As early as 50 years ago scientists discovered that genes, DNA fragments containing instructions for protein synthesis, can be split. The protein-coding sequence (exon) is often interrupted by a non-coding segment called “intron”. Such a sequence is excised (spliced) from a maturing ribonucleic acid (RNA) molecule to assemble a mature template for the protein translation.

It is commonly believed that introns can be found only in eukaryotes (animals, plants and fungi), and genes of prokaryotic microorganisms are continuous. Even when the first split bacterial or viral genes were discovered, most scientists considered them an oddity, not a rule. Many biology handbooks still distinguish between eukaryotic genes with exons and introns and uninterrupted coding sequences found in the genomes of microorganisms or viruses.

The results of analysis conducted by our team contradict this view. They indicate that introns can be found in many genes of viruses infecting bacteria called bacteriophages. The program we have developed recognizes interrupted genes and distinguishes introns from random gaps in the coding sequence, which may be the result of mutations or sequencing errors. Obtained results showed that introns are a relatively common part of the viral genomes. Preliminary laboratory experiments confirm some of those results and show that predicted introns are often capable of self-excision from coding sequences.

The aim of this project is to assess the true prevalence of bacteriophage introns and to understand the mechanism of their splicing. To achieve this goal we plan to answer the following questions:

- How many bacteriophage genes contain introns?
- How viral genes interrupted by introns are assembled into a continuous RNA molecule?
- Which host factors are involved in RNA splicing?
- What conditions allow for efficient intron splicing?
- What role do introns play during the phage infection?

We expect that finding the answers to these questions will allow us to discover new groups of intervening sequences, elucidate their function, and finally, understand the evolution of genomes better.