Dynamics of the human gut phageome during the antibiotics therapy

Anna Górska*†, Silke Peter‡§, Matthias Willmann‡§, Robert Schlaberg¶‖ and Daniel Huson*

*ZBIT Center for Bioinformatics, University of Tübingen, Sand 14, 72076, Tübingen, Germany
†International Max Planck Research School From Molecules to Organisms,
Max Planck Institute for Developmental Biology and Eberhard Karls University Tübingen,
Spemannstr. 35 - 39, 72076, Tübingen, Germany
‡Institute of Medical Microbiology and Hygiene,
University of Tübingen, Elfriede-Aulhorn-Str. 6, 72076, Tübingen, Germany
§German Center for Infection Research (DZIF), partner site Tübingen, Tübingen, Germany
¶Department of Pathology, University of Utah, UT, Salt Lake City, USA
‖ARUP Institute for Clinical and Experimental Pathology, UT, Salt Lake City, USA

Abstract—**Antibiotics are widely used in all sorts of the medical procedures. Without them, modern medical practice is not possible. As much as they are inevitable, they have severe adverse effects in promoting the antibiotic resistance. Each antibiotic therapy has a profound impact on the human gut. Firstly antibiotics select resistant bacteria what results in rapid changes of the taxonomical composition of the human gut. Secondly, the antibiotic pressure prompts the horizontal gene transfer including transduction that is intermediated by the phages, what advances spread of the resistance. Although the taxonomical composition restores after a month from the therapy, the bacteria get enriched with the genes conferring resistance to antibiotics. We have analyzed changes in the microbiome and phageome of the two people throughout the antibiotic therapy and subsequent recovery time. Analysis of the phage genomes is bioinformatically tricky as phages have complex genomic structures and there are no reliable reference databases. Therefore, we developed a novel comprehensive pipeline for the time-series analysis of the metagenomic datasets focusing on the function and changes in the phage sequences. We observed the phages respond to the antibiotic therapy, and are enriched with antibiotic-resistant genes.**

I. INTRODUCTION

Antibiotics are amongst the most successful drugs in saving human life. They have been used in medicine since their discovery in 1928. Those 90 years of extensive antibiotics usage caused the emergence of antibiotic resistance [1] - i.e., there are more and more bacterial strains, including pathogenic, that cannot be eradicated from the patient's organism. In 2014 WHO classified antibiotic resistance as a severe threat to the human life [2].

No antibiotic is ideally specific towards the pathogen. Therefore, an antibiotic therapy affects a substantial portion of the person's gut bacteria and causes rapid changes in the composition of the gut microbiome. Although the taxonomical structure of the healthy microbiome can restore in a month, the therapy causes enrichment of the antibiotic-resistant genes within the genomes of the gut bacteria [3], [4]. Consequently, with each antibiotic therapy it is easier for the pathogenic bacteria to become resistant and it is harder for the microbiome to restore its taxonomic structure [5].

The antibiotic-resistant genes can pass between bacterial cells via means of the horizontal gene transfer (HGT), that includes the three mechanisms: natural transformation, conjugation involving plasmids and transduction involving bacterial viruses [6], [7].

The bacterial viruses (phages) infect bacteria and integrate their genomes into the bacterial chromosomes. Prompted by stress, phages can steer bacteria into the production of the phage particles loaded with the nucleic acid molecule of the phage genome [8]. The phages are often enriched with the bacterial genes located near to their integration side on the genome. This process is influenced by antibiotics as they select for resistant bacteria and consequently phages carrying resistant genes [9].

The human gut works as a bioreactor enabling faster HGT causing emergence and spread of the resistance. Therefore, analysis of the HGT is necessary to obtain a full picture of the dynamics within the gut microbiome and perhaps to control the emergence of the resistance in the future. This project aims to analyze in detail changes of the human gut phageome and genes transferred by phages during the antibiotic therapy and to lie down work for the comprehensive pipeline for analysis of the HGT within microbiomes.

II. STUDY

Two healthy volunteers took ciprofloxacin twice a day orally. Ciprofloxacin is a broad-spectrum fluoroquinolone antibiotic that inhibits DNA separation during the cell division via inhibiting the gyrases and topoisomerases [10]. The stool samples were collected six times: before the therapy, on the 1st, 3rd, and the last day of the therapy (6th), and on the subsequent +2 and +28 days after the therapy [11]. For each of the twelve samples (two participants and the six time-points), two types of sequencing protocols were carried out. First sequencing was a standard deep whole-genome sequencing of entire metagenome (MiSeq at GATC Biotech AG Constance,
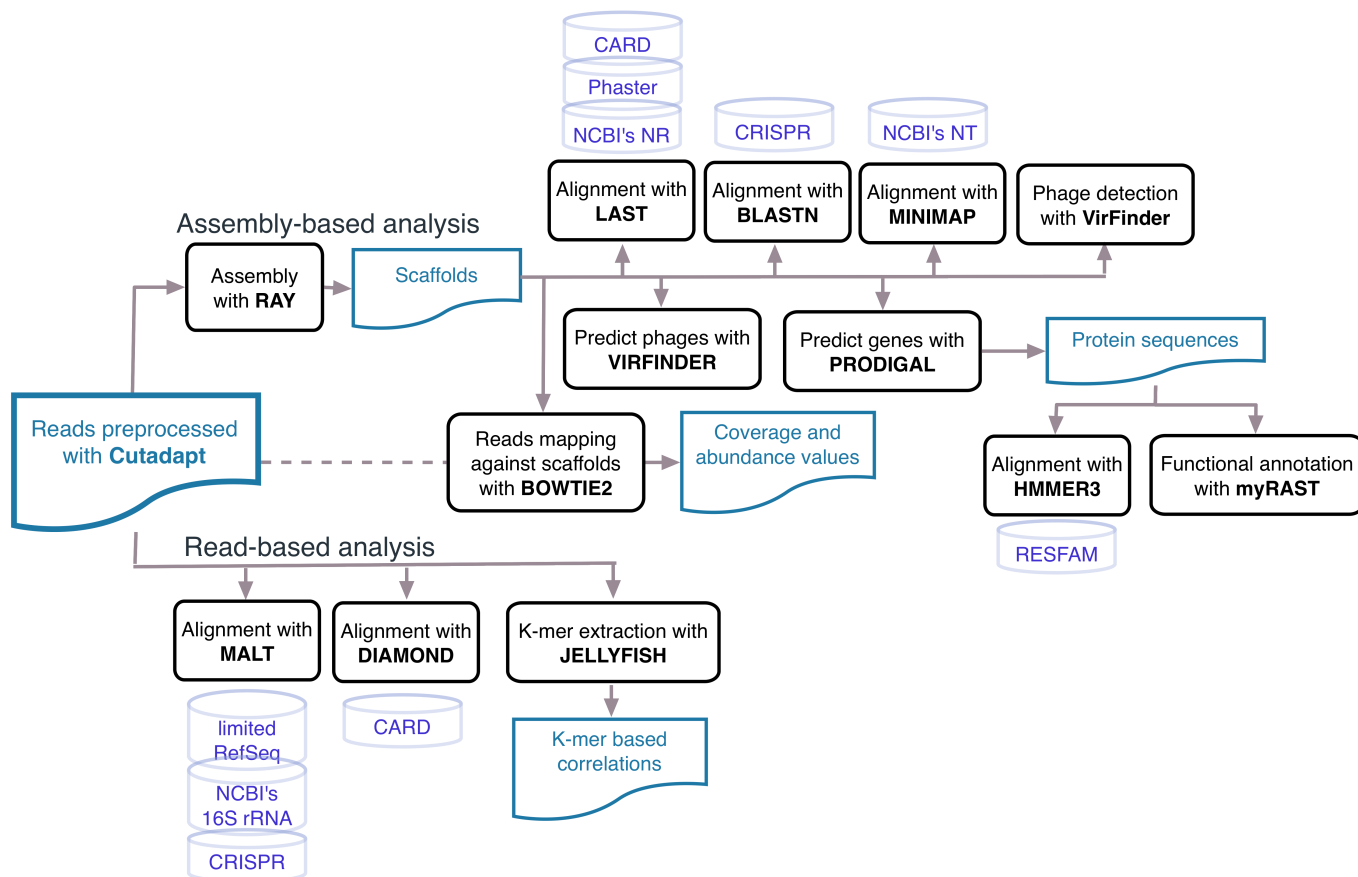
Figure 1: Data analysis pipeline. Blue elements denote inputs and results, black programs and violet the databases.

Germany). Second sequencing regarded only the viral-like particles as described in the [12]. It was performed on the NextSeq 500 System (mid-output kit, 2x150). Therefore, for each of the samples, two sequencing datasets are available: whole-metagenomic sequencing encapsulating both bacteria and phages (Microbiome set), and bacteriophage sequencing (Phage set).

III. METHODS

Metagenomic sequencing generates large numbers of the short reads belonging to a wide range of the micro-organisms. The classical metagenome analysis pipeline relies on the alignment of the reads against the most suitable sequence database. Based on those alignments the taxonomic or functional profiles are computed. Therefore, the analysis is only as reliable as the reference databases.

In the case of the Phage set, the classical approach is not possible. There are only ∼2,000 complete phage genomes in the NCBI database what is an order of magnitude smaller than the number of the bacterial genomes (∼35,000 January 2018). Additionally, the phage genomes are believed to be much more versatile and variable. Therefore, rather than using the classical approach the reads were assembled and annotated. To ensure comparability, the samples of the Microbiome set underwent the same analysis pipeline.

The assembly and the subsequent annotation is a complex pipeline involving ten different databases and total of sixteen computational tools (Fig. 1). It is not only expensive in terms of the computational time, but also a complicated procedure regarding the installation of the programs with all their dependencies, and storage. The entire pipeline was deployed on the computational cluster. A bioinformatician's job besides the design and execution of the analysis pipeline includes visualization of the results. That is yet another complex computational task performed on the cluster. The plotting was done with python Matplotlib package [13].

A. Basic analysis

The analysis pipeline consists of two paths: read-based and assembly based analysis (Fig. 1). The first path provides a contamination assessment and outlook of the global changes in the microbiome and phageome. The second path enables high-resolution gene-specific analysis of the feature behaviors over the course of the antibiotic therapy and following recovery.

First, the reads were preprocessed with Cutadapt [14]. Next, based on the number of k-mer count vectors extracted from the preprocessed reads the correlation heatmaps (Jellyfish [15]) and PCA plots were computed. The alignment of reads against the 16S rRNA NCBI database was used to determine possible bacterial contaminations in the Phage samples (MALT [16]).
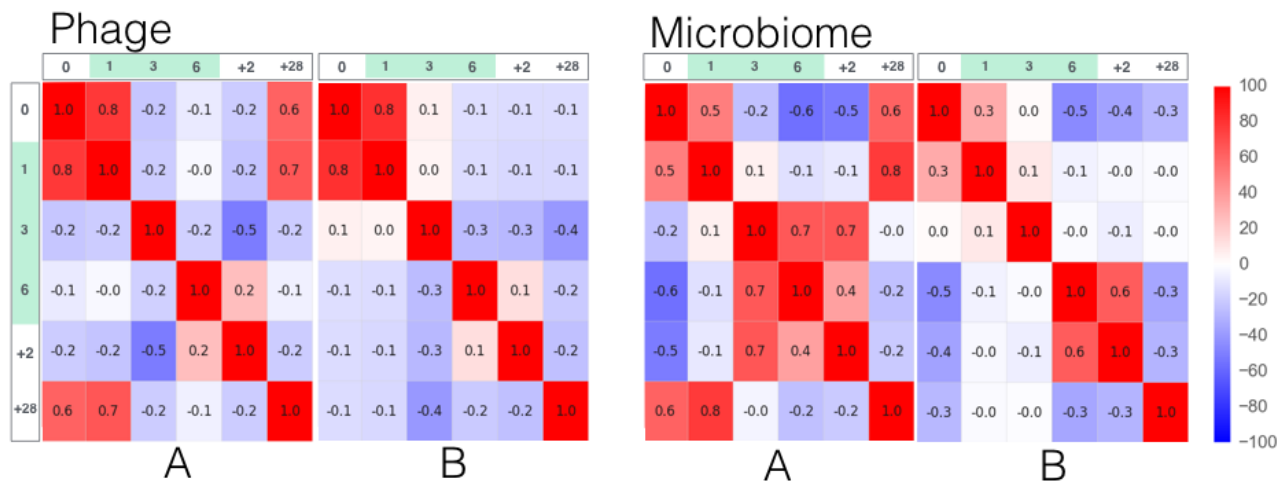
Figure 2: Correlation matrices for both datasets and both participants. Green color denotes days of the antibiotic therapy.

The assembled scaffolds were aligned against the multiple databases: CARD [17] to check if they carry the antibiotic-resistant genes, PHASTER [18] to find viral proteins, nucleotide database to enable MEGAN taxonomic annotation [19] and against the previously collected CRISPR-spacer database to identify phages [20], [21]. Next, genes were predicted, and the protein sequences were aligned with HMMER3 against the HMM profiles of the antibiotic resistance genes to enable sensitive detection of the antibiotic resistance. Finally, myRAST [22] was used for functional annotation of the extracted protein sequences.

Based on the reads mapping (Bowtie2 [23]), the coverage was computed, and then each scaffold was assigned an abundance value. That enables the abundance values of the scaffolds sharing particular annotations to be summed up. If we do it separately per each time-point, the result is an abundance trajectory. Therefore, abundance trajectories for different features in a context of the antibiotic therapy could be analyzed.

B. Phage detection

The Microbiome set comprises of the bacterial, plasmid and the phage scaffolds. We employed numerous strategies for the phage detection, e.g., based on the presence of CRISPR spacer, or of the phage protein. The methods pointed at disjoined sets of the scaffolds. Therefore, the Random Forest (RF) classifier was trained and used to enrich the selections. We used VirFinder [24] results (p-value ≤ 0.05) as the positive labeling, and training based on the other features such as GC content, gene density and a number of overlapping ORFs. In total, there were fourteen features used for the RF training.

IV. RESULTS AND DISCUSSION

K-mer based correlation plots (Fig. 2) show that both the phageome and the microbiome for both participants change as a result of the antibiotic pressure. In all variants, the 0 and 1st days are positively correlated with each other suggesting there is no change on the 1st day of therapy; and

negatively correlated with the following three days. Hence the structure of phageome and microbiome changes as a result of the therapy, starting on the 3rd day. In the case of the Microbiome set, participant A 3rd to +2 days correlate with each other, however, in both of the Phage sets, the 6 and +2 days are correlated. That means the gut microbiomes reach other stable states under the antibiotic pressure. Finally, for the Microbiome and Phage sets of the participant A, the last day correlates with the 0 and 1st days, i.e., microbiome and phageome structures restore after the therapy. That is not a case for the participant B, suggesting that only the A participant's microbiome is healthy.

As predicted, we observed the antibiotic influences foremost bacteria. Abundance of the both bacterial genomic and the plasmid scaffolds carrying antibiotics-resistant genes increase as a result of the therapy. We observed an increase of the abundance of the scaffolds with an alignment to the antibiotic-resistant HMM profile, suggesting phages pick up fragments of the unspecific resistant genes (Fig. 3). As showed by the k-mer profiles, the levels do return to the starting values in the case of the participant A but not B. Ovearll trajectory patterns for the phage sets are weaker. What is probably caused by a lower quality and depth of the sequencing of this dataset in a comparison to the Microbiome set.

V. CONCLUSION

The study has a pilot character, aimed at developing and testing of the pipeline as the two participants do not constitute a representative biological sample, the strong biological conclusions cannot be drawn. However, we have shown that the pipeline enables extensive analysis of the multiple aspects of such combined time series data. It is easily adaptable for analysis of other HGT mechanisms such as conjugation involving plasmids. Nevertheless, it is a complicated pipeline suitable for deployment only on the sizeable computational cluster such as BiNAC.
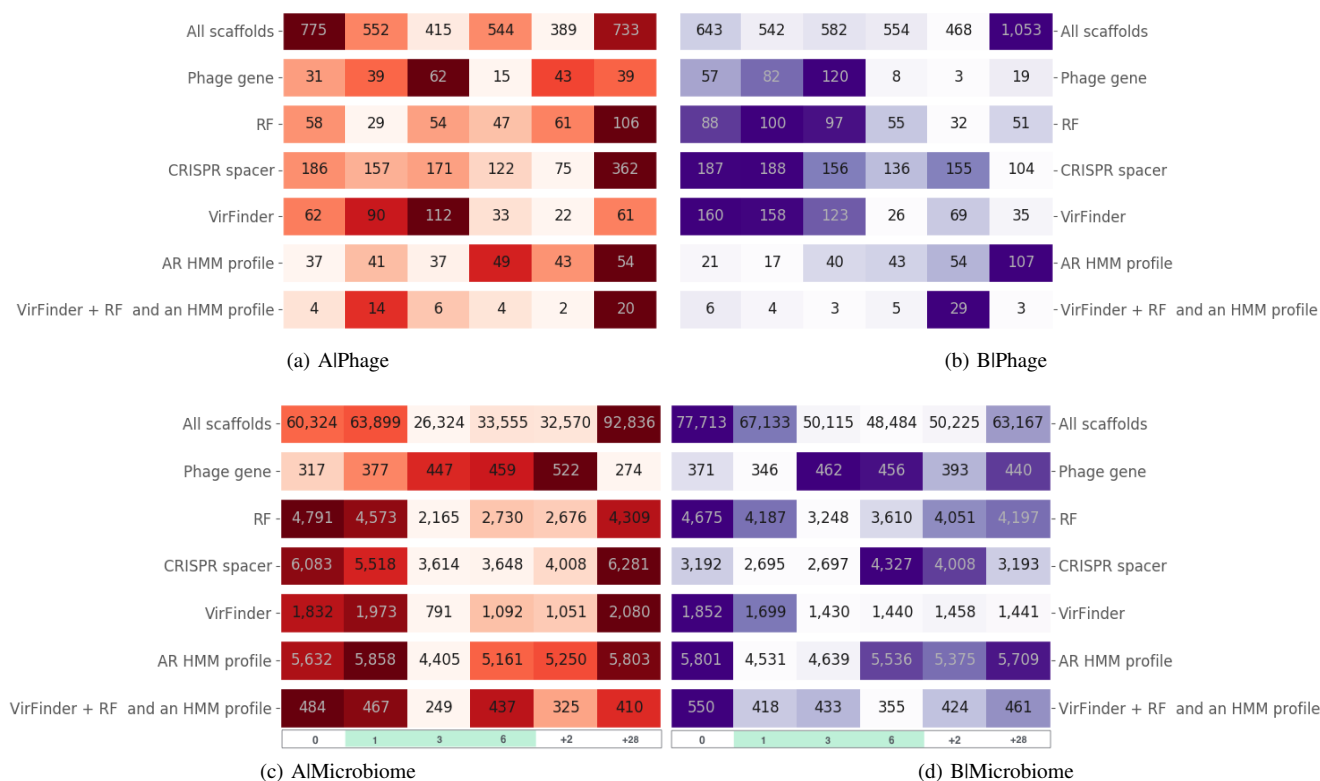
Figure 3: Feature abundance trajectories for different selections of the scaffolds: rows denote features and columns the consecutive sampling days. The green color marks the days of the antibiotic therapy. Four features are speaking to a possible phage origin of the scaffolds: Phage gene, CRISPR spacer, VirFinder and RF. HMM profiles of antibiotics informs about antibiotic resistance genes. Colors correspond to each horizontal trajectory separately.

REFERENCES

[1] H. Nikaido, "Multidrug Resistance in Bacteria," *Annu Rev Biochem.*, no. 2, pp. 119–146, 2009.

[2] WHO, "Antimicrobial resistance. Global Report on Surveillance," *Bulletin of the World Health Organization*, vol. 61, no. 3, pp. 383–94, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22247201

[3] Y. Hu, X. Yang, J. Qin, N. Lu, G. Cheng, N. Wu, Y. Pan, J. Li, L. Zhu, X. Wang, Z. Meng, F. Zhao, D. Liu, J. Ma, N. Qin, C. Xiang, Y. Xiao, L. Li, H. Yang, J. Wang, R. Yang, G. F. Gao, J. Wang, and B. Zhu, "Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota." *Nature communications*, vol. 4, p. 2151, 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23877117

[4] A. A. Salyers, A. Gupta, and Y. Wang, "Human intestinal bacteria as reservoirs for antibiotic resistance genes," *Trends in Microbiology*, vol. 12, no. 9, pp. 412–416, 2004.

[5] L. Dethlefsen and D. A. Relman, "Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation." *Proceedings of the National Academy of Sciences of the United States of America*,

no. Suppl 1, pp. 4554–61, 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20847294

[6] P. J. Hastings, S. M. Rosenberg, and A. Slack, "Antibiotic-induced lateral transfer of antibiotic resistance," *Trends in Microbiology*, vol. 12, no. 9, pp. 401–404, 2004.

[7] W. V. Schaik, "The human gut resistome," *Royal Society Publisher*, vol. 370, pp. 1–9, 2015.

[8] T. A. Brown, "Mapping genomes," in *Genomes*, 2002, pp. 125–162.

[9] S. R. Modi, H. H. Lee, C. S. Spina, and J. J. Collins, "Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome." *Nature*, vol. 499, no. 7457, pp. 219–22, 2013.

[10] B. Jia, A. R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K. K. Tsang, B. A. Lago, B. M. Dave, S. Pereira, A. N. Sharma, S. Doshi, M. Courtot, R. Lo, L. E. Williams, J. G. Frye, T. Elsayegh, D. Sardar, E. L. Westman, A. C. Pawlowski, T. A. Johnson, F. S. Brinkman, G. D. Wright, and A. G. McArthur, "CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database," *Nucleic Acids Research*, vol. 45, no. D1, pp. D566–D573, 2017.

[11] M. Willmann, M. El-Hadidi, D. H. Huson, M. Schütz, C. Weidenmaier, I. B. Autenrieth, and S. Peter, "Antibiotic selection pressure determination through sequence-based metagenomics," *Antimicrobial Agents and Chemotherapy*, vol. 59, no. 12, pp. 7335–7345, 2015.

[12] R. V. Thurber, M. Haynes, M. Breitbart, L. Wegley, and F. Rohwer, "Laboratory procedures to generate viral metagenomes." *Nature protocols*, vol. 4, no. 4, pp. 470–483, 2009.

[13] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 99–104, 2007.

[14] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *EMBnet.journal*, vol. 17, no. 1, pp. pp. 10–12, 2011.

[15] G. Marçais and C. Kingsford, "A fast, lock-free approach for efficient parallel counting of occurrences of k-mers," *Bioinformatics*, vol. 27,

no. 6, pp. 764–770, 2011.

[16] A. Herbig, F. Maixner, K. I. Bos, A. Zink, J. Krause, and D. H. Huson, "MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman," *bioRxiv*, p. 50559, 2016. [Online]. Available: http://biorxiv.org/content/early/2016/04/27/050559.abstract

[17] A. G. McArthur, N. Waglechner, F. Nizam, A. Yan, M. A. Azad, A. J. Baylay, K. Bhullar, M. J. Canova, G. De Pascale, L. Ejim, L. Kalan, A. M. King, K. Koteva, M. Morar, M. R. Mulvey, J. S. O'Brien, A. C. Pawlowski, L. J. V. Piddock, P. Spanogiannopoulos, A. D. Sutherland, I. Tang, P. L. Taylor, M. Thaker, W. Wang, M. Yan, T. Yu, and G. D. Wright, "The comprehensive antibiotic resistance database," *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 7, pp. 3348–3357, 2013.

[18] D. Arndt, J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, and D. S. Wishart, "PHASTER: a better, faster version of the PHAST phage search tool," *Nucleic Acids Research*, vol. 44, no. W1, pp. W16–W21, 2016. [Online]. Available: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw387

[19] D. H. Huson, S. Beier, I. Flade, A. Górska, M. El-Hadidi, S. Mitra, H.-J. Ruscheweyh, and R. Tappu, "MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data," *PLOS Comput Biol*, vol. 12, no. 6, pp. 4–12, 2016.

[20] R. Jansen, J. D. A. Van Embden, W. Gaastra, and L. M. Schouls, "Identification of genes that are associated with DNA repeats in prokaryotes," *Molecular Microbiology*, vol. 43, no. 6, pp. 1565–1575, 2002.

[21] A. F. Andersson and J. F. Banfield, "Virus population dynamics and acquired virus resistance in natural microbial communities." *Science (New York, N.Y.)*, vol. 320, no. 5879, pp. 1047–50, 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18497291

[22] E. M. Glass and F. Meyer, "The Metagenomics RAST Server: A Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes," *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, vol. 8, pp. 325–331, 2011.

[23] B. Langmead, "Aligning short sequecning reads with Bowtie," *Current Protocols Bioinformatics*, pp. 1–24, 2011.

[24] J. Ren, N. A. Ahlgren, Y. Y. Lu, J. A. Fuhrman, and F. Sun, "VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data," *Microbiome*, vol. 5, no. 1, p. 69, 2017.