In the majority of DNA sequencing experiments the first step of analysis consists of mapping sequencing reads onto a so-called *reference genome*, which represents the consensus of genomic sequence of the species of interest. Currently reference genomes are available for thousands of species and much effort is devoted to the analysis of genomic diversity among them. This is especially visible in the case of human genomics, where the development is driven by the perspective of application to personalized medicine. However, current pipelines of sequencing data analysis are unable to utilize this knowledge to reduce the bias and the noise caused by differences between reference and actual genomes.

The objective of the current project is to address this problem. We will introduce the concept of *reference multi-genome* that will model multiple variants of particular genomic loci. Furthermore, we will design and implement tools incorporating this concept into current sequencing analysis pipelines. In order to save for further analysis reads that don't match a consensus reference genome, we will focus on the algorithmically challenging read mapping step. Finally, we will illustrate the advantages of our approach in a case study.

Summarizing, our project will provide a complete set of tools for processing DNA sequencing reads that (1) utilizes available genome variation knowledge, (2) is ready to use by non-computer scientists within various data analysis pipelines and (3) proves its superiority over traditional approaches in real applications. The proposed approach will be advantageous for a wide range of research projects benefiting from DNA sequencing technology, including cancer genomics and personalized medicine.