The DNA sequence of the human genome was published for the first time in 2003. The *Human Genome Project* took 13 years of work of research centres from all over the world and costed nearly 3 billion dollars. Nowadays, there are technologies allowing human genome to be sequenced in several days at a cost less than thousand dollars. As a result, availability of both, genomic and proteomic data, has increased rapidly opening new opportunities in many fields of science. Multiple sequence alignment is one of the most important analyses conducted on sequence data. In general, it consists in identifying events in the evolution of sequences. These include simple phenomena (insertions and deletions of fragments or symbol substitutions) as well as more complex events (rearrangements, duplications, horizontal gene transfers). Sequence alignment is often represented in a graphical form (Fig.1.).

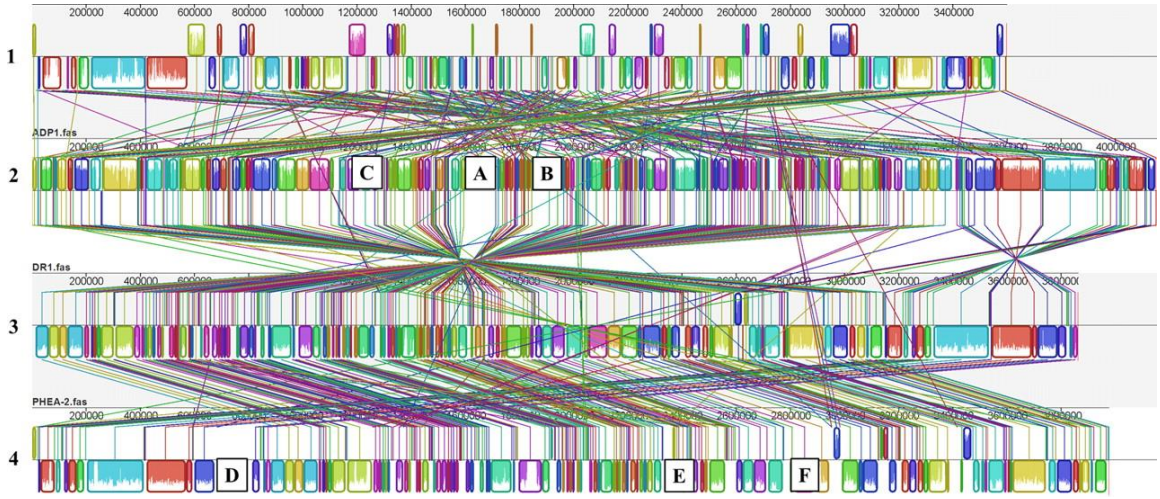

Fig.1. Graphical representation of genome alignment of four species (given in rows). The presence of complex evolutionary events like rearrangements or duplications is visible. Corresponding blocks in different organisms (homologous regions) are denoted with colours and connected with lines (Jung et al., 2011).

Multiple sequence alignment is of crucial significance for understanding processes ongoing in living organisms. For instance, it gives insight into structure and function of proteins in human body or extends knowledge about gene expression by identifying binding sites of regulatory molecules. Therefore, it is very important in establishing causes of many diseases and developing therapies. Another field of application is finding evolutionary relationships between organisms, giving a contribution to discovering the history of speciation.

The project is focused on developing alignment algorithms superior to existing strategies, with two aspects being of particular interest. These are the result quality evaluated on the basis of manually curated reference alignments and the computational time. The latter is especially important in the face of rapidly increasing availability of sequence data. The problem is the number of sequences (as in the case of protein families containing tens of thousands of members), as well their lengths and the presence of complex evolutionary events (as in the case of genomes whose chromosomes consist of hundreds millions of nucleotides). Due to dissemination of multiple sequence alignment in biological analyses, results of our work may find a use not only in biology or biotechnology, but also in medicine, pharmacology, agriculture, chemical industry, and many other domains