# Establishing a universal pangenome model

The term *pangenome* was initially proposed as a single data structure for joint analysis of a group of bacterial genes. In the presence of a variety of whole genome sequences available it has evolved and currently it refers to a model of joint analysis of genomes of related organisms. Such a model is expected to support various operations, including construction, visualization, annotation, processing experimental data etc.

Numerous pangenome models were proposed up to date, ranging from unstructured collections of genome sequences to sophisticated models that require complex preprocessing of sequence data. The most popular solutions include various types of *sequence graphs*, in which particular genomes are represented by paths sharing identical or similar fragments and forking on dissimilar sequence areas. These models offer different functionalities and are subject to specific restrictions. The choice of a model determines the range of available tools, which in turn affects the quality of analysis results, its effectiveness, and even the possibility of conducting it.

The aim of the current project is to address this problem. We will develop data analysis pipelines that cross the boundaries between different models and combine their advantages. This goal will be achieved in the following steps:

1. **Design and implementation of transformation methods of pangenome models.** We will focus on transformations from models that are relatively easy to build to other models.

2. **Implementation of algorithms and data structures for transferring annotations between pangenome models.** Pangenome annotations include individual genomes, inherited genome annotations and experimental data. They will be transferred according to the transformations of the model, which will require efficient representation of the latter.

3. **Development of effective data analysis pipelines based on transformations of pangenome models.** We will perform benchmarks of tools performing similar tasks on different pangenome models. Then, we will propose pipelines combining various tools and models.

Summarizing, our project will lead to a universal pangenome model.