Genomics of phages with therapeutic potential

Genomics of phages with therapeutic potential

Bacteriophages, viruses that prey on bacteria, have been applied since the 1920’s to treat and prevent bacterial infection. After the discovery of antibiotics, this route was however largely abandoned. Now, with antimicrobial resistance in human-pathogenic bacteria on the rise and a dire need for alternatives, phage therapy once again takes center stage.

Phage therapy holds the promise of substantial benefits both from the economic as well as the public health perspective but also holds distinct challenges. The aim of this PhD was to address how bioinformatics tools, specifically genomics and mathematical modelling, can be applied to move the field towards a future of actual phage therapy in humans. It is composed of three related research projects.

The first part of this thesis is an introduction to various topics and methods relevant to the research projects that jointly make up this PhD. Chapters 1 - 3 deal with phages, their use in therapy and the nosocomial pathogen Staphylococcus aureus. Following that, Chapter 4 and 5 provide an overview of Next Generation Sequencing as well as commonly employed genomics tools, while Chapter 6 details basics of Machine Learning.

The second part, divided into three chapters, presents the three research projects. In project 1, an important commercial phage cocktail with a long history was sequenced and its component phages analyzed. It was found that the cocktail is composed of at least 23 different phage types, which were present in differing abundances. Some of these phage types were successfully amplified on a collection of in-house bacteria corresponding to the cocktail's stated bacterial targets. Further, no harmful genes were detected in the cocktail.

Project 2 deals with phage communities in sewage by comparing samples from around the world to each other as well as to databases of available phage genomes. It revealed a great diversity in the sequences, many of which were distant from all known phages. The phage content of the different sample locations exhibited a rather stable genomic distance that was not influenced by whether the locations were geographically close or not.

Project 3 had the goal of identifying gene families in the extensive accessory genome of the hospital pathogen Staphylococcus aureus that influence its susceptibility to clinical phage preparations. This was done by phage testing a set of patient-derived S. aureus isolates against a panel of phage preparations. We then sought to model the results using the bacteria's genetic background as features. Doing so, we built nine models with sufficient explanatory power over the susceptibility outcome and from them identified a set of 167 gene families relevant for phage susceptibility.

The third part of the thesis consists of conclusive remarks and a critical reflection on how each of these projects has impacted the field and how they are connected as well as pointing out directions for future investigations.

In summary, the work included in this thesis focuses on applying genomics and mathematical modelling to questions related to phage therapy.