Abstract

Helicobacter pylori is a gram-negative bacterium that is found in the human gastric mucosa in more than 50% of the world's population. It is a pathogen and infections can lead to chronic gastritis or gastric ulcers, but also a positive effect on asthma in children was observed. Since this bacteria has a high genomic variability it is important to sequence and assemble different *H. pylori* strains to determine intra- and intergenomic variability and the association with diseases.

This Master's thesis is divided into two main parts. First, different assembly tools were compared on a range of different bacterial sequencing data. In particular, the results of the popular GS De Novo Assembler (Newbler) were compared to the results of a previously published assembler benchmark. The performances of the assemblers were quite different depending on the genomic data. In general *MaSuRCA*, *Cabog* and *SPAdes* performed best while *SGA* and *Abyss* got the lowest scores. Newbler achieved good results especially in relation to the reference coverage, less overlapping bases and low rate of mismatches and indels. It obtained better results with Illumina MiSeq than with HiSeq data.

The second and main part of this thesis covers the assembly of four *Helicobacter pylori* strains. These strains were sequenced using Illumina and PacBio sequencing technologies. Assemblies were performed with two different long-read assembly strategies. The first one was a hybrid approach where the long PacBio reads are corrected by mapping of the short Illumina reads before used for assembly. But fully finished and closed genomes were only achieved with the second method, a so called stand-alone-assembly approach of Canu assembler followed by a circularisation step and a consensus building step using Illumina reads. It could be shown that *H. pylori* is a bacterium that has a high genomic variability, including large inversions of about 400 kbp, different copy numbers of the cagA region and a high amount of local variations affecting different genes.