

Abstrakt

Identifikace virů je problematická, jelikož nemají žádnou sekvenci v genomu, která by byla pro všechny viry společná. Díky NGS a virové metagenomice byla získána spousta sekvencí, ale ne všechny se dosud podařilo identifikovat. Sekvenování s využitím hybridizačních sond se specializuje na vycytávání virových sekvencí právě pomocí specifických sond připravených podle virových genomů a je velmi citlivá. Při studii viromů ze vzorku stolice, krve a vzorku z dýchacích cest se ukázalo, že identifikace metagenomických dat je nejnáročnější ze vzorku stolice. Povedlo se ale objevit např. nový pikornavirus, nový lidských papillomavirus nebo crAssfág. K identifikaci a analýze těchto dat napomáhají bioinformatické nástroje jako jsou třeba SURPI, Taxonomer, RIEMS a MG-RAST.

Klíčová slova: virom, identifikace, neznámé sekvence, virový genom, „assembly“

Abstract

A virus identification is problematic, because viruses do not have any sequence in genome, which would be common for all of them. Thanks to NGS viral metagenomic is able to gain lots of sequences, but not all of them have been identified yet. Sequencing using hybridization probes specialises in capture of viral genomes and it is very sensitive. Studying viromes from stool and blood samples and samples from respiratory tract has shown that the hardest identification of metagenomic data is from stool samples. Despite this new viruses has successfully been discovered, for example new picornavirus, new human papillomavirus or crAssphage. Bioinformatic pipelines are helpful for identification and analysis of those data, such as pipelines like SRPI, Taxonomer, RIEMS and MG-RAST.

Key words: virome, identification, unknown sequences, viral genome, assembly