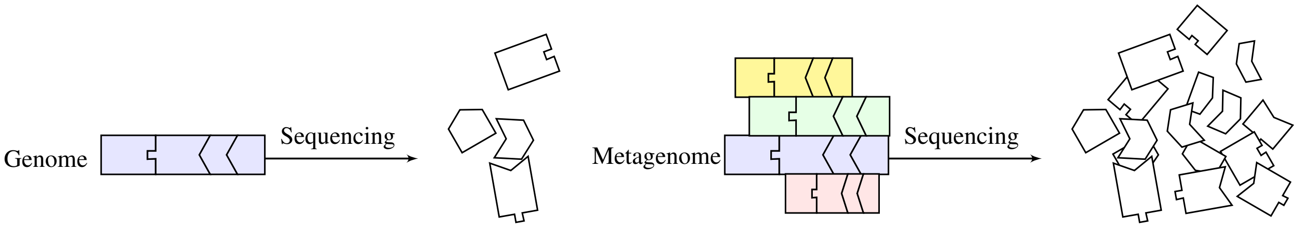
In the natural world, the information system that coordinates the processes of reproduction and vital functions regulation, is underpinned with the *genome* encoded using the deoxyribonucleic acid (DNA). The precise order of nucleotides within a DNA molecule may be determined using DNA sequencing. As it is impossible to read the entire DNA sequences at once relying on the state-of-the-art technologies, these sequences are first broken up randomly into numerous short fragments. In the recent decade, the DNA sequencing methods were becoming cheaper and faster, hence the number of known sequences is increasing rapidly.

The genome is the entire set of the genetic material acquired from a single organism. It is also possible to analyse the genomes of all the organisms living in a given environment at once. Such an acquired set of genomes is called the *metagenome*, and it may be subject to the same sequencing procedure as the genome derived from a single organism. During the metagenome sequencing, a collection of mixed reads is obtained, which is derived from the DNA sequences of all the microorganisms in a single environmental sample. The difference between genome and metagenome sequencing is illustrated in the figure below.



The biggest advantage of the metagenomic analysis is that it is not necessary to isolate and culture organisms in the laboratory to study them. Therefore, it is possible to investigate the species that previously have been usually neglected due to the lack of laboratory-grown cultures. Metagenomic analyzes can help in solving numerous practical challenges in medicine, engineering, agriculture, and ecology. In the recent years, the methods for metagenomic reads analysis are being actively developed due to their practicality. The most methods consist in classifying the studied metagenomic fragments to the specific taxonomy categories or grouping them based on their mutual similarity.

It is worth noting that the human organism carries a hundred times more bacterial genes than our inherited human genome. There are many studies showing the dependence between the health condition (physical and psychic) and the microbiota composition (mainly in the intestines). For example, the recent studies demonstrated the correlation between the children's obesity and the occurrence of changes in their intestinal microflora while they were infants. However, little research has been aimed at making it possible to exploit this knowledge without performing cultures, but exclusively on the basis of metagenomic sample analysis.

The goal of this project is to develop new methods for efficient supervised and unsupervised classification of the reads generated from metagenome sequencing. These methods will be helpful in a variety of areas that employ the metagenomic analysis, such as engineering, agriculture, ecology, and medicine. Specifically, the created algorithms may occur helpful in solving particular bioinformatic tasks, such as:

- ∞ characterizing qualitatively and quantitatively the composition of the environment,
- ∞ investigating the relationships between species composition and environmental conditions,
- ∞ detecting new species (especially in case of the organisms living in places with extreme conditions for life),
- ∞ studying the health condition of a patient based on his metagenomic sample.

The last task may help in replacing the lengthy process of microbiological analysis and accelerate the diagnosis of many diseases, which in turn could result in increasing the effectiveness of treatment.