MGmapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads

An increasing amount of species and gene identification studies rely on the use of next generation sequence analysis of either single isolate or metagenomics samples. Several methods are available to perform taxonomic annotations and a previous metagenomics benchmark study has shown that a vast number of false positive species annotations are a problem unless thresholds or post-processing are applied to differentiate between correct and false annotations. MGmapper is a package to process raw next generation sequence data and perform reference based sequence assignment, followed by a post-processing analysis to produce reliable taxonomy annotation at species and strain level resolution. An in-vitro bacterial mock community sample comprised of 8 genuses, 11 species and 12 strains was previously used to benchmark metagenomics classification methods. After applying a post-processing filter, we obtained 100% correct taxonomy assignments at species and genus level. A sensitivity and precision at 75% was obtained for strain level annotations. A comparison between MGmapper and Kraken at species level, shows MGmapper assigns taxonomy at species level using 84.8% of the sequence reads, compared to 70.5% for Kraken and both methods identified all species with no false positives. Extensive read count statistics are provided in plain text and excel sheets for both rejected and accepted taxonomy annotations. The use of custom databases is possible for the command-line version of MGmapper, and the complete pipeline is freely available as a bitbucked package (https://bitbucket.org/genomicepidemiology/mgmapper). A web-version (https://cge.cbs.dtu.dk/services/MGmapper) provides the basic functionality for analysis of small fastq datasets.

General information
Publication status: Published
Organisations: Department of Bio and Health Informatics, Metagenomics, National Food Institute, Research group for Genomic Epidemiology, Genomic Epidemiology, Department of Systems Biology, Center for Biological Sequence Analysis
Number of pages: 13
Publication date: 2017
Peer-reviewed: Yes

Publication information
Journal: P L o S One
Volume: 12
Issue number: 5
Article number: e0176469
ISSN (Print): 1932-6203
Ratings:
BFI (2017): BFI-level 1
Scopus rating (2017): CiteScore 3.01 SJR 1.164 SNIP 1.144
Web of Science (2017): Indexed yes
Original language: English
Electronic versions:
journal.pone.0176469.pdf
DOIs:
10.1371/journal.pone.0176469
Source: FindIt
Source-ID: 2358325614
Research output: Contribution to journal › Journal article – Annual report year: 2017 › Research › peer-review