SHOGUN: Accurate, scalable microbiome quantification

Authors

- **Benjamin Hillmann**
  ⓘ 0000-0003-4276-1329 · ⬡ bhillmann

  Computer Science and Engineering, University of Minnesota

Abstract

SHOGUN is a software pipeline for simultaneous taxonomic and functional abundance profiling of metagenomics datasets with Bayesian redistribution of ambiguous mapping. The pipeline is built in a modular fashion so that it may be run in its entirety or indivual parts may be flexible allowing for user creation of a reference database and selection of the alignment tool that best fits a given users data and computational resources. The package allows users to efficiently go from quality-controlled sequences to abundance profiles consistently and accurately, enabling reproducible metagenomic sequencing research.

Introduction

Example figure reference, Figure 1B shows that quality controlled reads are aligned against the reference database. Adding another line for watchdog. Another watchdog change with removing the symbolic link. Hellow world. Still having conflicts. Why are there still conflicts? Still having conflicts?
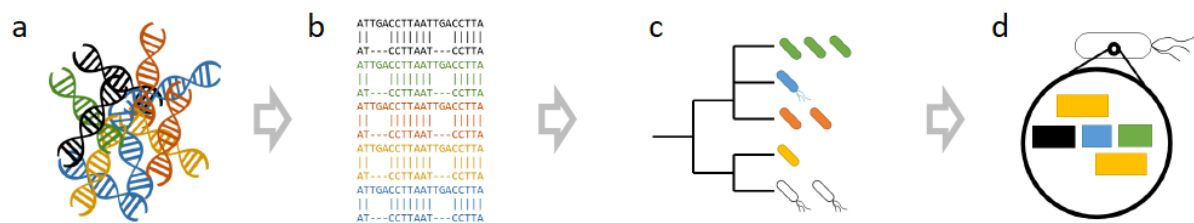

Figure 1: (a) Raw reads from the sequencing machine are demultiplexed into individual samples. Then, each read is quality controlled with by removing adapters, low quality bases and contaminates such as host reads. Optionally, the read pairs can be stitched. (b) The quality-controlled reads are aligned against a database of known genomes to identify each read's most likely source taxon. (c) The taxa that are hit are filtered out and summarized at a specific level. These processing steps include last common ancestor assignment, genome coverage analysis, and redistribution of reads to a specific taxonomic level. (d) After the taxonomic prediction is set, the full functional repertoire of genes is directly observed through a bag-of-genes approach or predicted through a per microbe approach

References