Genomics Proteomics Bioinformatics 13 (2015) 296–303

HOSTED BY

ELSEVIER

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

RESOURCE REVIEW

Web Resources for Metagenomics Studies

Pravin Dudhagara [1,*,a], **Sunil Bhavsar** [1,b], **Chintan Bhagat** [1,c], **Anjana Ghelani** [2,d], **Shreyas Bhatt** [3,e], **Rajesh Patel** [3,f]

[1] *Department of Biosciences, Veer Narmad South Gujarat University, Surat 395007, India*
[2] *Department of Microbiology, Shree Ramkrishna Institute of Computer Education and Applied Sciences, Surat 395001, India*
[3] *Department of Life Sciences, Hemchandracharya North Gujarat University, Patan 384265, India*

Received 6 August 2015; revised 9 October 2015; accepted 13 October 2015
Available online 18 November 2015

Handled by Vladimir B. Bajic

KEYWORDS

Metagenomics;
Metagenomes;
Web resources;
Software tools;
Synthetic metagenome

Abstract The development of next-generation sequencing (NGS) platforms spawned an enormous volume of data. This explosion in data has unearthed new scalability challenges for existing bioinformatics tools. The analysis of metagenomic sequences using bioinformatics pipelines is complicated by the substantial complexity of these data. In this article, we review several commonly-used online tools for **metagenomics** data analysis with respect to their quality and detail of analysis using simulated **metagenomics** data. There are at least a dozen such **software tools** presently available in the public domain. Among them, MGRAST, IMG/M, and METAVIR are the most well-known tools according to the number of citations by peer-reviewed scientific media up to mid-2015. Here, we describe 12 online tools with respect to their web link, annotation pipelines, clustering methods, online user support, and availability of data storage. We have also done the rating for each tool to screen more potential and preferential tools and evaluated five best tools using **synthetic metagenome**. The article comprehensively deals with the contemporary problems and the prospects of **metagenomics** from a bioinformatics viewpoint.

Introduction

Metagenomics has become an emerging and indispensable field of microbial ecology since the first use of the concept in 1998 [1]. It deals with the study of genetic material recovered directly from environmental samples. Originally, metagenomics studies depended on the cultivation of clonal cultures followed by functional expression screening [1], which are not able to represent the total community profile and may overlook the vast majority of the microbial biodiversity [2]. The advent of next-generation sequencing (NGS) technologies has made it possible to perform sequencing-based

---

[*] Corresponding author.
E-mail: dudhagarapr@gmail.com (Dudhagara P).
[a] ORCID: 0000-0002-4041-6751.
[b] ORCID: 0000-0002-7525-0624.
[c] ORCID: 0000-0002-2318-5998.
[d] ORCID: 0000-0003-4638-6925.
[e] ORCID: 0000-0002-4364-2244.
[f] ORCID: 0000-0002-4712-1921.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

http://dx.doi.org/10.1016/j.gpb.2015.10.003
1672-0229 © 2015 The Authors. Production and hosting by Elsevier B.V. on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

metagenomic analyses. NGS platforms allow researchers to retrieve high-throughput data at substantially lower cost, fueling the development of sequence-based metagenomics analyses aimed at decoding the entire microbial community [3].

Two distinct metagenomics approaches are commonly used: the first is referred to as marker-gene metagenomics or targeted metagenomics, and the second is shotgun metagenomics. The former involves the use of marker genes (including conserved markers such as 16S rRNA [4,5], 18S rRNA, and ITS [6]) and an amplicon sequencing method to reassemble the taxonomic structure of a microbial community [7,8]. For the latter approach, the random nature of shotgun sequencing ensures adequate coverage to assess the entire community's structure [9,10] and opens new avenues for discovering structural and functional novelties [11]. At present, thousands of marker-gene assisted and shotgun metagenomics projects have been undertaken, comprising millions of samples available in the public domain. After sequencing of a metagenomics sample, the primary task is to analyze the vast amount of data in order to determine the identities and roles of the microbial species present in the sample [12]. Several offline tools are available for classifying metagenomic reads against known reference datasets [13]. However, the assembly of metagenomic sequence data and the identification of the operational taxonomic units (OTUs) are major technical challenges that persist in metagenomic data analysis pipelines [14]. Moreover, the high complexity of metagenomics data is a critical barrier in analyzing these metadata using stand-alone tools. On the other hand, various web-based workflows may be able to offer in-depth exploration without the need for highly configured computers.

In this article, we review the most useful online software tools for metagenomics data analysis, including a brief description of how they work in the course of the analysis. More importantly, we have rated the tools on the basis of overall users' experience in five vital criteria: (1) ease of data uploading, (2) availability of online user support, (3) spectrum of data analyses available, (4) number of citations in scientific literature, and (5) size of the stored data (**Figure 1**, **Table 1**). While such an evaluation can never be completely thorough, we nevertheless hope that our evaluation of these web resources for metagenomics studies will provide vital information for budding scientists, assisting them to make the best choice for their preferred applications.

MG-RAST

The software package Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST) is a user-friendly open-source server initially constructed on the SEED project (http://www.theseed.org/) [15] framework for metagenomics studies [16–18]. Debuting in 2007, MG-RAST was one of the earliest online metagenomics analysis tools [16]. A new version (GitHub embedded version 3.6) is currently available, which does not entirely depend on SEED technology but uses the SEED subsystem instead as a preferred data source, to enable taxonomic and functional classification of metagenomes. The new version is encapsulated and separated from the data store, allowing far greater scalability and much faster computation than previous versions. Currently, MG-RAST contains 215,773 metagenomic datasets, which have been accessed by more than 12,000 registered users. Other than data access, registered users can also submit their own raw metagenomic data in FASTA, FASTQ, and SFF formats along with detailed sample information. The uploaded data are processed by a multi-step workflow that includes quality control, automated annotation, and analysis of prokaryotic metagenomic shotgun samples as well as targeted samples. Automated annotation is done using numerous reference datasets. The server supports a variety of user-driven analyses, including phylogenetic, functional, metabolic, and comparative analyses of two or more metagenomes. MG-RAST also provides the facility to select a database for quantitative insights into microbial populations based on sequencing data. It also offers the ability to output data in multiple clustering forms that are downloadable in flat

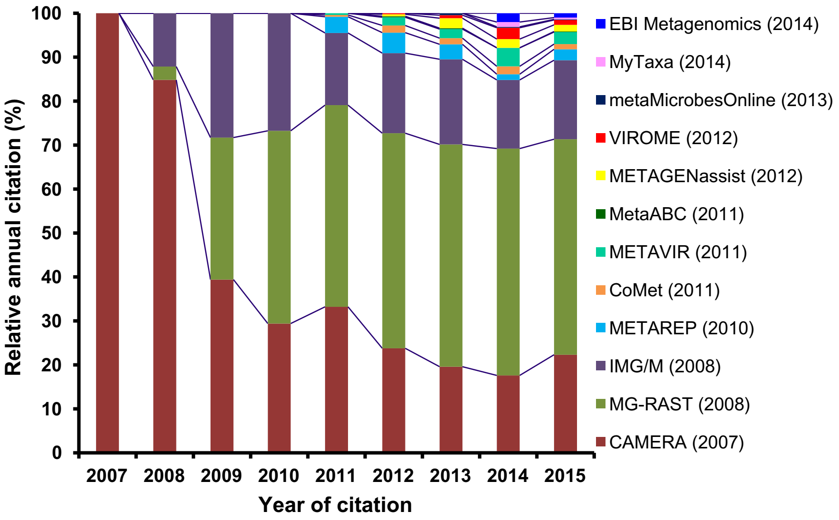

Figure 1 Relative citation of the metagenomics software tools from articles published in peer-reviewed scientific journals by October 2015
Year in the bracket in the legend box indicates the year of original release of the respective tool. Citation of each tool was tracked from Google Scholar. Total citation of all tools is considered as 100% in each year and relative percentage of citations of each tool per year was calculated in relative to the total citations of all tools in the respective year.

Download English Version:

https://daneshyari.com/en/article/2822379

Download Persian Version:

https://daneshyari.com/article/2822379

Daneshyari.com