HostPhinder: A Phage Host Prediction Tool

The current dramatic increase of antibiotic resistant bacteria has revitalised the interest in bacteriophages as alternative antibacterial treatment. Meanwhile, the development of bioinformatics methods for analysing genomic data places high-throughput approaches for phage characterization within reach. Here, we present HostPhinder, a tool aimed at predicting the bacterial host of phages by examining the phage genome sequence. Using a reference database of 2196 phages with known hosts, HostPhinder predicts the host species of a query phage as the host of the most genomically similar reference phages. As a measure of genomic similarity the number of co-occurring k-mers (DNA sequences of length k) is used. Using an independent evaluation set, HostPhinder was able to correctly predict host genus and species for 81% and 74% of the phages respectively, giving predictions for more phages than BLAST and significantly outperforming BLAST on phages for which both had predictions. HostPhinder predictions on phage draft genomes from the INTESTI phage cocktail corresponded well with the advertised targets of the cocktail. Our study indicates that for most phages genomic similarity correlates well with related bacterial hosts. HostPhinder is available as an interactive web service [1] and as a stand alone download from the Docker registry [2].

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark
Number of pages: 22
Publication date: 2016
Peer-reviewed: Yes

Publication information
Journal: Viruses
Volume: 8
Issue number: 5
Article number: 116
ISSN (Print): 1999-4915
Ratings:
Web of Science (2019): Indexed yes
Web of Science (2018): Indexed yes
Scopus rating (2017): CiteScore 3.88 SJR 1.805 SNIP 1.13
Web of Science (2017): Impact factor 3.761
Web of Science (2017): Indexed yes
Scopus rating (2016): CiteScore 3.6 SJR 1.747 SNIP 1.02
Web of Science (2016): Impact factor 3.465
Web of Science (2016): Indexed yes
Scopus rating (2015): CiteScore 3.74 SJR 1.832 SNIP 1.034
Web of Science (2015): Impact factor 3.042
Web of Science (2015): Indexed yes
Scopus rating (2014): CiteScore 3.8 SJR 1.906 SNIP 1.098
Web of Science (2014): Impact factor 3.353
Scopus rating (2013): CiteScore 3.41 SJR 1.642 SNIP 0.979
Web of Science (2013): Impact factor 3.279
Scopus rating (2012): CiteScore 2.67 SJR 1.152 SNIP 0.686
Web of Science (2012): Impact factor 2.509
Scopus rating (2011): CiteScore 1.63 SJR 0.72 SNIP 0.439
Web of Science (2011): Impact factor 1.5
Scopus rating (2010): SJR 0.446 SNIP 0.21
Web of Science (2010): Impact factor 1
Original language: English
Keywords: Prediction, Genome, K-mers, "Host specificity"
Electronic versions:
HostPhinder.pdf
DOIs:
10.3390/v8050116

Bibliographical note
This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.