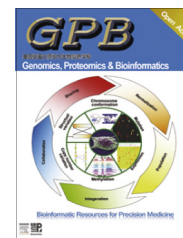




Genomics Proteomics Bioinformatics

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



RESOURCE REVIEW

Web Resources for Microbial Data



Qinglan Sun ^{#,a}, Li Liu ^{#,b}, Linhuan Wu ^c, Wei Li ^d, Quanhe Liu ^e,
Jianyuan Zhang ^f, Di Liu ^g, Juncai Ma ^{*,h}

Information Network Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

Received 9 January 2015; revised 22 January 2015; accepted 27 January 2015

Available online 23 February 2015

Handled by Fangqing Zhao

KEYWORDS

Microorganism;
Web resource;
Bioinformatics tools;
Rating

Abstract There are multitudes of web resources that are quite useful for the microbial scientific research community. Here, we provide a brief introduction on some of the most notable microbial web resources and an evaluation of them based upon our own user experience.

Introduction

Microorganisms are found virtually everywhere in the environment and serve as important biological resources. Some microorganisms have become widely utilized in industrial production. With the rapid development of high-throughput sequencing technology in recent years, it is considerably easier to obtain whole genome sequence of microorganisms. To date, thousands of microbes have been sequenced. Due to the difficulties of isolating single microbial species, high-throughput

sequencing technology has been applied to mixtures of environmental microbes to obtain metagenome sequencing data. One metagenomics endeavor is the Human Microbiome Project (HMP), which aims to obtain metagenomic sequencing data from a large number of human subjects to enhance our understanding of the relationship between the microbiome and human health. In this review, we provide a brief introduction on some of the most useful microbial web resources (Table 1), including information on collection of microbial cultures, species identification, literature, patent resources, microbial genomics, and metagenomics, as well as tools for analyzing genomic and metagenomic data. In addition, we have evaluated their function and provided ratings for each resource based on our own user experience.

Integrated Microbial Genomes

The Integrated Microbial Genomes (IMG, <http://img.jgi.doe.gov/>) system [1] is a combined web resource of microbial genome datasets, including genome sequences and gene annotations. The aim of the IMG is to provide free microbial genomic data, together with integrated annotations and comparative analysis services, to scientists worldwide. At the end

* Corresponding author.
E-mail: ma@im.ac.cn (Ma J).

Equal contribution.

^a ORCID: 0000-0002-8451-760X.

^b ORCID: 0000-0001-6977-1004.

^c ORCID: 0000-0002-5255-1846.

^d ORCID: 0000-0003-3364-8690.

^e ORCID: 0000-0001-8263-9446.

^f ORCID: 0000-0003-4215-5214.

^g ORCID: 0000-0003-3693-2726.

^h ORCID: 0000-0001-6382-8014.

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.2015.01.008>

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-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1 Important web resources for microbial research

Name	Link	Main features	Rating	Refs.
IMG	http://img.jgi.doe.gov/	The integrated microbial genomes and annotation	★★★★★	[1,2]
RAST	http://img.jgi.doe.gov/	Gene prediction and annotation tool for microbial genome	★★★★★	[3,4]
MG-RAST	http://metagenomics.anl.gov/	Analysis and annotation tool for metagenome	★★★★★	[5]
HMP-DACC	http://metagenomics.anl.gov/	Data center for the Human Microbiome Project	★★★★★	[6]
PATRIC	http://atricbre.org/portal/portal/patric/Home	The Integrated pathogen genomes and annotation	★★★★★	[7]
WDCM-CCINFO	http://www.wfcc.info/ccinfo	Worldwide culture collection for microorganisms	★★★★☆	
GCM	http://gcm.wfcc.info	Global catalog of microorganism resource and storage information	★★★★☆	[8]
ABC	http://www.wfcc.info/abc	Papers and patents of microorganisms	★★★★☆	
Website for avian flu information	http://www.avian-flu.info	Data center of avian flu	★★★★☆	[9]
GMS-DBD	http://www.boldmirror.net	DNA barcode data for species identification	★★★★☆	[10]

Note: More stars in the rating column indicate higher importance and usefulness.

of 2014, there were over 10,000 microbial genomes in IMG. Approximately 9000 of them were from GenBank, with the remaining ones submitted by registered users or institutes. Genome sequencing data stored in IMG are annotated using several functional references, such as the Clusters of Orthologous Groups (COG), Pfam, KEGG, TIGRfam, MetaCyc, and the Gene Ontology (GO), providing valuable information for registered users. In addition to integrating microbial genome data, IMG also provides the tool for analyzing publicly-available environmental microbial samples and metagenomes, which is named IMG/M [2].

Rapid Annotation using Subsystem Technology

The Rapid Annotation using Subsystem Technology (RAST, <http://rast.nmpdr.org/>) [3,4] is an automated online tool for analyzing microbial genomes. RAST screens the genome sequence submitted by the registered users and predicts gene-coding sequences, rRNA, and tRNA using glimmer2, search_for_rnas, and tRNAscan-SE, respectively. RAST then automatically produces two classes of functional annotations for the predicted gene sequences. One is a subsystem-based functional annotation that depends on recognition of functional variants of subsystems. The other is a nonsubsystem-based functional annotation, which considers combined evidence from a number of tools. RAST integrates these two annotations and provides exceptionally strong gene annotation results. RAST also uses the gene annotation results for metabolic pathway reconstruction, which makes the resource useful for comprehensive annotation efforts. RAST also excels at making collections of functionally related protein families (FIGfams). When a new genome is submitted to RAST and made public by the submitter, annotated genes are compared and added to the FIGfams collection. The expanding FIGfams collection has proven to be a robust and scalable solution to microbial genome annotation efforts. To date, over 12,000 users have registered in RAST and have submitted over 60,000 distinct microbial genomes for annotation.

Metagenomics RAST

The Metagenomics RAST (MG-RAST, <http://metagenomics.anl.gov/>) server [5] is an open online system for management and comparative analysis of metagenomic data. Registered users can submit raw sequencing data in FASTA format, along with sampling information into the system. The uploaded sequences are subsequently processed and analyzed. Summaries of the raw data as well as the analyzed results are generated automatically. The MG-RAST server can manage many different types of data including phylogenetic and metabolic data. In addition to analyzing single-sample datasets, the MG-RAST can also provide comparative analysis for multiple metagenomes and genomes. The metagenome annotation and comparative analysis systems are designed such that tools and/or new data can be added or replaced at any stage of the analysis process as needed to accommodate new methods. In order to protect data privacy, user access is controlled. Registered users retain full control of their data. Nonetheless, collaboration and sharing of data between multiple users are both possible and encouraged within the MG-RAST server.

HMP – Data Analysis and Coordination Center

The Data Analysis and Coordination Center (DACC, <http://hmpdacc.org/>) is the data center for the HMP [6]. It includes sampling information as well as microbial genome and metagenome sequences. The HMP was launched in 2008 and funded by the National Institutes of Health (NIH). The main objective of this resource is to produce microbial data, which enhances our understanding of the relationship between the human microbiome and human health. The HMP has investigated the microorganisms from multiple locations on the human body, including the gastrointestinal tract, oral cavity, nasal passages, skin, and urogenital tract. The Project has produced a significant amount of genomic and metagenomic data.

Download English Version:

<https://daneshyari.com/en/article/2822455>

Download Persian Version:

<https://daneshyari.com/article/2822455>

[Daneshyari.com](https://daneshyari.com)