



Desenvolvimento de protocolo modular para leitura e classificação metagenômica usando dicionários de Armazenamento para transferir e avaliar a taxonomia e a anotação funcional de microbiota.

Igor Ferreira Brito ¹, Matheus Augusto de Bittencourt Pasquali ²

RESUMO

As análises metagenômicas exploram dados obtidos de amostras ambientais para desvendar detalhes sobre sua composição taxonômica, a função desempenhada por um agente microbiano, ou ambos. As etapas necessárias para conduzir uma análise metagenômica são bastante abrangentes, incluindo pré-processamento de dados, alinhamento em relação a um banco de dados de referência, análises taxonômicas e funcionais. Selecionamos um conjunto de ferramentas de código aberto de última geração para análise metagenômica. Incluí fastp, a ferramenta de pré-processamento mais rápida e completa, bowtie2, a ferramenta mais precisa para remoção de sequências hospedeiras e BASTA, uma ferramenta flexível para análise taxonômica. Também foi criado um script python para transferir anotações para resultados de alinhamento em análises funcionais. Aplicamos o protocolo a um conjunto de dados para validação do seu uso, simulados para avaliar resultados taxonômicos e funcionais. As métricas dos resultados do conjunto de dados simulados mostram que o protocolo identificou as espécies esperadas em análise, indicando que o protocolo identifica espécies raras, e é particularmente útil para realizar análises funcionais usando National Center for Biotechnology Information (NCBI). O protocolo dá a liberdade de criar dicionários construídos para quaisquer identificadores funcionais. Todas as ferramentas utilizadas aqui estão disponíveis em um ambiente conda, facilitando a aquisição e configuração de software.

Palavras-chave: Metagenômica, bioinformática, protocolo, classificação taxonômica.

¹Igor Ferreira Brito; Graduando de Engenharia de Alimentos; Unidade Acadêmica de Engenharia de Alimentos, UFCA, Campina Grande, PB, e-mail: igor.ferreira@estudante.ufcg.edu.br

²Matheus Augusto de Bittencourt Pasquali Ph.D.; Orientador, Unidade Acadêmica de Engenharia de Alimentos, UFCA, Campina Grande, PB, e-mail: matheus.augusto@professor.ufcg.edu.br



Development of modular protocol for reading and metagenomic classification using storage agreement to transfer and evaluate taxonomy and functional annotation of microbiota.

ABSTRACT

Metagenomic analyses explore data obtained from environmental samples to uncover details about their taxonomic composition, the function performed by a microbial agent, or both. The steps required to conduct a metagenomic analysis are quite comprehensive, including pre-processing, alignment with a reference database, taxonomic and functional analyses. We selected a set of state-of-the-art open source tools for metagenomic analysis. Includes fastp, the fastest and most complete preprocessing tool, bowtie2, the most accurate tool for removal of host sequences and BASTA, a flexible tool for taxonomic analysis. A python script has also been created to transfer annotations to alignment results in functional analyses. We applied the protocol to a set of data to validate its use, simulated to evaluate taxonomic and functional results. The simulated data set results metrics show that the protocol identified the expected species under analysis, indicating that the protocol identifies rare species, and is particularly useful for performing functional analyses using National Center for Biotechnology Information (NCBI). The protocol gives you the freedom to create dictionaries built for any functional identifiers. All the tools used here are available in a conda environment, facilitating the acquisition and configuration of software.

Keywords: Metagenomics, bioinformatics, protocol, taxonomic classification.