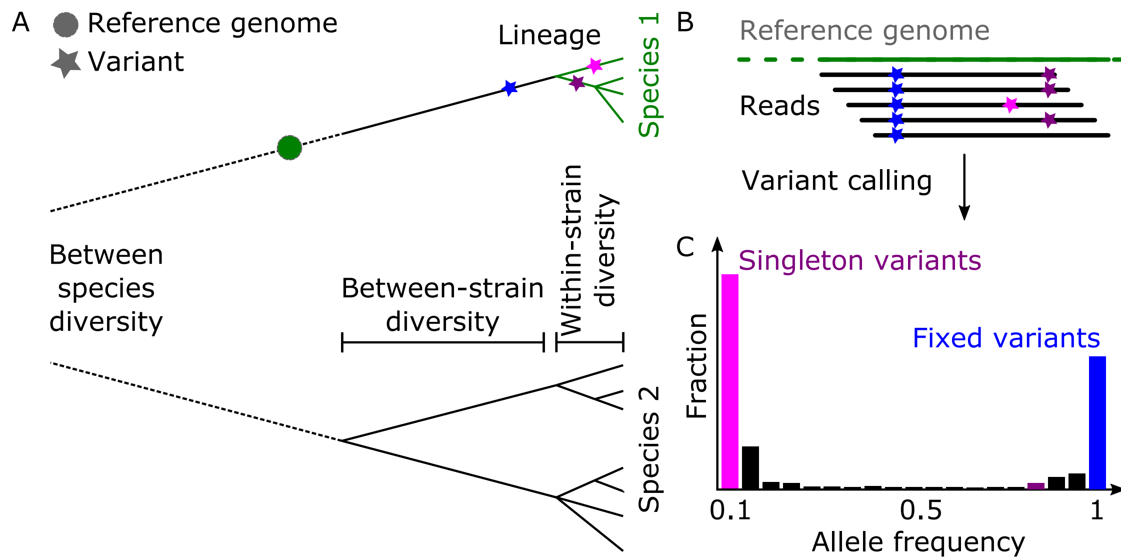

Figure 1: Evolutionary Processes in the microbiome likely manifests in variants observable of different frequencies relative to a reference genome A) Genetic diversity of two genetically distinct lineages of a bacterial species that have colonized a host. Within a lineage, substitutions (stars) likely have accumulated and can have profound impact on the host health outcomes. In many study settings, a reference genome has been identified e.g. via bacterial cultivation efforts followed by NGS (green point). B) The distribution of variants can be estimated by mapping the metagenomic reads to the reference genome, often via accounting for technical biases such as sequencing errors. C) The site frequency spectrum can be represented by the allele frequency of observed variants in the alignment. This typically consists of substitutions that have arisen since colonization to the host and are less frequent or singleton variants which constitute the low-frequency bin in the allele frequency plot. Diverged lineages are often represented in the alignment by variants with a higher allele frequency. Variants accumulated shortly after obtaining the reference sequence or mutations that exhibit a high fitness benefit might be fixed and have a AF of ~ 1.

PHASTER (Arndt et al. 2016), the successor of PHAST (Zhou et al. 2011).

In principle, shotgun sequencing approaches using metagenomics sequence virus-like particles alongside the chromosomal bacterial DNA; they can also be used to survey the complete viral community, without selection based on host or sequence similarity to known viruses (Rosario et al. 2009). However, early attempts to study phage diversity through metagenomics found that the majority of phage sequences could not be classified into any known viral taxonomy, highlighting that the current databases underestimate phage diversity (Simmonds et al. 2017; Turner, Kropinski, and Adriaenssens 2021). Moreover, studies using metagenomic techniques have uncovered abundant novel viruses in the human gastrointestinal (GI) tract that had not been detected previously, such as the so-called crAssphages (Edwards et al. 2019), which are viruses that may have coevolved with the human lineage and are an integral part of the normal human gut virome. This suggests a yet undetected diversity, often referred to as "viral dark matter", which remains to be quantified (Hatfull 2015; Krishnamurthy and D. Wang 2017). Attempts have been made to quantify the phage diversity collected from metagenomes in databases such as the "Gut Phage Database" (Unterer, Khan Mirzaei, and Deng 2021) which includes over >140,000 non-redundant gut phages obtained by analysing $> 28,000$ shotgun metagenomic datasets. Other studies of the human gut metagenome have been mined for phage genomes (Gregory et al. 2019) or combined data generated from cultivated approaches and metagenomes (Paez-Espino et al. 2019). However, further research is needed to understand the role of phages, particularly in the context of human health, since the role of phages present in the human GI tract is still unclear (Camarillo-Guerrero et al. 2021).

Although the performance of software tools analysing metagenomic samples for their bacterial content has increased (e.g., metagenomic assemblers and taxonomic binners), these tools have been found to underperform, especially on viral genomes (Sczyrba et al. 2017; Meyer et al. 2021). This gap might be filled by a new generation of machine learning classifiers based on deep learning (Bzhalava et al. 2018; Liu et al. 2020; Ren et al. 2020) or optimization of traditional approaches for reconstructing viromes (e.g., by analysing subgraphs of the metagenomic assembly graph) (Antipov et al. 2020). Furthermore, the co-evolutionary properties could be analysed more directly without assembly; for example, by analysing the CRISPR spacer content within metagenomes that could indicate recent encounters between bacteria and phages.

5. CRISPR–Cas immunity is an example of adaptive evolution

As a product of the evolutionary arms race, a fraction of the genome of most bacteria is dedicated to defense against phages (Koonin, Makarova, and Zhang 2017). Broadly, these systems can be categorized into "innate defense" including the restriction–modification systems that discrimination between self and non-self DNA via methylation (Tock and Dryden 2005; Goldfarb et al. 2015; Ofir et al. 2018), and abortive infection (Abi) systems that shut down the bacterial cell cycle upon infection (Fineran et al. 2009) and adaptive defense systems. One well-studied adaptive system involves clustered regularly interspaced short palindromic repeats (CRISPR), which consist of 23–50 nt repeated sequences interspaced by the so-called spacers (**Fig. 2**). These sets of spacers construct a memory of past encounters with foreign DNA and

allow bacteria to counteract attacks by phages with matching genomic loci. Although they were first described in *E. coli* (Nakata, Amemura, and Makino 1989), they are found in ~40% of the bacterial strains present in public databases (Makarova, Haft, et al. 2011; Pourcel et al. 2020).

CRISPR-associated (Cas) proteins play a role in the three steps of CRISPR—Cas immunity, namely (i) the acquisition of new spacers in the CRISPR loci, (ii) the maturation of the CRISPR transcript into small CRISPR RNAs and (iii) interference, which involves targeting and cleavage of the invading genome (Haft et al. 2005; Koonin, Makarova, and Zhang 2017). It has been observed that specific sets of *cas* genes exist in genomes, which guided the classification of *cas* types and subtypes (Makarova, Haft, et al. 2011; Makarova, Wolf, Alkhnbashi, et al. 2015; Makarova, Wolf, Iranzo, et al. 2020). This is of interest, since these subsets might correspond to different modes of action or differences in their preferred targets. Some subsets are potentially relevant for the development of new genome editing tools, as these come with a limited set of *cas* genes required for functioning, such as the Cas9 subtype (Jinek et al. 2012; Hsu, Lander, and Zhang 2014).

Although the general mechanistic principles of CRISPR–Cas-based defence are known, there are several ongoing questions in CRISPR biology, such as their natural distribution and variation, their role in functions other than defence (such as regulatory roles), their interactions with plasmids and their contribution to bacterial evolvability as a whole. For example, the CRISPR–Cas system may also interfere with the acquisition of genes that could confer fitness benefits to the bacteria. CRISPR–Cas could limit the uptake of AB resistance genes in environments with high AB pressure (Shehreen et al. 2019).

In general, the CRISPR-Cas system is a rapidly evolving defense system that is in a continuous arms race with bacteriophages, resulting in high evolutionary pressure on *cas* gene sequences (Koonin 2019a), which can be observed via the d_N/d_S statistics (Münch, Stecher, and McHardy 2017). It also can be seen as the mechanism of a specific form of genome evolution (i.e., a system that directly facilitates directed genome evolution). It has been discussed as evidence of the presence of the quasi-Lamarckian mode of evolution (Koonin 2019b; Wideman et al. 2019). However, this system remains to be studied, since the degree to which the self-targeting effects of CRISPR-Cas systems might be functionally relevant are unclear, such as in the context of quorum regulation, repair processes and others (Shmakov et al. 2018; Newsom et al. 2020).

6. Antibiotics and "Red Queen" dynamics

Evolutionary forces act not only to provide reproductive advantages in response to abiotic factors, but also to allow organisms to adapt to competing organisms and their competitive strategies resulting in the so-called "Red Queen" dynamics (Van and Van Valen 1973). One product of these coevolutionary Red Queen interactions is the evolution of microbial defense systems, such as AB compounds and bacteriocins. Survival under conditions with high AB pressure might select for bacteria with a higher mutation rate, since this can accelerate adaptations (Taddei et al. 1997). These mutations might alter AB targets, such as the target proteins, or cell

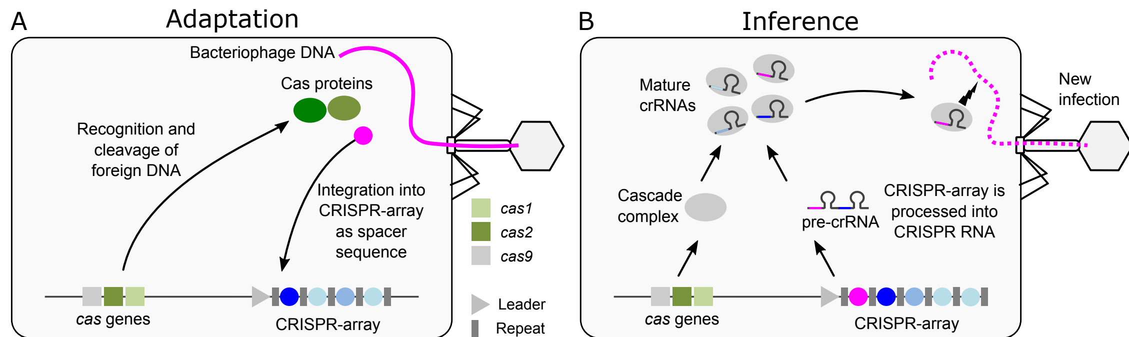

Figure 2: Schematic representation of CRISPR–Cas adaptation and interference. A) Incorporation of foreign DNA as spacers in the CRISPR-array from the invader (e.g. plasmid or phage). B) The CRISPR interference process degrades foreign DNA that was previously incorporated as spacer sequences.

permeability (Heisig 1996). Besides mutations, many resistance developments can be explained by the acquisition of resistance genes via horizontal gene transfer events (Robicsek, Jacoby, and Hooper 2006). However, developing resistance to AB imposes a high fitness cost on the bacteria, which could be limited by compensatory mutations that restore fitness without the loss of resistance (Levin et al. 1997). A further evolutionary strategy for surviving high doses of AB is the stress-induced activation of several regulatory networks, such as the SOS response (Cirz et al. 2007), which was found to be activated in bacteria in response to ciprofloxacin, beta-lactams and fluoroquinolones (Ames and Gold 1990; Miller et al. 2004; Pérez-Capilla et al. 2005). This response can increase the rate of mutations through the induction of error-prone DNA polymerases, thus increasing the rate of acquiring potential mutations conferring resistance to ABs. Furthermore, some evidence has suggested that AB stress can increase the bacterial recombination rate (López and Blázquez 2009), which is a further major driving force in bacterial evolution and survival (Guttman and Dykhuizen 1994), since elements taken up via horizontal gene transfer are integrated into the bacterial chromosome by such recombination events.

Human-induced factors drive coevolution between the host and pathogens in agriculture, aquaculture, and forestry (Percy 2007). Direct public health consequences can arise from this evolutionary interaction in the form of (multi-) drug-resistant strains (Barbosa and Levy 2000) and further drive biological evolution. Understanding these evolutionary processes is of profound importance for human prosperity, since it would enable the development of predictive models that could be used, for example, to estimate the potential of a community to develop resistance against certain ABs. In particular, the human gut is overly exposed to antibiotics and harbours a wide variety of resistance genes (Salyers, Gupta, and Y. Wang 2004).

7. The gastrointestinal system is a melting pot of coevolutionary interactions

In humans, the GI tract has a high microbe density. Metagenomic analyses of the human distal gut microbiome have been performed for around 15 years (Gill et al. 2006) boosting the development of metagenomic approaches (Robinson, Bohannan, and Young 2010). The gut microbiome describes the collection of microorganisms in the human GI tract and their collective interacting genomes. The study of these interactions was motivated by the observation that shifts in microbiome composition occur between healthy and diseased individuals, as well as between age groups and between humans from different geographic locations (Yatsunenko et al. 2012). Recently, similar shifts have even been described in human disease conditions as diverse as autism, cancer and diabetes.

One major problem in interpreting the results of observational studies is a limitation in their ability to prove causation. For example, epidemiological studies established that human health is largely impacted by socioeconomic factors such as educational attainment (Doubeni et al. 2012). Further, diet quality seems to be tightly linked to socioeconomic status (Darmon and Drewnowski 2008) and there seem to be direct relationships between dietary factors, such as dietary energy density, and human diseases (Swinburn et al. 2004). Since changes in the diet were shown to directly alter microbiome composition (Singh et al. 2017), microbiome studies should adequately address these potentially confounding effects. However, even when these are addressed carefully, these confounding effects are likely to persist. Fortunately, some recent advances in research methodologies are less prone to confounding and might clarify which microbiome–disease interactions are causal. One potential method is the application of Mendelian randomization techniques to microbiome studies, but these require large genome–wide association studies highlighting human genome-bacteria associations. Another methodological advance for detecting the causal relationships between the microbiome and disease is the development of model systems, such as gnotobiotic (greek; gnotos: known; bios: life) mouse models. In contrast to normal mice, these mice harbour microorganisms or defined minimal microbial communities that have been applied to germ-free mice (e.g., through gavaging with pure cultures of the bacteria of interest). The composition of these communities can be modulated, their genome sequences are known and the bacteria can also be studied *ex vivo* in the laboratory. Within this framework, immunological and ecological mechanisms caused by individual bacterial species can be studied with high statistical power.

To generate gnotobiotic mouse models, germ-free mice are colonised with microbial communities of varying complexity and origins at defined stages of their life to determine how much of the donor phenotype is transferable to the resulting conventionalised mice via the microbiota (Gordon and Pesti 1971; Faith et al. 2010). Examples of such communities include the Altered Schaedler Flora, where causative roles of single microorganisms have been linked to disease severity in the context of immune mediation (Dewhirst et al. 1999). Variations of this Altered Schaedler Flora consortium have been developed, often with the intent of triggering a specific disease phenotype in mice. Another example is the Oligo Mouse Microbiota model (OMM^{12}), which is a bacterial community that provides colonisation resistance against

infection with specific human pathogens such as Salmonella enterica serovar Typhimurium (Brugiroux et al. 2016). Through this model, causative relationships have been demonstrated, for example, in the case of *Mucispirillum schaedleri*, which provides protection against Salmonella-induced colitis (Herp et al. 2019), or *E. coli*, which provides colonisation resistance against this pathogen (Eberl, Weiss, et al. 2021). The OMM^{12} model has also proven to be a valuable tool for studying evolutionary adaptations (Yilmaz et al. 2021). As part of my dissertation, I applied the OMM^{12} model to study the evolutionary adaptations of the community during repeated rounds of AB exposure.

List of Figures

1 Evolutionary Processes in the microbiome likely manifests in variants observable of different frequencies relative to a reference genome A) Genetic diversity of two genetically distinct lineages of a bacterial species that have colonized a host. Within a lineage, substitutions (stars) likely have accumulated and can have profound impact on the host health outcomes. In many study settings, a reference genome has been identified e.g. via bacterial cultivation efforts followed by NGS (green point). B) The distribution of variants can be estimated by mapping the metagenomic reads to the reference genome, often via accounting for technical biases such as sequencing errors. C) The site frequency spectrum can be represented by the allele frequency of observed variants in the alignment. This typically consists of substitutions that have arisen since colonization to the host and are less frequent or singleton variants which constitute the low-frequency bin in the allele frequency plot. Diverged lineages are often represented in the alignment by variants with a higher allele frequency. Variants accumulated shortly after obtaining the reference sequence or mutations that exhibit a high fitness benefit might be fixed and have a AF of ~ 1. xxv

2 Schematic representation of CRISPR–Cas adaptation and interference. A) Incorporation of foreign DNA as spacers in the CRISPR-array from the invader (e.g. plasmid or phage). B) The CRISPR interference process degrades foreign DNA that was previously incorporated as spacer sequences. xxviii

bibibliography.bib

Bibliography

Amann, R I, W Ludwig, and K H Schleifer (Mar. 1995). "Phylogenetic identification and in situ detection of individual microbial cells without cultivation". en. In: *Microbiol. Rev.* 59.1, pp. 143–169.

Ames, B N and L S Gold (Aug. 1990). "Too many rodent carcinogens: mitogenesis increases mutagenesis". en. In: *Science* 249.4972, pp. 970–971.

Antipov, Dmitry et al. (Aug. 2020). "Metaviral SPAdes: assembly of viruses from metagenomic data". en. In: *Bioinformatics* 36.14, pp. 4126–4129.

Arndt, David et al. (July 2016). "PHASTER: a better, faster version of the PHAST phage search tool". en. In: *Nucleic Acids Res.* 44.W1, W16–21.

Barbosa, Teresa M and Stuart B Levy (Oct. 2000). "The impact of antibiotic use on resistance development and persistence". en. In: *Drug Resist. Updat.* 3.5, pp. 303–311.

Boyd, E Fidelma and Harald Brüssow (Nov. 2002). "Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved". en. In: *Trends Microbiol.* 10.11, pp. 521–529.

Brock, David C and Gordon E Moore (2006). *Understanding Moore's Law: Four Decades of Innovation*. en. Chemical Heritage Foundation.

Brown, Tom B et al. (May 2020). "Language Models are Few-Shot Learners". In: arXiv: `2005.14165 [cs.CL]`.

Brugiroux, Sandrine et al. (Nov. 2016). "Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium". en. In: *Nat Microbiol* 2, p. 16215.

Bulgarelli, Davide et al. (Mar. 2015). "Structure and function of the bacterial root microbiota in wild and domesticated barley". en. In: *Cell Host Microbe* 17.3, pp. 392–403.

Buschiazzo, Emmanuel et al. (Jan. 2012). "Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms". en. In: *BMC Evol. Biol.* 12, p. 8.

Bzhalava, Zurab et al. (Sept. 2018). "Machine learning for detection of viral sequences in human metagenomic datasets". en. In: *BMC Bioinformatics* 19.1, p. 336.

Camarillo-Guerrero, Luis F et al. (Feb. 2021). "Massive expansion of human gut bacteriophage diversity". en. In: *Cell* 184.4, 1098–1109.e9.

Canchaya, Carlos et al. (Aug. 2003). "Phage as agents of lateral gene transfer". en. In: *Curr. Opin. Microbiol.* 6.4, pp. 417–424.

Caro-Quintero, Alejandro and Konstantinos T Konstantinidis (Feb. 2012). "Bacterial species may exist, metagenomics reveal". en. In: *Environ. Microbiol.* 14.2, pp. 347–355.

Cirz, Ryan T et al. (Jan. 2007). "Complete and SOS-mediated response of Staphylococcus aureus to the antibiotic ciprofloxacin". en. In: *J. Bacteriol.* 189.2, pp. 531–539.

Ciufo, Stacy et al. (July 2018). "Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI". en. In: *Int. J. Syst. Evol. Microbiol.* 68.7, pp. 2386–2392.

Darmon, Nicole and Adam Drewnowski (May 2008). "Does social class predict diet quality?" en. In: *Am. J. Clin. Nutr.* 87.5, pp. 1107–1117.

DeSalle, Rob (Sept. 2000). "Adaptive evolution of genes and genomes". en. In: *Heredity* 85.3, pp. 303–303.

Dewhirst, F E et al. (Aug. 1999). "Phylogeny of the defined murine microbiota: altered Schaedler flora". en. In: *Appl. Environ. Microbiol.* 65.8, pp. 3287–3292.

Doubeni, Chyke A et al. (July 2012). "Socioeconomic status and the risk of colorectal cancer: an analysis of more than a half million adults in the National Institutes of Health-AARP Diet and Health Study". en. In: *Cancer* 118.14, pp. 3636–3644.

Eberl, Claudia, Diana Ring, et al. (2019). "Reproducible colonization of germ-free mice with the Oligo-Mouse-Microbiota in different animal facilities". en. In: *Front. Microbiol.* 10, p. 2999.

Eberl, Claudia, Anna S Weiss, et al. (Sept. 2021). "E. coli enhance colonization resistance against Salmonella typhimurium by competing for galactitol, a context-dependent limiting carbon source". en. In: *Cell Host Microbe*.

Edwards, Robert A et al. (Oct. 2019). "Global phylogeography and ancient evolution of the widespread human gut virus crAssphage". en. In: *Nat Microbiol* 4.10, pp. 1727–1736.

Endo, T, K Ikeo, and T Gojobori (May 1996). "Large-scale search for genes on which positive selection may operate". en. In: *Mol. Biol. Evol.* 13.5, pp. 685–690.

Faith, Jeremiah J et al. (Sept. 2010). "Creating and characterizing communities of human gut microbes in gnotobiotic mice". en. In: *ISME J.* 4.9, pp. 1094–1098.

Fineran, Peter C et al. (Jan. 2009). "The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 106.3, pp. 894–899.

Gill, Steven R et al. (June 2006). "Metagenomic analysis of the human distal gut microbiome". en. In: *Science* 312.5778, pp. 1355–1359.

Goldfarb, Tamara et al. (Jan. 2015). "BREX is a novel phage resistance system widespread in microbial genomes". en. In: *EMBO J.* 34.2, pp. 169–183.

Gordon, H A and L Pesti (Dec. 1971). "The gnotobiotic animal as a tool in the study of host microbial relationships". en. In: *Bacteriol. Rev.* 35.4, pp. 390–429.

Gouy, M and C Gautier (Nov. 1982). "Codon usage in bacteria: correlation with gene expressivity". en. In: *Nucleic Acids Res.* 10.22, pp. 7055–7074.

Grantham, R et al. (Jan. 1980). "Codon catalog usage and the genome hypothesis". en. In: *Nucleic Acids Res.* 8.1, r49–r62.

Gregory, Ann C et al. (July 2019). "The human gut virome database". en.

Guttman, D S and D E Dykhuizen (Nov. 1994). "Clonal divergence in Escherichia coli as a result of recombination, not mutation". en. In: *Science* 266.5189, pp. 1380–1383.

Hacquard, Stéphane et al. (May 2016). "Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi". en. In: *Nat. Commun.* 7, p. 11362.

Haft, Daniel H et al. (Nov. 2005). "A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes". en. In: *PLoS Comput. Biol.* 1.6, e60.

Harald, Brüssow, Canchaya Carlos, and Hardt Wolf-Dietrich (Sept. 2004). "Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion". In: *Microbiol. Mol. Biol. Rev.* 68.3, pp. 560–602.

Hatfull, Graham F (Aug. 2015). "Dark Matter of the biosphere: the amazing world of bacteriophage diversity". en. In: *J. Virol.* 89.16, pp. 8107–8110.

Heisig, P (Apr. 1996). "Genetic evidence for a role of parC mutations in development of high-level fluoroquinolone resistance in Escherichia coli". en. In: *Antimicrob. Agents Chemother.* 40.4, pp. 879–885.

Herp, Simone et al. (May 2019). "Mucispirillum schaedleri antagonizes Salmonella virulence to protect mice against colitis". en. In: *Cell Host Microbe* 25.5, 681–694.e8.

Hsu, Patrick D, Eric S Lander, and Feng Zhang (June 2014). "Development and applications of CRISPR-Cas9 for genome engineering". en. In: *Cell* 157.6, pp. 1262–1278.

Jain, Chirag et al. (Nov. 2018). "High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries". en. In: *Nat. Commun.* 9.1, p. 5114.

Jinek, Martin et al. (Aug. 2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity". en. In: *Science* 337.6096, pp. 816–821.

Jumper, John et al. (July 2021). "Highly accurate protein structure prediction with AlphaFold". en. In: *Nature*.

Kaplan, Jared et al. (Jan. 2020). "Scaling Laws for Neural Language Models". In: arXiv: `2001.08361 [cs.LG]`.

Koonin, Eugene V (Jan. 2019a). "CRISPR: a new principle of genome engineering linked to conceptual shifts in evolutionary biology". en. In: *Biol. Philos.* 34.1, p. 9.

– (Feb. 2019b). "Lamarckian or not, CRISPR-Cas is an elaborate engine of directed evolution". In: *Biol. Philos.* 34.1, p. 17.

Koonin, Eugene V, Kira S Makarova, and Feng Zhang (June 2017). "Diversity, classification and evolution of CRISPR-Cas systems". en. In: *Curr. Opin. Microbiol.* 37, pp. 67–78.

Krishnamurthy, Siddharth R and David Wang (July 2017). "Origins and challenges of viral dark matter". en. In: *Virus Res.* 239, pp. 136–142.

Kunin, Victor et al. (Dec. 2008). "A bioinformatician's guide to metagenomics". en. In: *Microbiol. Mol. Biol. Rev.* 72.4, 557–78, Table of Contents.

Lennon, Jay T and Kenneth J Locey (Mar. 2020). "More support for Earth's massive microbiome". en. In: *Biol. Direct* 15.1, p. 5.

Lenski, R E and M Travisano (July 1994). "Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 91.15, pp. 6808–6814.

Levin, B R et al. (Jan. 1997). "The population genetics of antibiotic resistance". en. In: *Clin. Infect. Dis.* 24 Suppl 1, S9–16.

Liu, Fu et al. (Dec. 2020). "RNN-VirSeeker: a deep learning method for identification of short viral sequences from metagenomes". en. In: *IEEE/ACM Trans. Comput. Biol. Bioinform.* PP.

Locey, Kenneth J and Jay T Lennon (May 2016). "Scaling laws predict global microbial diversity". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 113.21, pp. 5970–5975.

López, E and J Blázquez (2009). "Effect of subinhibitory concentrations of antibiotics on intrachromosomal homologous recombination in Escherichia coli". In: *Antimicrob. Agents Chemother.* 35.8, pp. 3411–3415.

Lynch, Michael et al. (Oct. 2016). "Genetic drift, selection and the evolution of the mutation rate". en. In: *Nat. Rev. Genet.* 17.11, pp. 704–714.

Makarova, Kira S, Daniel H Haft, et al. (May 2011). "Evolution and classification of the CRISPR–Cas systems". en. In: *Nat. Rev. Microbiol.* 9.6, pp. 467–477.

Makarova, Kira S, Yuri I Wolf, Omer S Alkhnbashi, et al. (Nov. 2015). "An updated evolutionary classification of CRISPR-Cas systems". en. In: *Nat. Rev. Microbiol.* 13.11, pp. 722–736.

Makarova, Kira S, Yuri I Wolf, Jaime Iranzo, et al. (Feb. 2020). "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants". en. In: *Nat. Rev. Microbiol.* 18.2, pp. 67–83.

Marais, Gabriel A B, Bérénice Batut, and Vincent Daubin (Oct. 2020). "Genome evolution: mutation is the main driver of genome size in prokaryotes". en. In: *Curr. Biol.* 30.19, pp. 1083–1085.

Merino, Nancy et al. (Apr. 2019). "Living at the extremes: extremophiles and the limits of life in a planetary context". en. In: *Front. Microbiol.* 10, p. 780.

Meyer, F et al. (2021). "Critical assessment of metagenome interpretation - the second round of challenges". In: *bioRxiv*.

Miller, Christine et al. (Sept. 2004). "SOS response induction by beta-lactams and bacterial defense against antibiotic lethality". en. In: *Science* 305.5690, pp. 1629–1631.

Muir, Paul et al. (Mar. 2016). "The real cost of sequencing: scaling computation to keep pace with data generation". en. In: *Genome Biol.* 17, p. 53.

Münch, Philipp C, Bärbel Stecher, and Alice C McHardy (Oct. 2017). "EDEN: evolutionary dynamics within environments". en. In: *Bioinformatics* 33.20, pp. 3292–3295.

Murray, Connor S, Yingnan Gao, and Martin Wu (July 2021). "Re-evaluating the evidence for a universal genetic boundary among microbial species". en. In: *Nat. Commun.* 12.1, p. 4059.

Nakata, A, M Amemura, and K Makino (June 1989). "Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome". en. In: *J. Bacteriol.* 171.6, pp. 3553–3556.

Newsom, Sydney et al. (2020). "The CRISPR-Cas mechanism for adaptive immunity and alternate bacterial functions fuels diverse biotechnologies". en. In: *Front. Cell. Infect. Microbiol.* 10, p. 619763.

Nielsen, Rasmus (2005). "Molecular signatures of natural selection". en. In: *Annu. Rev. Genet.* 39, pp. 197–218.

Ochman, H, J G Lawrence, and E A Groisman (May 2000). "Lateral gene transfer and the nature of bacterial innovation". en. In: *Nature* 405.6784, pp. 299–304.

Ofir, Gal et al. (Jan. 2018). "DISARM is a widespread bacterial defence system with broad anti-phage activities". en. In: *Nat Microbiol* 3.1, pp. 90–98.

Paez-Espino, David et al. (Jan. 2019). "IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes". en. In: *Nucleic Acids Res.* 47.D1, pp. D678–D686.
Pedulla, Marisa L et al. (Apr. 2003). "Origins of highly mosaic mycobacteriophage genomes". en. In: *Cell* 113.2, pp. 171–182.
Percy, Diana M (Feb. 2007). "The geographic mosaic of coevolution". In: *Syst. Biol.* 56.1, pp. 149–151.
Pérez, Vilma et al. (June 2017). "Bacterial Survival under Extreme UV Radiation: A Comparative Proteomics Study of Rhodobacter sp., Isolated from High Altitude Wetlands in Chile". en. In: *Front. Microbiol.* 8, p. 1173.
Pérez-Capilla, Tatiana et al. (Feb. 2005). "SOS-independent induction of dinB transcription by beta-lactam-mediated inhibition of cell wall synthesis in Escherichia coli". en. In: *J. Bacteriol.* 187.4, pp. 1515–1518.
Petersen, Lise et al. (Sept. 2007). "Genes under positive selection in Escherichia coli". en. In: *Genome Res.* 17.9, pp. 1336–1343.
Pourcel, Christine et al. (Jan. 2020). "CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers". en. In: *Nucleic Acids Res.* 48.D1, pp. D535–D544.
Ranea, Juan A G et al. (Jan. 2005). "Microeconomic principles explain an optimal genome size in bacteria". en. In: *Trends Genet.* 21.1, pp. 21–25.
Reis, Mario dos, Renos Savva, and Lorenz Wernisch (Sept. 2004). "Solving the riddle of codon usage preferences: a test for translational selection". en. In: *Nucleic Acids Res.* 32.17, pp. 5036–5044.
Ren, Jie et al. (2020). *Identifying viruses from metagenomic data using deep learning*.
Robicsek, Ari, George A Jacoby, and David C Hooper (Oct. 2006). "The worldwide emergence of plasmid-mediated quinolone resistance". en. In: *Lancet Infect. Dis.* 6.10, pp. 629–640.
Robinson, Courtney J, Brendan J M Bohannan, and Vincent B Young (Sept. 2010). "From structure to function: the ecology of host-associated microbial communities". en. In: *Microbiol. Mol. Biol. Rev.* 74.3, pp. 453–476.
Rocha, Eduardo P C et al. (Mar. 2006). "Comparisons of dN/dS are time dependent for closely related bacterial genomes". en. In: *J. Theor. Biol.* 239.2, pp. 226–235.
Rodriguez-R, Luis M et al. (July 2021). "Reply to: "Re-evaluating the evidence for a universal genetic boundary among microbial species"". en. In: *Nat. Commun.* 12.1, p. 4060.
Rosario, Karyna et al. (Nov. 2009). "Metagenomic analysis of viruses in reclaimed water". en. In: *Environ. Microbiol.* 11.11, pp. 2806–2820.
Salyers, Abigail A, Anamika Gupta, and Yanping Wang (Sept. 2004). "Human intestinal bacteria as reservoirs for antibiotic resistance genes". en. In: *Trends Microbiol.* 12.9, pp. 412–416.
Sczyrba, Alexander et al. (Nov. 2017). "Critical Assessment of Metagenome Interpretation-a benchmark of metagenomics software". en. In: *Nat. Methods* 14.11, pp. 1063–1071.
Shehreen, Saadlee et al. (May 2019). "Genome-wide correlation analysis suggests different roles of CRISPR-Cas systems in the acquisition of antibiotic resistance genes in diverse species". en. In: *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374.1772, p. 20180384.

Shmakov, Sergey A et al. (June 2018). "Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 115.23, E5307–E5316.

Simmonds, Peter et al. (Jan. 2017). "Virus taxonomy in the age of metagenomics". en. In: *Nat. Rev. Microbiol.* 15.3, pp. 161–168.

Singh, Rasnik K et al. (Apr. 2017). "Influence of diet on the gut microbiome and implications for human health". en. In: *J. Transl. Med.* 15.1, p. 73.

Skennerton, Connor T, Michael Imelfort, and Gene W Tyson (May 2013). "Crass: identification and reconstruction of CRISPR from unassembled metagenomic data". en. In: *Nucleic Acids Res.* 41.10, e105.

Swinburn, B A et al. (Feb. 2004). "Diet, nutrition and the prevention of excess weight gain and obesity". en. In: *Public Health Nutr.* 7.1A, pp. 123–146.

Taddei, F et al. (June 1997). "Role of mutator alleles in adaptive evolution". en. In: *Nature* 387.6634, pp. 700–702.

Tock, Mark R and David T F Dryden (Aug. 2005). "The biology of restriction and anti-restriction". en. In: *Curr. Opin. Microbiol.* 8.4, pp. 466–472.

Turner, Dann, Andrew M Kropinski, and Evelien M Adriaenssens (Mar. 2021). "A Roadmap for Genome-Based Phage Taxonomy". en. In: *Viruses* 13.3.

Unterer, Magdalena, Mohammadali Khan Mirzaei, and Li Deng (May 2021). "Gut Phage Database: phage mining in the cave of wonders". en. In: *Signal Transduct Target Ther* 6.1, p. 193.

Van, Valen and L Van Valen (1973). "A new evolutionary law". In.

Vatanen, Tommi et al. (Mar. 2019). "Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life". en. In: *Nat Microbiol* 4.3, pp. 470–479.

Weiss, A S et al. (2021). "Exploring the interaction network of a synthetic gut bacterial community". In: *bioRxiv*.

West, Stuart A et al. (Dec. 2007). "The Social Lives of Microbes". In: *Annu. Rev. Ecol. Evol. Syst.* 38.1, pp. 53–77.

Wideman, Jeremy G et al. (Feb. 2019). "Mutationism, not Lamarckism, captures the novelty of CRISPR–Cas". In: *Biol. Philos.* 34.1, p. 12.

Yatsunenko, Tanya et al. (May 2012). "Human gut microbiome viewed across age and geography". en. In: *Nature* 486.7402, pp. 222–227.

Yilmaz, Bahtiyar et al. (Apr. 2021). "Long-term evolution and short-term adaptation of microbiota strains and sub-strains in mice". en. In: *Cell Host Microbe* 29.4, 650–663.e9.

Zhou, You et al. (July 2011). "PHAST: a fast phage search tool". en. In: *Nucleic Acids Res.* 39.Web Server issue, W347–52.

Appendix A: Unpublished Manuscript III

Pulsed antibiotic treatments of gnotobiotic mice manifests in complex community dynamics and resistance effects

Philipp C. Münch[1,2,3], Claudia Eberl[2], Diana Ring[2], Simon Wölfel[2], Adrian Fritz[1], Simone Herp[2], Robert Geffers[1], Eric A. Franzosa[3], Curtis Huttenhower[3], Alice C. McHardy[1,4] and Bärbel Stecher[2,4]

[1] Department for Computational Biology of Infection Research, Helmholtz Center for Infection Research, 38124 Braunschweig, Germany
[2] Max von Pettenkofer-Institute for Hygiene and Clinical Microbiology, Ludwig-Maximilian University of Munich, 80336 Munich, Germany
[3] Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA
[4] These authors share senior authorship

Introduction

A strong body of evidence supports the crucial role of the gut microbiota in human health and disease (Lloyd-Price, Abu-Ali, and Huttenhower 2016; Human Microbiome Project Consortium 2012). Microbes have co-adapted through various molecular mechanisms with mammalian hosts to form symbiotic stable relationships (Ley, Peterson, and Gordon 2006), and the ability to tolerate and respond to a certain amount of environmental perturbations (Sommer et al. 2017). This results in a fairly stable and host-specific microbial community where the majority of strains persist over several years (Faith et al. 2013). Yet, long-lasting disruptions of the microbiota have been observed, for example in response to extreme dietary changes (Turnbaugh et al. 2008) or antibiotic therapy (Lange et al. 2016). Antibiotics (AB) are likely the largest known driver of compositional changes in the gut microbiome, which are, in particular when they occur early in life, associated with development of several diseases such as obesity (Nagpal et al. 2018), allergic disorders (Droste et al. 2000) and infections (Croswell et al. 2009). Furthermore, increase in the development of antibiotic resistances frequently occurs in response to antibiotic overuse which is recognised by the WHO as one of the top ten threats to global health (Thangaraju and Venkatesan 2019).

Despite the immense impact that antibiotics have on the human microbiome, only a few studies have been performed to characterize both the short- and long-term responses of the human microbiome to antibiotic therapy (Nobel et al. 2015; Vrieze et al. 2014). In particular, it is insufficiently understood which factors determine the ability of the microbiota to return to the pre-antibiotic state, a process which is termed microbial resilience. It appears that resilience is mainly determined by the type of antibiotic used, as different antibiotic classes have distinct inhibition spectra ranging from strain-specific to broad-spectrum effects (Maier et al. 2021). Moreover, environmental factors, including diet, environmental reservoirs and microbial community context influence recovery of the microbiota after antibiotic perturbations (Ng et al. 2020). It is unclear however, to what extent other mechanisms like antibiotic detoxification (Klünemann et al. 2021), antibiotic tolerance or the evolution of antibiotic resistant microbial populations could also be involved in mediating community resilience. As the effect of drugs on single strains may vastly differ between *in vitro* and *in vivo* communities - these effects are

mainly caused by altered bioavailability of the drug bacterial physiology and microbial community effects like cross-feeding and competition - in vivo studies are needed.

In complex populations, such as the human gut, the analysis of the evolutionary dynamics of different bacterial populations is challenging and currently limited to organisms that can be re-isolated from fecal samples (Zhao et al. 2019). Instead, the study of bacterial model organisms in mice enables long-term tracking of within-host bacterial evolution as mutants can be re-isolated, and their genome sequence be compared to the initial population (Barroso-Batista et al. 2020). Recently, within-host evolution of a defined bacterial community, the Oligo-Mouse Microbiota (OMM^{12}) has been analyzed at SNP scale over a 6-year period. This revealed the emergence of stably-co-existing genotypes in the different bacterial populations (Yilmaz et al. 2021) which colonize the murine gut over years (Eberl et al. 2019).

We used stably associated OMM^{12} mice to study effects of three different classes of antibiotics on this microbial community in vivo. This community is composed of 12 publically available, sequenced strains isolated from mice and represents 5 bacterial phyla that are naturally abundant in the murine gut. Our study design was optimized to analyze short-term microbiota perturbation by the AB treatment as well as the recovery of the microbiota during a “resilience” phase in between different treatments and the response to a secondary AB challenge to discover potential adaptive community effects. We found that (i) there is, as expected, an antibiotic-dependent effect to the OMM^{12} community, (ii) strain-specific perturbations do not always correlate with in vitro determined MIC values determined (iii) the severity of community perturbation after repeated antibiotic courses decreases over time, pointing at the evolution of “resilience effects” within the community; (iv) we identify that become unaffected by the AB in the secondary challenge. In select cases we identify *de novo* mutations in antibiotic-target genes and which correlated with increased MIC to the AB(v) we solved several problems in read mapping of metagenomics data to a defined consortium, for example, shared conserved regions between and within genome lead to an overestimation of variants when not addressed carefully.

Results

Community resilience and a wide variety of response patterns in members of a defined bacterial community in response to repeated antibiotic treatment

We studied community dynamics of the OMM^{12} during repeated antibiotic applications for 80 days and performed shotgun metagenomics and strain-specific qPCR to track community composition and evolution. In order to observe adaptation of the community to antibiotic challenge, the study design included a total of 4 5-days treatment periods ($t_1 - t_4$), interspaced by resilience phases of varying lengths ranging from 5 days up to 35 days (**Fig. 1A**), which allowed us to study the influence of repeated applications of AB to the microbial community dynamics. In this setting, gnotobiotic OMM^{12} adult mice caged under germ-free conditions (n = 5 per group) were treated with physiologically relevant concentrations (daily dose 25mg/kg) of antibiotics vancomycin, tetracyclin and ciprofloxacin via the drinking water (**see Methods**). During the study period of 80 days, fecal samples from individual mice were collected on a weekly basis, amounting to a total of 307 samples.

Fecal bacterial community composition showed strong antibiotic-dependent responses, measured by strain-specificqPCR (**Fig. 1BC**). Overall composition dynamics across treatment groups were non-uniform, with a subset of bacterial members exhibiting decreased abundance after antibiotic exposure. Notably, the impact of AB treatment on community composition is

much clearer observable on absolute (**Fig. 1B**), compared to relative abundances (**Fig. 1C**), emphasizing that such signal could be lost when applying transformation to relative abundances or when measurements are only performed on a relative scale, which is practised in many metagenomic studies due to the absence of absolute quantification such as cell counts. Of the initial 12 bacterial members in the community, 10 were reliably detected by qPCR (missing *A. muris* and *B. animalis*), which has been observed in previous studies using the model (Eberl et al. 2019). Co-caged mice that undergo the same treatment exhibited a homogenous response to AB treatment (**Fig. 1BC**), which is likely attributable to coprophagy (Ng et al. 2020). We quantified the impact of AB applications to the community composition by measuring community stability of time points before and after the AB challenge using the Bray-Curtis distance metric. For all AB groups, the disruptions in community compositions are getting less pronounced with increased study time (**Fig. 1D**), indicating a resilience effect through adaptations of community members.

Permutational analysis of variance (PERMANOVA) of Bray-Curtis dissimilarity indicated that the treatment group explained 18% of microbial taxonomic variation (permutation test, $P < 0.001$), followed by the treatment status on case groups (13%, $P < 0.001$) and sampling date (4%, $P < 0.001$). The mouse ID is not significantly associated with variation in the taxonomic composition, indicating a high within-group homogeneity of the study population. This is in line with a PERMANOVA analysis of gene-length and coverage normalized (CPM) abundances of UniRef90 families (Bray-Curtis) recovered from the whole genome shotgun sequencing (WGS) samples (again with treatment group and sampling date accounting for 21% and 1.3% of variation, respectively; $P < 0.001$). Interestingly, we observe a higher variance explained by treatment status using UniRef90, compared to qPCR and WGS abundance (21%, 13%, 9%, respectively), which is indicative of the robustness of the measure.

We identified 5 major response patterns: (i) A decrease in absolute abundance at the consecutive time points after treatment for members such as *B. coccoides* in the tetracycline and vancomycin group,; A decrease in abundance similar to (i) but with more pronounced decrease after the first AB challenge t_1 compared to later ones, i.e. such as *E. faecalis* in the ciprofloxacin group. This may indicate that the community or the individual members evolve towards increased AB tolerance or resistance; (iii) the near extinction of a species after treatment over longer time periods, with re-appearance during the resilience phase, such as *E. faecalis* or *M. intestinale* after vancomycin treatment; (iv) and more complex patterns, for example of *E. faecalis* during tetracycline treatment that might arise from multiple sources including co-dependence of bacterial members (cross feeding) and the presence of multiple bacteria phenotypes at timpoint i. This high diversity of response patterns motivated us to study factors such as evolution of AB potentially influencing resistance or tolerance, such as bacterial growth, phage activation and selection of substrains.

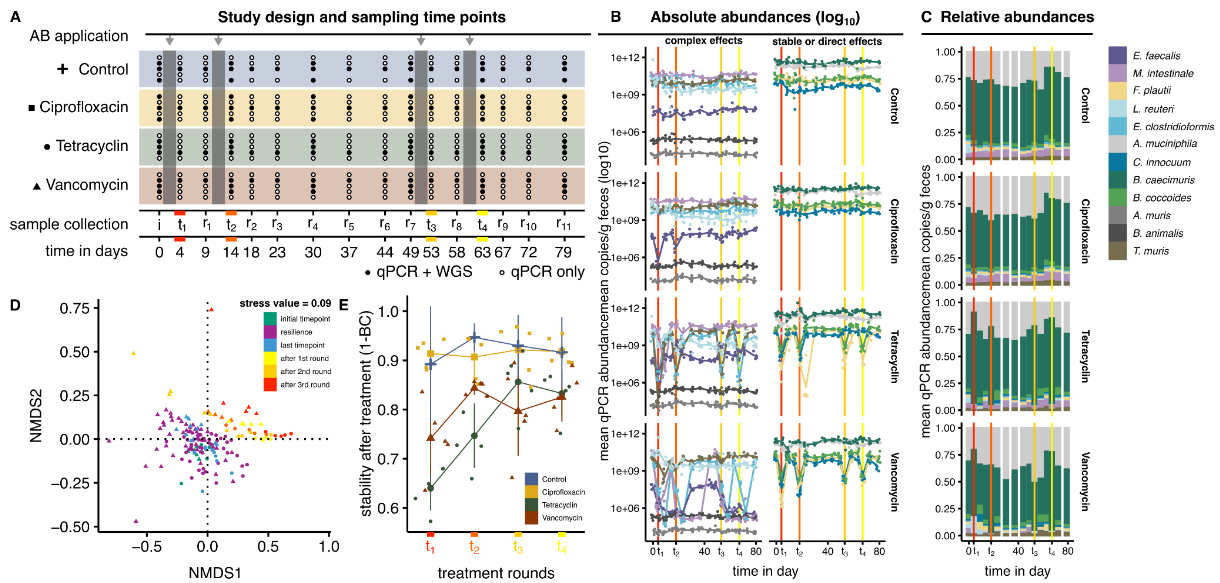

Figure 1: Stability of community over time and treatment regimes measured via species specific qPCR. A) Schematic representation of study design. 3 treatment groups and one control group (housed together in one cage per group with 5 mice each) undergo repeated treatment of antibiotics via drinking water (grey vertical bars) between sample collection time points (circles). Sample i denotes initial time point. t_1-t_4 denotes the time points directly after treatment, while r_1-r_4 denotes recovery time points where no treatment was applied in respect to the previous time point. Filled points denote fecal samples which were collected for each mouse individually. B) Community abundance of all observed community members of the OMM[12] community that show either direct response or no response to treatment (right) or a complex response pattern (left) as an absolute measure ($\log_{10}$) scaled. For each time point, the median community abundance is shown. C) Relative abundance of OMM[12] members at different sample time points D) Distance between OMM[12] composition of mice from the tetracycline and vancomycin group using non-metric multidimensional scaling (NMDS) of community abundance using Bray-Curtis distances (stress value = 0.09). Initial (i) communities show high similarity, while a bigger distance between samples is introduced by treatment. E) Successive treatments are less disruptive to the overall community distance within mice measured as deviations from previous time points using Bray-Curtis dissimilarities.

SNP occurrence associated with AB resistance pattern

During the time course of the experiment we observe treatment specific effects on the SNP scale (**Fig. 2**), measured by shotgun NovaSeq pyrosequencing (~6.8 Mbases per sample) on a subset of 112 samples on selected time points (**Fig. 1A**). This subset consists of all time points of one mouse per group plus two biological replicates that originate from mice co-housed on selected time points. In brief, we mapped quality controlled to the full reference OMM[12] collection via BWA mem with stringent conditions and filtered out duplicated reads and ones that map to multiple positions since we observed that this would inflate SNP counts artificially, due to high degree of similarities within and between OMM[12] community members. Variant calling (SNPs) was done with LoFreq under stringent settings (**see Methods**). We observed a number of variants with an intermediate and non-changing AF around ~0.5 which are likely the effect of duplication events in the genomes relative to the reference sequence, where one duplicated loci acquired a SNP during the study or on parental generations. We further confirmed the presence of these regions via the sequence of 90 colonies picked from the plates of the *in vivo* evolved community (*E. faecalis*, *C. innocuum* and *B. coccoides*) which also exhibit an intermediate AF of ~0.5 (**Fig. 5A**). Since this would inflate observed SNP load, we filtered variants that have a median AF between 0.1 and 0.7 and a low interquartile range (IQR) of smaller than 0.4, resulting in a final set of 505 SNPs observed in the whole study (**Supplementary Table 1**).

In samples taken from the tetracycline group, we observe SNPs on genes encoding for ribosomal apparatus, such as the 30S ribosomal protein S11 (q = $9.9*10^{-8}$ on *A. muciniphila,* **Supplementary Table 1**). Furthermore, in *B. caecimuris*, we observe two SNPs that are present on S9 gene (q-value = $1.4*10^{-8}$ and $1.9*10^{-8}$). As tetracyclines act as a protein synthesis inhibitor and mutations within the 30S gene are known AB binding, these mutations may lead to tetracyclin resistance. Another mode to acquire resistance is via AB export out of the cell such as RND-type drug efflux pumps. We observe mutations in the efflux RND transporter permease subunit on *M. intestinale* in the ciprofloxacin group (q-value = 0.008). In other bacteria, efflux pumps are associated with resistance against fluoroquinolones (Jiang et al. 2018; Godreuil et al. 2003). We further found a variety of SNPs in genes with no known link to resistance mechanism, such as the gene encoding for the transporter auxin efflux carrier (AEC) family transporter on *T. muris* in the ciprofloxacin and vancomycin group but not on the tetracycline group (q = $6.6*10^{-7}$ and $3.2*10^{-5}$, respectively, **Fig. 2**). It was shown that proteins with helicase activities such as the DEAH/RHA protein, is required to modulate bacterial survival under diverse antibiotics treatments (Grass et al. 2021). Both, in *B. caecimuris* and *T. muris*, we find mutations in the DEAD/DEAH box helicase family protein (**Fig. 2**), both with a q-value < 0.005 in the vancomycin group (**Supplementary Table X2**).

We observe variants associated with genes preservation of cell wall integrity, especially in the vancomycin group. Vancomycin is a glycopeptide antibiotic known to inhibit peptidoglycan synthesis and bind to acyl-D-alanyl-D-alanine and inhibits therefore the addition of new units to the peptidoglycan wall. In *T. muris*, we identified 3 possible related SNPs in the gene that encodes the cell envelope integrity protein (*tolA*) which are found in the ciprofloxaxin and vancomycin group (q-value < 0.0009). Furthermore we observe mutations in the *zapD* gene (**Fig. 2**), encoding essential functions for cell division. This *zapD* gene is one of the Z-associated genes, whose products are known as FtsZ-regulatory proteins that exhibit roles in stabilizing FtsZ-ring assembly (Durand-Heredia et al. 2012), which in turn is an antibiotic target (Silber et al. 2020; Casiraghi et al. 2020; Ur Rahman et al. 2020). Furthermore, *zapD* is linked to biofilm formation (H.-H. Chen et al. 2020) and bacteria within biofilms are refractory to host immune responses and antibiotic treatment (Arciola, Campoccia, and Montanaro 2018).

Beside these SNPs that are potentially in line with the mode of action of the antibiotics, we detect a wide range of mutations that are associated with resistance development, but not in particular with the antibiotic class we used in our study. For example, there is evidence that glycosyltransferase inhibitors could be an alternative therapeutic strategy to current antibiotics (Schutzbach and Brockhausen 2009; El Qaidi et al. 2018). Interestingly, in treatment groups find many SNPs associated with glycosyltransferase inhibition, such as “glycosyltransferase inhibitors, “glycoside hydrolase family 76 protein”, “glycosyltransferase family 1 protein” and “glycosyltransferase family 2 protein”, mostly in *T. muris* and *B. caecimuris*. In *M. intestinale*, we observe SNPs on the “GH92 family glycosyl hydrolase” (2 SNPs with q = $2.5*10^{-8}$ and $4.6*10^{-7}$ Tet., 1 SNP with q = 0.001 in Vanco.). Putative explanation for this could be the presence of substrains that provide a broad-spectrum resistance to a wide range of potential stressors, and these get selected by the application of specific antibiotics. Not all SNPs we observe to be selected should be considered to be associated with an antibiotic target, due to hitchhiking effects. However, some of them are in genes with hitherto unknown connection to AB resistance and show a time and application-dependent response, such as genes encoding for the RagB/SusD family nutrient uptake membrane protein (**Fig. 2**) which we detect to be associated with all AB types we have tested in *B. caecimuris*, and with vancomycin treatment in *M. intestinale* and requires further investigation.

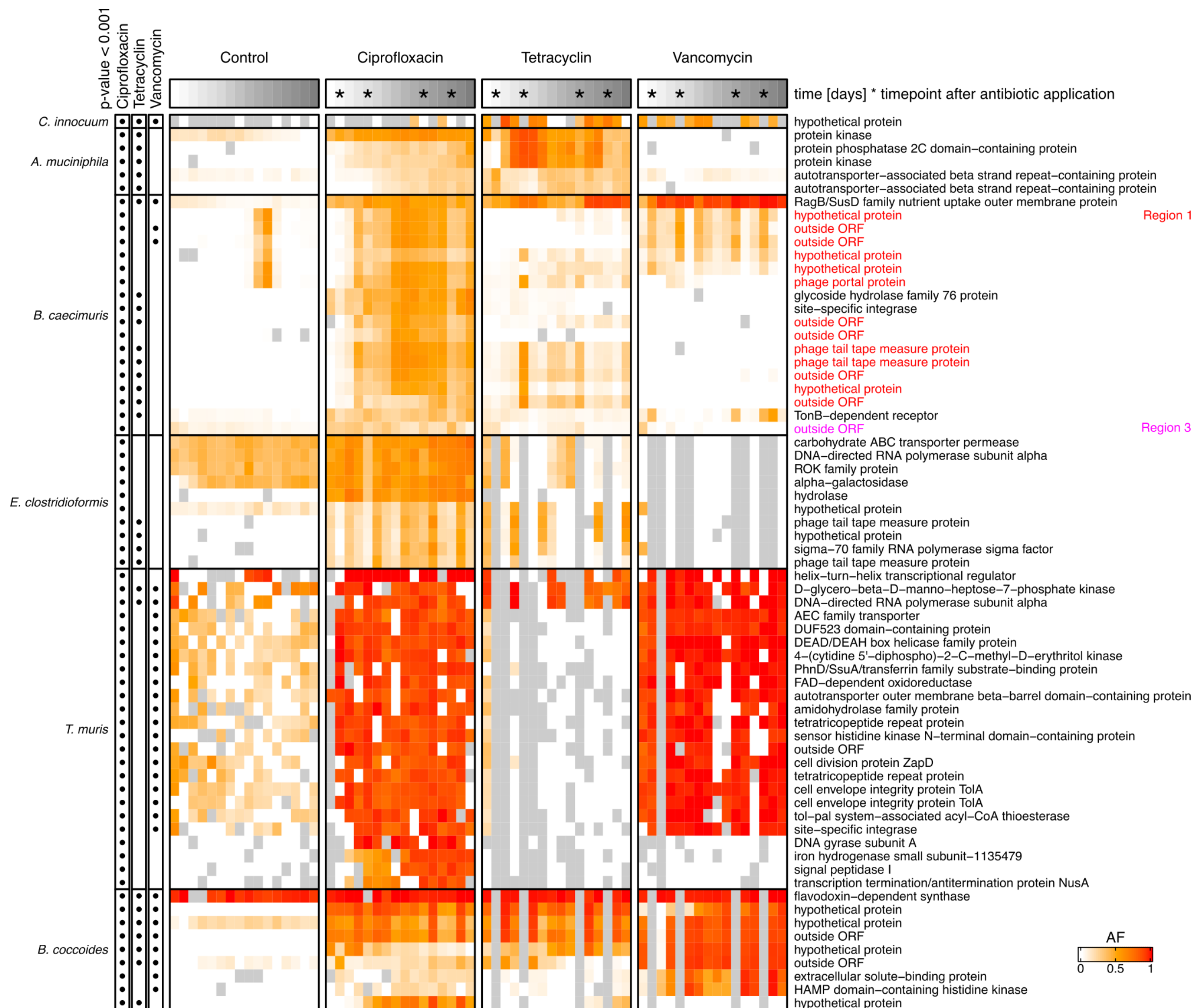

Figure 2: Allele frequency of 67 SNPs that are varying significantly between control and at least one treatment group, indicating substrain selection or *de novo* SNP acquisition. Technical zeros, defined when the read count on a location is lower than 25, are shown as grey. SNPs that are present in less than 8 samples, have an effect size below 0.1, or have a greater q-value of 0.01 or belong to the intermediate AF group are filtered out. Text in color indicates that the locations are on one of the differentially coverage regions (see **Fig. 4**)

AB treatment modulates bacterial growth

SNP analysis identified potential resistance-conferring mutations that could explain qPCR patterns of selected bacteria such as the resistance development in *C. inoccum*, it seems that SNP presence/absence alone cannot explain the more complex response pattern we see after treatment in community members such as *A. muciniphila*. Since we had the controlled experimental setting and high-quality reference genomes at hand, we searched for other types of signatures that could relate to development of AB tolerance, such as change in growth rate. Since DNA replication starts from a fixed origin in the circular genome and proceeds bi-directionally, the coverage distribution of a metagenomic sample mapped to the reference genome can provide insights into changes in bacterial growth rate (**Fig 3A)** (Korem et al. 2015).

We used an in-house developed R package (*OligoMMR*) to identify differences in bacterial growth rates of specific OMM[12] community members using metagenomic sequence data. This method is based on determining sequencing coverage between the peak (origin of replication)

and trough (opposite region in the circular bacterial genome) and provides a quantitative measure of a species' growth (P2T; **Fig. 3A**) (Korem et al. 2015). Using this P2T analysis, we found evidence that antibiotic treatment modulates growth of OMM[12] community members (**Fig. 3**). While there are less pronounced differences during the study duration on the estimated growth rates for bacteria such as *F. plautii* and *B. caecimuris* we observe a treatment-dependent effect for others (**Fig. 3BC**). Especially during vancomycin treatment, growth rates drastically decline on the adjacent time point after AB application. For *A. muciniphila*, while the bacteria seem largely unaffected by the antibiotics according to the qPCR profile (**Fig. 1B**), differences in growth rate are observed in response to AB treatments (**Fig. 3B**). For example, tetracycline treatment, *A. muciniphila* P2T declines sharply and recover during the resilience phase While the drop in bacterial abundance is often mostly pronounced in the adiacent time point (after 1 week of continuous AB treatment), drop in P2T seems to persist even longer, where the growth rate is lowest after the second time point we took after treatment (~2 weeks). In the vancomycin group, we see after treatment a complete stop in bacterial growth in the two next time points taken after treatment (**Fig. 1** bottom, **B**), which manifests in a drop in absolute qPCR in these time points. Interestingly, all following treatment regimes influence growth rate positively, and no clear drop in bacterial cell count is observed, indicating that selection effects affect not only bacterial survival but also growth rate. Taken together this illustrates the requirement that to understand adaptations processes to antibiotics, community dynamics should be taken into account.

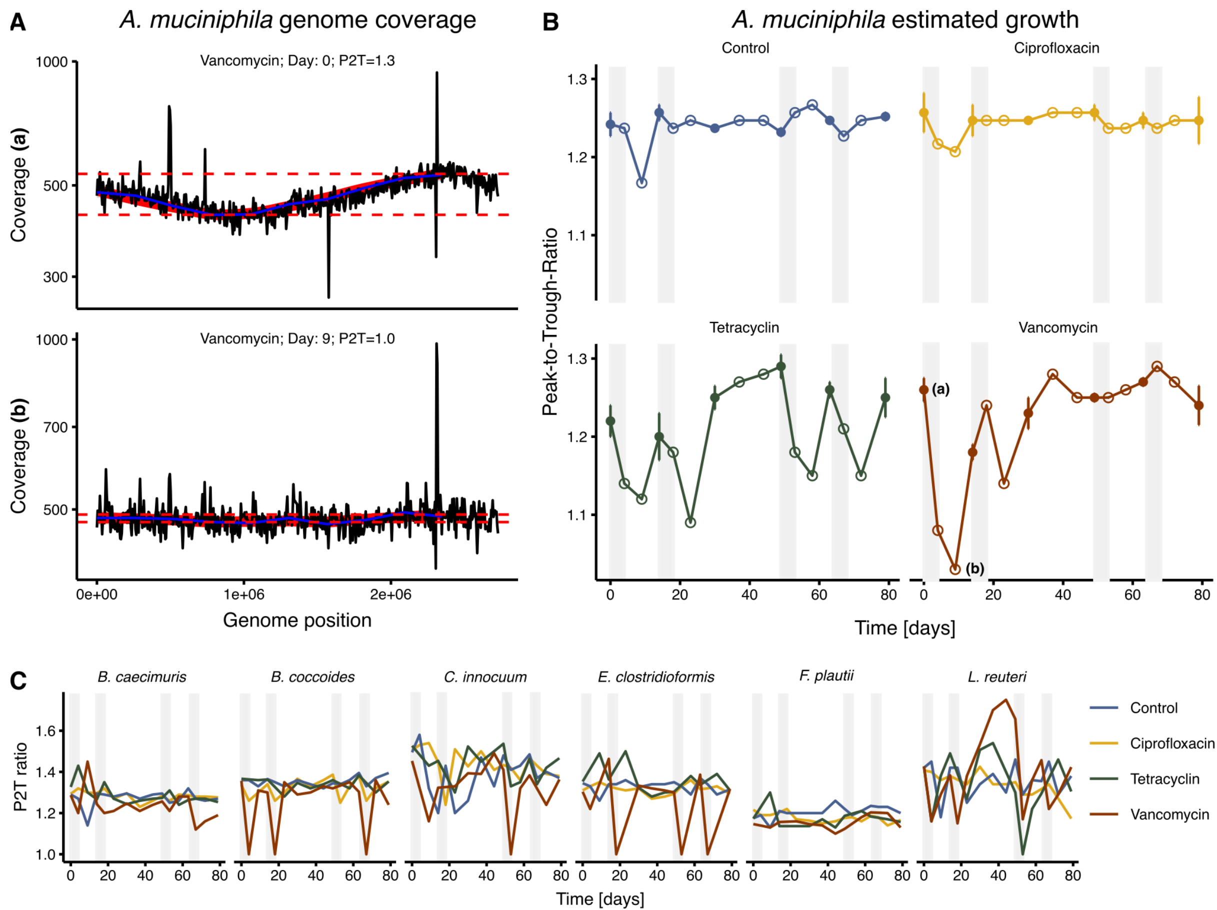

Figure 3: Estimated bacterial growth changes during study time in response to treatment regimes. A) Coverage profile of the metagenomic reads mapped to the *A. muciniphila* reference genome. The coverage distribution (top) shows a clear coverage pattern skewed by the DNA replication process. On a subset of samples,

this coverage distribution is less pronounced (bottom), possible due to reduced growth. Red lines indicate “peak” and “through” points based on the smoothed coverage profile (solid red and blue lines).The peak-to-trough ratio (P2T) is a correlated for bacterial growth. B) Aggregated P2T ratios for *A. muciniphila* stratified by treatment group. While for the control group treatment regimes (grey areas) are not associated with change in growth rate, for tetracycline and vancomycin the growth rate is typically lower after treatment. C) Aggregated P2T ratios for selected community members under different treatment regimes.

Regions with differential coverage

We have shown that community adaptation to antibiotics is linked to SNP acquisition and selection and change in growth rates. Next, we also wanted to address if changes in prophage induction (Modi et al. 2013). Lysogenic phages are integrated in the bacterial genome as prophage but get activated in response to diverse environmental stimuli including antibiotics to excise from the genome and produce phage progeny (ref). Using similarity search of the reference OMM[12] collection to known phage proteins (Arndt et al. 2016), we identified 58 putative prophage regions (found across 11 strains), of which 5 were marked as intact, 43 as incomplete and 10 as incomplete, **Supplementary Table 1**).

We observe abnormal low read coverage near some of the predicted prophage regions (**Fig 4A**, grey bars), which could be explained by loss of prophage regions *in vivo* within the study time or parental generations in respect to the reference genome, highlighting that prophage population OMM[12] is vastly evolving. Prophage activation would be indicated by higher read coverage in a predicted prophage region. We found several examples for predicted prophages within *B. caecimurius*. (**Fig. 4A**). Interestingly, an increase in coverage is observed at time points collected after AB-treatments, for example at day 4 and day 9, while coverage of the region was not increased at other time points (initial time point and time points after treatment)(**Fig. 4B**). We quantified the difference between coverage inside the loci vs. the adjacent region of the bacterial chromosome by calculating the fold-change in coverages (**Fig. 4B**, red dotted line and black dotted line).

We identified two further loci where we observe an increase in coverage in respect to the reference genome, again potentially due to prophage induction in the bacterial population(**Fig. 4A**). This time, fold-change of these regions are associated with specific treatment regimes (**Fig. 4C**), for example, the fold-change of region 2 seems stable in the control, ciprofloxacin and vancomycin group, but increased during study time in the tetracycline group, indicating that the presence of multiple copies are advantageous for the bacterial fitness only under tetracycline treatment. Fold change differences of region 3 are particularly pronounced at the last time points in the vancomycin group (**Fig. 4C**) and seem associated with a drop in growth rate (**Fig. 3C**). In total, region 1 contains 68 ORFs, some of which are bacteriophage associated, such as “phage tail tape measure protein”, “phage portal protein”, methyltransferases and reverse transcriptases (**Supplementary Table 2**). However, the remaining content of the loci is poorly annotated, but this prophage further contains “ParB/RepB/Spo0J family partition protein” usually associated with low copy number plasmids and “N-acetylmuramoyl-L-alanine amidase” which have found to be associated with other prophages and play a major role in lysis (Longchamp, Mauël, and Karamata 1994).

We identified a region on the chromosome exhibiting a differential and treatment-group dependent coverage, (Region 1) that encode a ADP-ribosylglycohydrolase family protein, essential for ADP-ribosylation in viruses and bacteria (Catara et al. 2019) linked to defence to antibiotic stress (Baysarowich et al. 2008) as well as linked to the toxin-antitoxin (TA) systems that are required for the development of persister cells that survive antibiotic stress by downregulating bacterial growth and upregulating beneficial phenotypes in these environments such as biofilm formations (Catara et al. 2019). This finding is particularly interesting in the

context of change in bacterial growth quantified earlier since low growth rates could highlight the presence of persister cells.

We further identified 2 regions with differentially coverage, Region 2 and 3, encoding 32 and 39 proteins, respectively. While region 2 encodes a range of transposase, it is likely associated with a mobile genetic element, region 3 contains no known phage gene for transposase gene, leaving the mode of propagation unknown. Interestingly, there are several similarities between region 2 and 3, for example both encode a efflux RND transporter permease and a efflux RND transporter periplasmic adaptor subunit and a TolC family protein. These efflux pumps are known to confer multidrug resistance on bacteria by transporting a wide spectrum of structurally diverse antibiotics (Venter et al. 2015) via the outer-membrane channel TolC (Pradel and Pagès 2002). In region 2, further membrane associated genes belong to RagB/SusD as well as SusF/SusE, and TonB-dependent receptors as well as MATE efflux transporters, while region 3 contains ABC transporters. Taken together, this indicates that subpopulations exist with varying numbers of copies of these regions 1-3 producing cells with a different number of membrane transporters, and cells that have multiple copies in their genome are more resistant under antibiotic stress due to optimized transport of the antibiotic compound out of the cell.

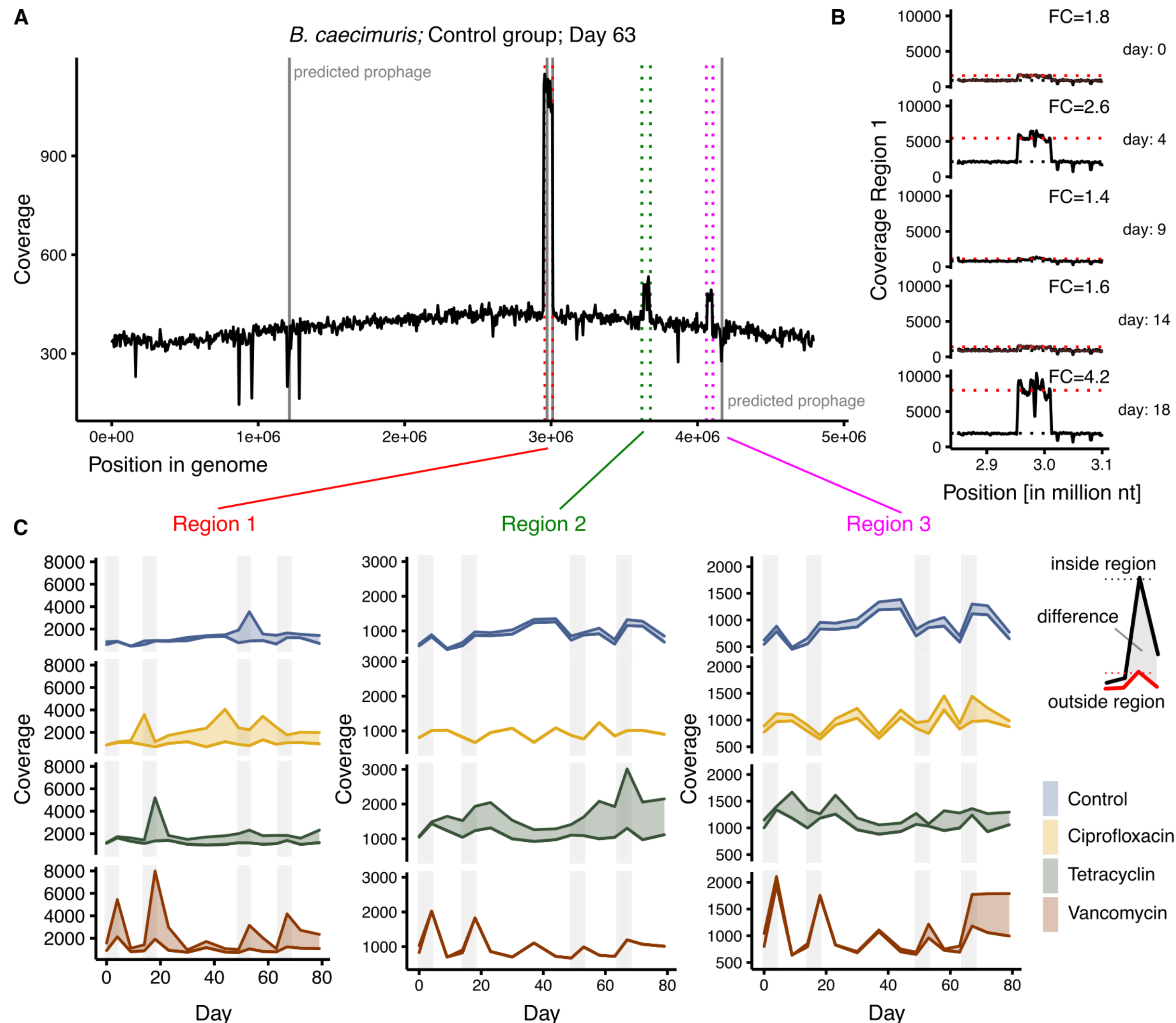

Figure 4: Coverage of prophage loci changes during treatment. A) On *B. caecimuris* we identified 3 regions with high similarity to known prophages (grey vertical lines) that align with regions where we see a particular high or low read coverage. Based on the coverage profile we identified 3 loci where the abundance is high on a subset of samples (colored vertical lines). B) For region 1 we found on the genome (marked red) shows a differential coverage where high changes in coverage of the loci in respect to its neighborhood are present after vancomycin treatment (i.e. on day 4 and 18) but a low fold change is present before treatment (day 0) or during resilience (day 9 and 14). C) Fold changes (FC) in coverage of the loci compared to the coverage adjacent to the loci is shown as filled region, displayed with coverage outside the region (lower line) and coverage of the loci (upper line). While region 1 (red) shows a high fold-change after treatment in the vancomycin group, region 2 (green) shows a higher and time-dependent fold change in the tetracycline group. Region 3 shows a high fold change on the vancomycin group after the 4th treatment round.

Reisolation of evolved bacterial community members confirm changes in antibiotic sensitivity

To assess if the evolved strains exhibit altered sensitivity to the respective antibiotics(i.e. acquired resistance to the applied antibiotics during the course of study), we picked 45 single colonies from the evolved community taken from the last time point from the cecal content for *E. faecalis*, *C. innocuum* and *B. coccoides*. Variant calling was performed using the same workflow as for the WGS samples but with mapping against the respective individual reference genome. Next generation sequencing retrieved on average 15.3M reads per sample.

We expected that AF of the observed variants (SNPs) in these samples were either near zero or 1. However, especially in *C. innocuum* we observed a multitude of variants with an intermediate AF of ~0.5. Interestingly, the coverage within these regions was elevated by the factor ~2 (**Fig. 5A**) which is indicative of a duplication event, where one copy has evolved and acquired a SNP but the other remained unchanged in the reference genome. A value slightly below 2 may arise due to the fact that one of the duplicated regions accumulated more mutations and due to the stringent mapping, not all reads were mapped to the reference genome. AF values within these regions are also intermediate in the metagenomic shotgun samples. We conclude thatAF around 0.5 should be interpreted with caution and have therefore filtered all metagenomic shotgun samples for the respective genomic regions.

Variant counts differ substantially between the colonies and we identified on average ### variants per sample for *E. faecalis* (**Fig. 5B**). This high variability is mainly driven by a subset of *E. faecalis* samples that have an abnormal high SNP load. On *C. innocuum*, we observe an overlap of variants that we detect in the MSG and variants that we detect when processing the isolates alone. As abundance of E. faecalis was too low in MGS, we could not compare the datasets side by side. Phenotypic characterization of a subset of isolates revealed that they exhibit different MIC values. In particular, SNPs in the 30S ribosomal protein gene S10 in *C.*

innocuum correlated with increased MIC against tetracycline.

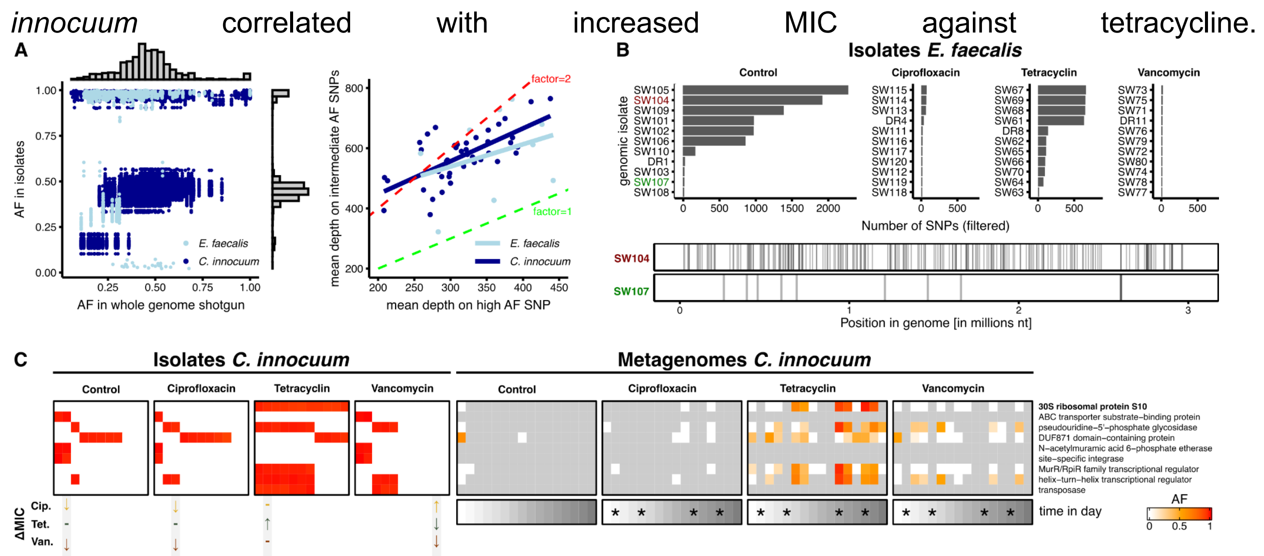

Figure 5: SNPs observed on the 30S ribosomal protein lead to a de novo acquired resistance against tetracycline. A) A subset of variants exhibit an intermediate AF on the isolates (y axis) also exhibit an intermediate AF in the WGS samples (left), leading to an inflated estimate of variants if not controlled appropriately. On the isolates, the coverage of the problematic variants is higher compared with variants near 1 or zero (right), indicating that these problematic regions are duplicated regions. B) Number of SNPs observed in the isolates differ substantially (top), and samples such as SW104 that have a high number of SNPs have these scattered throughout the genome (bottom). C) A subset of mutations that are present in the isolates (left) and metagenomes (right). Grey regions indicate the presence of a technical zero (read count on a location is lower than 25). MIC subplot indicates if the inhibitory concentration of the evolved community is higher (arrow up) or lower (arrow down) or similar ("-") compared to unevolved stock samples.

Set of highly correlated SNPs can be decomposed into genotypes using NMF

Major allele frequency distributions of variants observed over the study time (**Fig 2**) show evidence that some of the OMM[12] strains appear to be non-monoclonal and species have adapted within the host over generations, resulting in sets of linked variants. Furthermore, due to often low coverage and low sequence divergence, direct assembly of substrains show no clear separation of genomes on the substrain boundary. However, the non-random linkage of SNPs over time points, and oftentimes a positive correlation of allele frequency (AF) of variants between study groups (**Fig. 2**) support validity of the detected variants and the presence of haplotypes in the study population. To detect haplotypes we developed a method based on sparse non-negative matrix factorization (**Methods**) which detects co-occurring linear combinations of the SNPs profile of a bacteria in multiple samples (i.e. genotypes) and calculates the fraction of explained variance for each decomposed haplotype to the overall sample's variance in AF. Using this approach we successfully recovered co-occurring SNP combinations. The reconstructed haplotypes show a strong time and treatment-dependent effect.

Discussion

In this study, we present the community effects of a defined mouse microbiota to repeated applications to antibiotics. We identified selection effects and potential *de novo* mutation on the SNP level. Furthermore, estimated growth rates show a drop in growth rate after AB application, potentially due to the selection of persister cells that are evolved to exhibit a slow growth rate

but are physiologically more resistant to a wide range of stressors. Finally, re-isolation and sequencing of cells from the evolved community show that some SNPs reach fixation, and that carriers of SNPs are resistant to antibiotics and exhibit a higher MIC concentration, potentially due to mutations in ribosomal genes that prevent antibiotic binding (Jacoby 2005). This is one of the first strain-level genome based analyses on the evolution of AB resistance in the gut microbiome.

We present a range of caveats that are important to consider when performing similar experiments using low-complex community models such as the OMM^{12}. First, we observe a high similarity within chromosomal regions of the OMM^{12} community that would result in a highly inflated SNP profile. One solution to account for these effects would be masking of regions with high similarity to other regions or by limiting the mapping to unique locations (e.g. not consider reads that would map to multiple positions). Furthermore, we found that absolute abundance measures, as provided by qPCR measures (especially on a $\log_{10}$ scale), are intuitive to understand and response effects are harder to understand on a relative scale (**Fig. 1BC**). Duplication events complicate SNP analysis, since we observe that duplication events lead to an inflated SNP count whereby many of the SNPs exhibit intermediate AF. However, we see that copy number variations of specific loci such as genes seem to be a biological process linked with AB selection and filtering of these variants are non-trivial.

It was shown that bacterial persisters form a subpopulation of antibiotic-tolerant cells that form within a population of genetically antibiotic-susceptible bacteria and multiple genetic pathways involved in persister formation were identified (Harms et al. 2017), such as RelA, SpoT, the protease Lon, and toxin-antitoxin (TA) modules (Cohen, Lobritz, and Collins 2013), which release toxins to inhibit cellular processes in turn slowing down protein synthesis (Korch, Henderson, and Hill 2003). Interestingly, we see a SNP within the coding region of the toxin-antitoxin (TA) system in *F. plautii*, but, however, no significant difference of the SNP in the treatment groups in respect to control. Furthermore, some genes encoded on the region 1 where we observe differences in read counts, potentially due to phage activation, encode genes related to the TA system which could trigger a persister phenotype. A reduced growth rate after treatment could be attributable to the selection of persister cells that are resistant but exhibit a significantly reduced growth rate. The growth rates we observe would be in line with the selection of persister cells after treatment.

Observed ABR sensitivity of community members on the population level is in line with MIC concentrations observed in monocultures (**Methods, Supplementary Figure S1**). Difference in growth rate is a further factor that could cause the observed response patterns. Bacteria with relatively slow growth rates in monocultures (**Supplementary Table 5**), for example, *B. caecimuris* and *A. muciniphila* show intrinsic resistance against all three AB used in this study. ***

It was shown that AB enhances the genome-wide mutation rates (Long et al. 2016). We observe this phenomenon on the isolates of evolved community members of *E. faecalis*, but also on the control group and tetracycline group but not on ciprofloxacin and vancomycin groups. ***

, potentially due to the presence of subpopulations with a hypermutator phenotype where the mutations are scattered around the whole genome (**Fig. 5B** bottom) or be of a technical origin.

Understanding the factors contributing to resilience and dysbiosis are crucial for the success of faecal transplantation and therapeutic manipulation of the gut microbiome (Kootte et al. 2012). Furthermore, this study will provide guidance on how to tackle technical challenges that arise when analysing metagenomic content of gnotobiotic mouse models. Previous studies

characterized the short and long term effect of compositional and functional changes caused by antibiotic treatment in mice in the context of high-fat diet and obesity (Nobel et al. 2015) and identified effects of AB to high-fat diet induced obesity. The effects of AB treatment have also been studied and showed that antibiotic pressure resulted in reproducible, long-lasting alterations of the gut microbiome, including a decrease in overall diversity (Antonopoulos et al. 2009) and in the context of susceptibility to infections (Theriot et al. 2014). While the above mentioned studies are valuable, current studies on AB are using conventional mouse microbiomes, which has been shown to be a confounding factor, since mice raised in different research institutions or obtained from different vendors can exhibit profound differences in microbiota composition (Stecher et al. 2010). Gnotobiology is a method to mechanistically investigate microbiota functioning and to assess causality in alterations of gut microbiota composition. The use of gnotobiotic mouse models could be valuable to study community adaptation and the consequences of AB treatment in a mechanistic manner, for example, by controlling for interaction networks, such as cross-feeding or competition (Weiss et al. 2021). In particular, it is unclear, to which extent the microbiota returns to the pre-antibiotic state and which factors play a role in mediating resilience such as environment such as diet and the role of bacteria acquiring AB resistance as a consequence.

Methods

Animal experiments

In order to effectively capture the adaptation of the bacterial community to the AB challenge, the study design included a total of 4 treatment periods, interspaced by resilience phases of varying lengths ranging from 5 days up to 35 days, which allowed us to study the influence of repeated applications of AB to the microbial community dynamics. We treated adult mice stably colonized by the defined OMM^{12} community with physiologically relevant concentrations of antibiotics, vancomycin (Roth, 0242.3), tetracyclin (Sigma-Aldrich, T7660) and ciprofloxacin (Sigma-Aldrich, 17850) via drinking water (0,167mg/ml) and housed them under germ-free conditions (5 mouse per study group) or control (**Fig. 1A**). We observed no difference in the quantity of water consumed between the groups. During the study period of 80 days, fecal samples from each mouse were collected on a weekly basis, amounting to a total of 307 samples. All animal experiments were approved by the local authorities (Regierung von Oberbayern) and an ethics committee, and were performed according to the legal requirements.

Quantitative PCR of bacterial 16S rRNA genes

We determined the absolute abundance of the individual strains of the OMM^{12} consortium at different time points via qPCR as outlined in Brugiroux et al. (Brugiroux et al. 2016).

Reisolation of evolved strains

Cecal content from OMM^{12} mice was taken and diluted with sterile PBS (dilution series). Several dilutions were spread on AAM agar plates to obtain single colonies Colonies were picked based on colony morphology and used for inoculation of liquid AAM cultures. KB1 clones were grown aerobically as this is the only bacterium in the community that can grow aerobically. I46 colonies were restreaked on fresh plates to obtain single colonies. Cultures were analyzed via Gram-staining to check for contaminations. Cultures were used to stock the isolated clones.

Residual 8 ml cultures were used for chromosomal DNA isolation using the Turnbaugh method (bead bashing + phenol chloroform extraction).

Genome sequencing, mapping and variant calling

We employed shotgun NovaSeq pyrosequencing (~6.8 Mbases) per sample on a subset of 112 samples on selected time points (WGS profile of all time points for one mouse per group plus two biological replicates per group of selected time points) (**Fig. 1A**). We analysed sequenced samples via a snakemake pipeline (Köster and Rahmann 2012). This includes QC'ing reads via fastp using default arguments (S. Chen et al. 2018). We mapped all WGS samples to the combined reference genome collection. This collection consists of high-quality reference genomes (Lamy-Besnier et al. 2021). These reference genomes are optimized using chromosome conformation capture (Hi-C) data to reorganize, close, and improve the quality of these 12 genomes, based on the reference genomes we have published recently (Garzetti et al. 2017). We use bwa mem with the "-x intractg" to map QC'ed reads to the reference genomes. Mapping fraction to the reference set was very high (>95%). We then removed duplicated reads and reads that map to multiple positions using "samtools fixmate - -" and " samtools view -bq 1". Variant calling was performed using LoFreq (Wilm et al. 2012) using the options "lofreq viterbi" and "lofreq indelqual --dindel" and "lofreq call-parallel --sig 1E-4 --min-cov 25 --min-bq 25 --min-alt-bq 25 --min-mq 60 -d 10000".

Coverage quantification

We calculated coverage per sample using the "bamCoverage" command on the "bam" files that are written by the "lofreq viterbi" and "lofreq indelqual" step. These .bam files contain the alignment information of reads after removing exact duplicates and reads that are mapping to multiple positions.

Acknowledgements

P.C.M. received funding from the German Research Foundation (Grant number 315980449).

References

Antonopoulos, Dionysios A., Susan M. Huse, Hilary G. Morrison, Thomas M. Schmidt, Mitchell L. Sogin, and Vincent B. Young. 2009. "Reproducible Community Dynamics of the Gastrointestinal Microbiota Following Antibiotic Perturbation." *Infection and Immunity* 77 (6): 2367–75.

Arciola, Carla Renata, Davide Campoccia, and Lucio Montanaro. 2018. "Implant Infections: Adhesion, Biofilm Formation and Immune Evasion." *Nature Reviews. Microbiology* 16 (7): 397–409.

Arndt, David, Jason R. Grant, Ana Marcu, Tanvir Sajed, Allison Pon, Yongjie Liang, and David S. Wishart. 2016. "PHASTER: A Better, Faster Version of the PHAST Phage Search Tool." *Nucleic Acids Research* 44 (W1): W16–21.

Barroso-Batista, João, Miguel F. Pedro, Joana Sales-Dias, Catarina J. G. Pinto, Jessica A. Thompson, Helena Pereira, Jocelyne Demengeot, Isabel Gordo, and Karina B. Xavier.

2020. “Specific Eco-Evolutionary Contexts in the Mouse Gut Reveal Escherichia Coli Metabolic Versatility.” *Current Biology: CB* 30 (6): 1049–62.e7.
Baysarowich, Jennifer, Kalinka Koteva, Donald W. Hughes, Linda Ejim, Emma Griffiths, Kun Zhang, Murray Junop, and Gerard D. Wright. 2008. “Rifamycin Antibiotic Resistance by ADP-Ribosylation: Structure and Diversity of Arr.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (12): 4886–91.
Brugiroux, Sandrine, Markus Beutler, Carina Pfann, Debora Garzetti, Hans-Joachim Ruscheweyh, Diana Ring, Manuel Diehl, et al. 2016. “Genome-Guided Design of a Defined Mouse Microbiota That Confers Colonization Resistance against Salmonella Enterica Serovar Typhimurium.” *Nature Microbiology* 2 (November): 16215.
Casiraghi, Andrea, Lorenzo Suigo, Ermanno Valoti, and Valentina Straniero. 2020. “Targeting Bacterial Cell Division: A Binding Site-Centered Approach to the Most Promising Inhibitors of the Essential Protein FtsZ.” *Antibiotics (Basel, Switzerland)* 9 (2). https://doi.org/10.3390/antibiotics9020069.
Catara, Giuliana, Annunziata Corteggio, Carmen Valente, Giovanna Grimaldi, and Luca Palazzo. 2019. “Targeting ADP-Ribosylation as an Antimicrobial Strategy.” *Biochemical Pharmacology* 167 (September): 13–26.
Chen, Hong-Han, Chien-Che Chang, Yu-Han Yuan, and Shwu-Jen Liaw. 2020. “A CpxR-Regulated zapD Gene Involved in Biofilm Formation of Uropathogenic Proteus Mirabilis.” *Infection and Immunity* 88 (7). https://doi.org/10.1128/IAI.00207-20.
Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics* 34 (17): i884–90.
Cohen, Nadia R., Michael A. Lobritz, and James J. Collins. 2013. “Microbial Persistence and the Road to Drug Resistance.” *Cell Host & Microbe* 13 (6): 632–42.
Croswell, Amy, Elad Amir, Paul Teggatz, Melissa Barman, and Nita H. Salzman. 2009. “Prolonged Impact of Antibiotics on Intestinal Microbial Ecology and Susceptibility to Enteric Salmonella Infection.” *Infection and Immunity* 77 (7): 2741–53.
Droste, J. H., M. H. Wieringa, J. J. Weyler, V. J. Nelen, P. A. Vermeire, and H. P. Van Bever. 2000. “Does the Use of Antibiotics in Early Childhood Increase the Risk of Asthma and Allergic Disease?” *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology* 30 (11): 1547–53.
Durand-Heredia, Jorge, Eugene Rivkin, Guoxiang Fan, Jorge Morales, and Anuradha Janakiraman. 2012. “Identification of ZapD as a Cell Division Factor That Promotes the Assembly of FtsZ in Escherichia Coli.” *Journal of Bacteriology* 194 (12): 3189–98.
Eberl, Claudia, Diana Ring, Philipp C. Münch, Markus Beutler, Marijana Basic, Emma Caroline Slack, Martin Schwarzer, et al. 2019. “Reproducible Colonization of Germ-Free Mice With the Oligo-Mouse-Microbiota in Different Animal Facilities.” *Frontiers in Microbiology* 10: 2999.
El Qaidi, Samir, Congrui Zhu, Peter McDonald, Anuradha Roy, Pradip Kumar Maity, Digamber Rane, Chamani Perera, and Philip R. Hardwidge. 2018. “High-Throughput Screening for Bacterial Glycosyltransferase Inhibitors.” *Frontiers in Cellular and Infection Microbiology* 8 (December): 435.
Faith, Jeremiah J., Janaki L. Guruge, Mark Charbonneau, Sathish Subramanian, Henning Seedorf, Andrew L. Goodman, Jose C. Clemente, et al. 2013. “The Long-Term Stability of the Human Gut Microbiota.” *Science* 341 (6141): 1237439.
Garzetti, Debora, Sandrine Brugiroux, Boyke Bunk, Rüdiger Pukall, Kathy D. McCoy, Andrew J. Macpherson, and Bärbel Stecher. 2017. “High-Quality Whole-Genome Sequences of the Oligo-Mouse-Microbiota Bacterial Community.” *Genome Announcements* 5 (42). https://doi.org/10.1128/genomeA.00758-17.
Godreuil, Sylvain, Marc Galimand, Guy Gerbaud, Christine Jacquet, and Patrice Courvalin. 2003. “Efflux Pump Lde Is Associated with Fluoroquinolone Resistance in Listeria

Monocytogenes." *Antimicrobial Agents and Chemotherapy* 47 (2): 704–8.
Grass, Lena M., Jan Wollenhaupt, Tatjana Barthel, Iwan Parfentev, Henning Urlaub, Bernhard Loll, Eberhard Klauck, Haike Antelmann, and Markus C. Wahl. 2021. "Large-Scale Ratcheting in a Bacterial DEAH/RHA-Type RNA Helicase That Modulates Antibiotics Susceptibility." *Proceedings of the National Academy of Sciences of the United States of America* 118 (30). https://doi.org/10.1073/pnas.2100370118.
Harms, Alexander, Cinzia Fino, Michael A. Sørensen, Szabolcs Semsey, and Kenn Gerdes. 2017. "Prophages and Growth Dynamics Confound Experimental Results with Antibiotic-Tolerant Persister Cells." *mBio* 8 (6). https://doi.org/10.1128/mBio.01964-17.
Human Microbiome Project Consortium. 2012. "Structure, Function and Diversity of the Healthy Human Microbiome." *Nature* 486 (7402): 207–14.
Jacoby, George A. 2005. "Mechanisms of Resistance to Quinolones." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 41 Suppl 2 (July): S120–26.
Jiang, Xiaobing, Tao Yu, Ping Xu, Xiaobo Xu, Shengdong Ji, Wujun Gao, and Lei Shi. 2018. "Role of Efflux Pumps in the in Vitro Development of Ciprofloxacin Resistance in Listeria Monocytogenes." *Frontiers in Microbiology* 9 (September): 2350.
Klünemann, Martina, Sergej Andrejev, Sonja Blasche, Andre Mateus, Prasad Phapale, Saravanan Devendran, Johanna Vappiani, et al. 2021. "Bioaccumulation of Therapeutic Drugs by Human Gut Bacteria." *Nature* 597 (7877): 533–38.
Kootte, R. S., A. Vrieze, F. Holleman, G. M. Dallinga-Thie, E. G. Zoetendal, W. M. de Vos, A. K. Groen, J. B. L. Hoekstra, E. S. Stroes, and M. Nieuwdorp. 2012. "The Therapeutic Potential of Manipulating Gut Microbiota in Obesity and Type 2 Diabetes Mellitus." *Diabetes, Obesity & Metabolism* 14 (2): 112–20.
Korch, Shaleen B., Thomas A. Henderson, and Thomas M. Hill. 2003. "Characterization of the hipA7 Allele of Escherichia Coli and Evidence That High Persistence Is Governed by (p)ppGpp Synthesis." *Molecular Microbiology* 50 (4): 1199–1213.
Korem, Tal, David Zeevi, Jotham Suez, Adina Weinberger, Tali Avnit-Sagi, Maya Pompan-Lotan, Elad Matot, et al. 2015. "Growth Dynamics of Gut Microbiota in Health and Disease Inferred from Single Metagenomic Samples." *Science* 349 (6252): 1101–6.
Köster, Johannes, and Sven Rahmann. 2012. "Snakemake—a Scalable Bioinformatics Workflow Engine." *Bioinformatics* 28 (19): 2520–22.
Lamy-Besnier, Quentin, Romain Koszul, Laurent Debarbieux, and Martial Marbouty. 2021. "Closed and High-Quality Bacterial Genome Sequences of the Oligo-Mouse-Microbiota Community." *Microbiology Resource Announcements* 10 (17). https://doi.org/10.1128/MRA.01396-20.
Lange, Kathleen, Martin Buerger, Andreas Stallmach, and Tony Bruns. 2016. "Effects of Antibiotics on Gut Microbiota." *Digestive Diseases* 34 (3): 260–68.
Ley, Ruth E., Daniel A. Peterson, and Jeffrey I. Gordon. 2006. "Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine." *Cell* 124 (4): 837–48.
Lloyd-Price, Jason, Galeb Abu-Ali, and Curtis Huttenhower. 2016. "The Healthy Human Microbiome." *Genome Medicine* 8 (1): 51.
Longchamp, P. F., C. Mauël, and D. Karamata. 1994. "Lytic Enzymes Associated with Defective Prophages of Bacillus Subtilis: Sequencing and Characterization of the Region Comprising the N-Acetylmuramoyl-L-Alanine Amidase Gene of Prophage PBSX." *Microbiology* 140 (Pt 8) (August): 1855–67.
Long, Hongan, Samuel F. Miller, Chloe Strauss, Chaoxian Zhao, Lei Cheng, Zhiqiang Ye, Katherine Griffin, et al. 2016. "Antibiotic Treatment Enhances the Genome-Wide Mutation Rate of Target Cells." *Proceedings of the National Academy of Sciences of the United States of America* 113 (18): E2498–2505.
Maier, Lisa, Camille V. Goemans, Jakob Wirbel, Michael Kuhn, Claudia Eberl, Mihaela

Pruteanu, Patrick Müller, et al. 2021. “Unravelling the Collateral Damage of Antibiotics on Gut Bacteria.” *Nature*, October, 1–5.
Modi, Sheetal R., Henry H. Lee, Catherine S. Spina, and James J. Collins. 2013. “Antibiotic Treatment Expands the Resistance Reservoir and Ecological Network of the Phage Metagenome.” *Nature* 499 (7457): 219–22.
Nagpal, Ravinder, Tiffany M. Newman, Shaohua Wang, Shalini Jain, James F. Lovato, and Hariom Yadav. 2018. “Obesity-Linked Gut Microbiome Dysbiosis Associated with Derangements in Gut Permeability and Intestinal Cellular Homeostasis Independent of Diet.” *Journal of Diabetes Research* 2018 (September): 3462092.
Ng, Katharine Michelle, Andrés Aranda-Díaz, Carolina Tropini, Matthew Ryan Frankel, William Van Treuren, Colleen T. O’Loughlin, Bryan Douglas Merrill, et al. 2020. “Recovery of the Gut Microbiota after Antibiotics Depends on Host Diet, Community Context, and Environmental Reservoirs.” *Cell Host & Microbe* 28 (4): 628.
Nobel, Yael R., Laura M. Cox, Francis F. Kirigin, Nicholas A. Bokulich, Shingo Yamanishi, Isabel Teitler, Jennifer Chung, et al. 2015. “Metabolic and Metagenomic Outcomes from Early-Life Pulsed Antibiotic Treatment.” *Nature Communications* 6 (June): 7486.
Pradel, Elizabeth, and Jean-Marie Pagès. 2002. “The AcrAB-TolC Efflux Pump Contributes to Multidrug Resistance in the Nosocomial Pathogen Enterobacter Aerogenes.” *Antimicrobial Agents and Chemotherapy* 46 (8): 2640–43.
Ramiro, Ricardo S., Paulo Durão, Claudia Bank, and Isabel Gordo. 2020. “Low Mutational Load and High Mutation Rate Variation in Gut Commensal Bacteria.” *PLoS Biology* 18 (3): e3000617.
Schutzbach, John, and Inka Brockhausen. 2009. “Inhibition of Glycosyltransferase Activities as the Basis for Drug Development.” *Methods in Molecular Biology* 534: 359–73.
Silber, Nadine, Cruz L. Matos de Opitz, Christian Mayer, and Peter Sass. 2020. “Cell Division Protein FtsZ: From Structure and Mechanism to Antibiotic Target.” *Future Microbiology* 15 (June): 801–31.
Sommer, Felix, Jacqueline Moltzau Anderson, Richa Bharti, Jeroen Raes, and Philip Rosenstiel. 2017. “The Resilience of the Intestinal Microbiota Influences Health and Disease.” *Nature Reviews. Microbiology* 15 (10): 630–38.
Stecher, Bärbel, Samuel Chaffron, Rina Käppeli, Siegfried Hapfelmeier, Susanne Freedrich, Thomas C. Weber, Jorum Kirundi, et al. 2010. “Like Will to like: Abundances of Closely Related Species Can Predict Susceptibility to Intestinal Colonization by Pathogenic and Commensal Bacteria.” *PLoS Pathogens* 6 (1): e1000711.
Thangaraju, Pugazhenthan, and Sajitha Venkatesan. 2019. “WHO Ten Threats to Global Health in 2019: Antimicrobial Resistance.” *Cukurova Medical Journal* 44 (3): 1150–51.
Theriot, Casey M., Mark J. Koenigsknecht, Paul E. Carlson Jr, Gabrielle E. Hatton, Adam M. Nelson, Bo Li, Gary B. Huffnagle, Jun Z Li, and Vincent B. Young. 2014. “Antibiotic-Induced Shifts in the Mouse Gut Microbiome and Metabolome Increase Susceptibility to Clostridium Difficile Infection.” *Nature Communications* 5: 3114.
Turnbaugh, Peter J., Fredrik Bäckhed, Lucinda Fulton, and Jeffrey I. Gordon. 2008. “Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome.” *Cell Host & Microbe* 3 (4): 213–23.
Ur Rahman, Mujeeb, Ping Wang, Na Wang, and Yaodong Chen. 2020. “A Key Bacterial Cytoskeletal Cell Division Protein FtsZ as a Novel Therapeutic Antibacterial Drug Target.” *Bosnian Journal of Basic Medical Sciences / Udruzenje Basicnih Mediciniskih Znanosti = Association of Basic Medical Sciences* 20 (3): 310–18.
Venter, Henrietta, Rumana Mowla, Thelma Ohene-Agyei, and Shutao Ma. 2015. “RND-Type Drug Efflux Pumps from Gram-Negative Bacteria: Molecular Mechanism and Inhibition.” *Frontiers in Microbiology* 6 (April): 377.
Vrieze, Anne, Carolien Out, Susana Fuentes, Lisanne Jonker, Isaie Reuling, Ruud S. Kootte,

Els van Nood, et al. 2014. “Impact of Oral Vancomycin on Gut Microbiota, Bile Acid Metabolism, and Insulin Sensitivity.” *Journal of Hepatology* 60 (4): 824–31.
Weiss, A. S., A. G. Burrichter, A. C. D. Raj, and A. von Strempel. 2021. “Exploring the Interaction Network of a Synthetic Gut Bacterial Community.” *bioRxiv*. https://www.biorxiv.org/content/10.1101/2021.02.25.432904v1.abstract.
Wilm, Andreas, Pauline Poh Kim Aw, Denis Bertrand, Grace Hui Ting Yeo, Swee Hoe Ong, Chang Hua Wong, Chiea Chuen Khor, Rosemary Petric, Martin Lloyd Hibberd, and Niranjan Nagarajan. 2012. “LoFreq: A Sequence-Quality Aware, Ultra-Sensitive Variant Caller for Uncovering Cell-Population Heterogeneity from High-Throughput Sequencing Datasets.” *Nucleic Acids Research* 40 (22): 11189–201.
Yilmaz, Bahtiyar, Catherine Mooser, Irene Keller, Hai Li, Jakob Zimmermann, Lars Bosshard, Tobias Fuhrer, et al. 2021. “Long-Term Evolution and Short-Term Adaptation of Microbiota Strains and Sub-Strains in Mice.” *Cell Host & Microbe* 29 (4): 650–63.e9.
Zhao, Shijie, Tami D. Lieberman, Mathilde Poyet, Kathryn M. Kauffman, Sean M. Gibbons, Mathieu Groussin, Ramnik J. Xavier, and Eric J. Alm. 2019. “Adaptive Evolution within Gut Microbiomes of Healthy People.” *Cell Host & Microbe* 25 (5): 656–67.e8.

Supplementary Data

Supplementary Figures

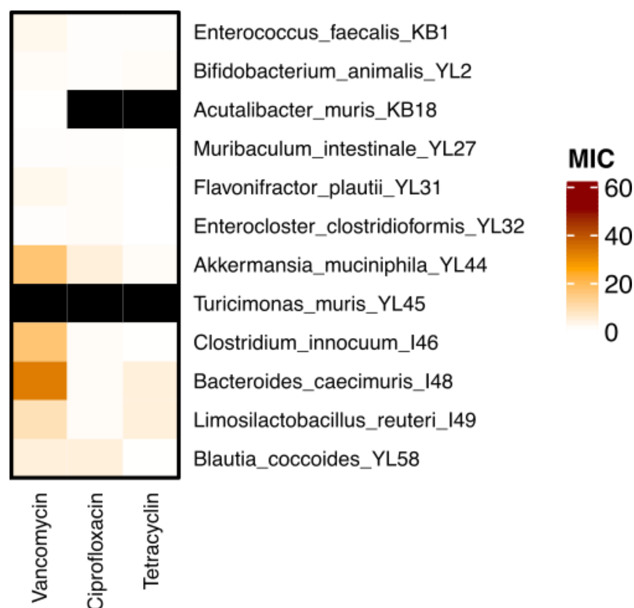

Figure S1: Minimal Inhibitory Concentration (MIC) of OMM[12] members in single cultures *in vitro*. Community members with a relatively higher MIC concentration exhibit a higher intrinsic resistance against the indicated compound.

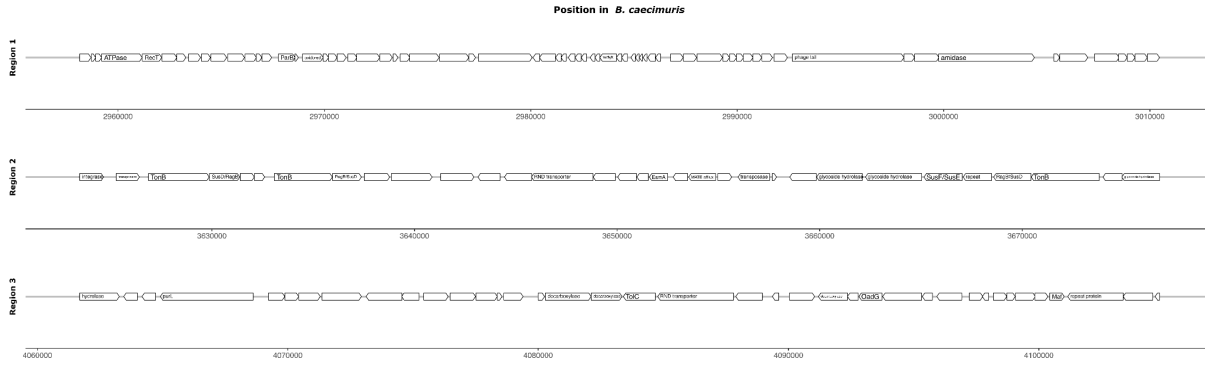

Figure S2: Loci with highly different read coverage dependent on treatment group on time points. Location and orientation of genes based on published genome annotation.

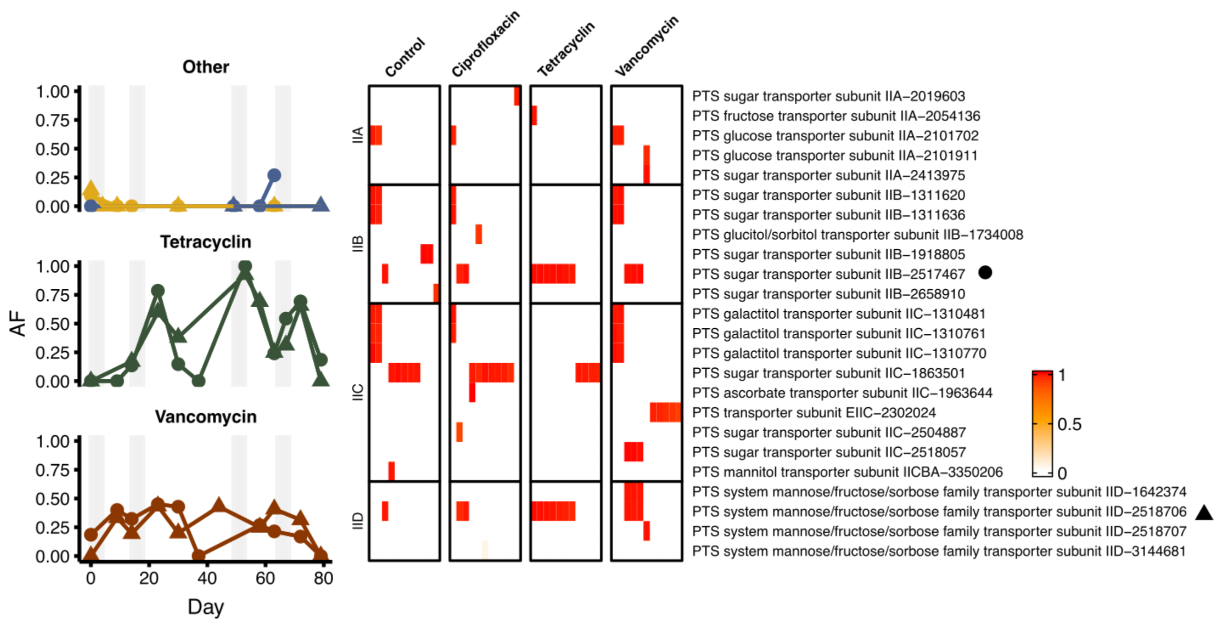

Figure S3: Change in AF of SNPs located in genes located on the PTS related genes on *C. innocuum* found in WGS (left) and isolates (right heatmap).