

Laboratorio 04 - Metagenómica

En este laboratorio vamos a familiarizarnos con la plataforma MG-RAST para el análisis de metagenomas.

[MG-RAST](#) (metagenomics Rapid Annotation using Subsystem Technology) es una plataforma en línea de libre acceso que provee un conjunto de herramientas para el análisis y la visualización de metagenomas. También, MG-RAST funciona como base de datos donde los usuarios pueden alojar sus datos metagenómicos de manera pública o privada.

Plataforma MG-RAST y exploración de metagenomas

Comencemos por ir a la página donde vamos a trabajar hoy día

- Ve a MG-RAST [aquí](#). Si no estás usando Mozilla te va a aparecer un aviso odioso que dice que el sitio ha sido optimizado para ese explorador. En la práctica no debería haber problemas con usar Safari o Chrome.

metagenomics.anl.gov

MG-RAST

metagenomics analysis server

Warning: This application has been optimized for the Firefox browser. Since you are using Safari, many features will not be available and / or behave incorrectly.
Firefox is freely available [here](#).

We have solved the database issue and are now performing cleanup. Some jobs will appear incomplete for a little while. We are fixing all jobs in the next few days.

Browse Metagenomes

[Register](#) [Contact](#) [Help](#) [Upload](#) [News](#)

About

MG-RAST (the Metagenomics RAST) server is an automated analysis platform for metagenomes providing quantitative insights into microbial populations based on sequence data.

# of metagenomes	245258
# base pairs	104.87 Tbp
# of sequences	821 billion
# of public metagenomes	34,929

The server primarily provides upload, quality control, automated annotation and analysis for prokaryotic metagenomic shotgun samples. MG-RAST was launched in 2007 and has over 12,000 registered users and 245258 data sets. The current server version is 3.6. We suggest users take a look at [MG-RAST for the impatient](#). Also available for download is the [MG-RAST manual](#).

- MG-RAST newsletter, August 2015
- Upcoming change to MG-RAST upload (early August 2015)
- MG-RAST API available
- MG-RAST newsletter, September 2014

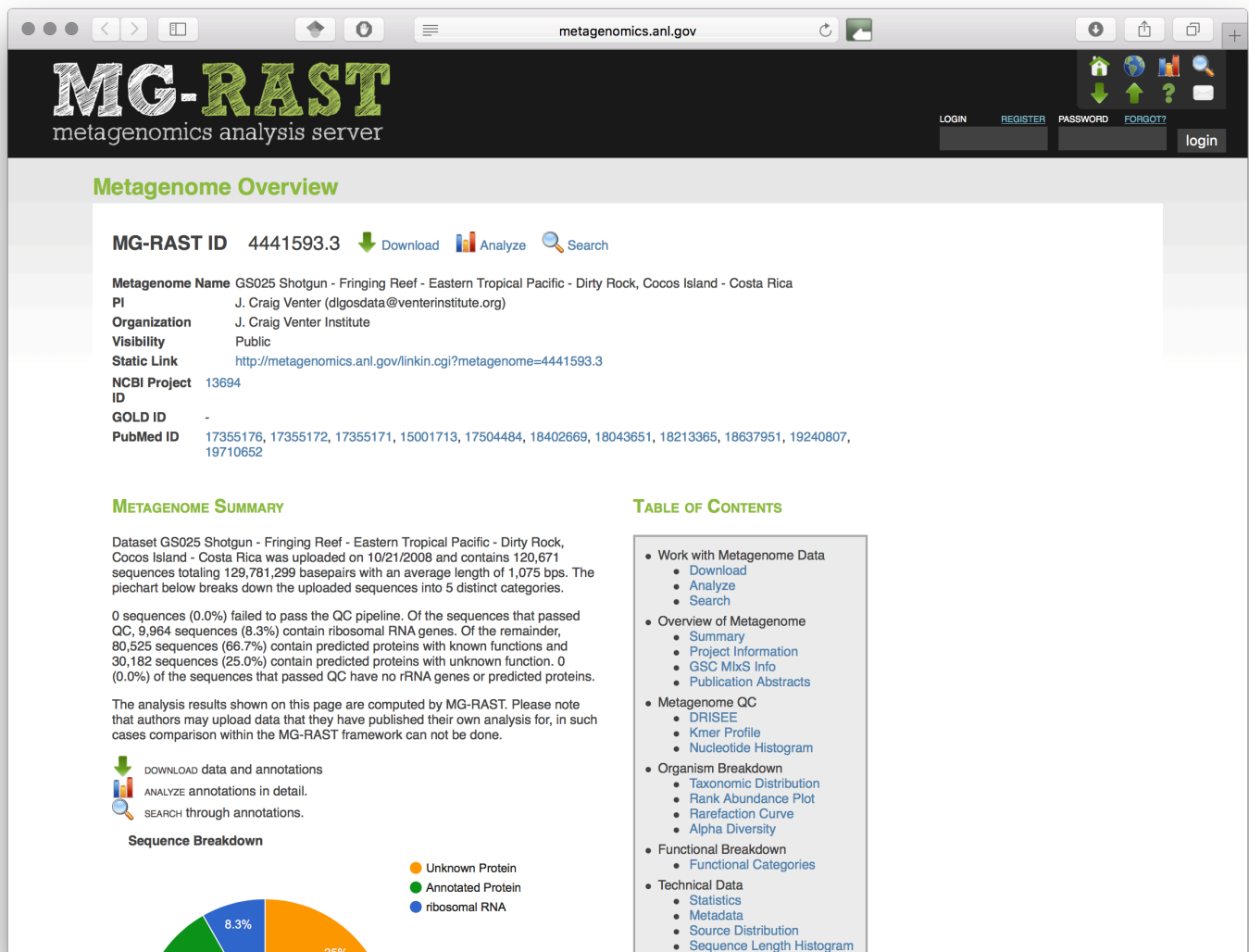
* login required

This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900040C.

This work was supported in part by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

[cite MG-RAST](#) [cite MG-RAST API](#)

- Dirígete hacia el ícono del planeta tierra en la esquina superior derecha
- La página resultante muestra una tabla con funciones básicas de búsqueda y filtrado. Tómate unos minutos para explorar la tabla y darte una idea de qué tipos de muestras están disponibles para análisis.
- Ahora vamos a ubicar y examinar dos proyectos metagenoma del [JVCI Global Ocean Sampling Expedition](#).
- Usando las casillas de búsqueda en la tabla, ubica y selecciona la muestra tomada del sitio de buceo [Dirty Rock](#) en las Islas Cocos de Costa Rica.



Responde las siguientes preguntas asociadas a la muestra:

¿Cuántas secuencias fueron subidas?

¿Qué tipo de plataforma de secuenciamiento fue usada para producir estos datos?

¿Cuántas secuencias pasaron el control de calidad? ¿A qué crees que se refiere esto?

¿Cuántos otros metagenomas están disponibles para el bioma "marine habitat"?

¿Cuántas reads fueron asignadas a eucariontes según MG-RAST?

¿Cuál es el filo (Phylum) más abundante en la muestra?

¿A cuántas secuencias se les asignó una función de acuerdo a la base de datos KEGG con un e-value entre 10^{-10} y 10^{-20} ?

¿Por qué hay tan pocas secuencias con funciones asignadas según SwissProt?

Análisis de metagenomas

Dentro de la página para la muestra Dirty Rock, haz clic en el ícono de gráfico de barras donde dice "Analyze". Deberías ver una ventana con las opciones como aparecen abajo.

MG-RAST
metagenomics analysis server

LOGIN REGISTER PASSWORD FORGOT? login

Metagenome Analysis

1 Data Type

ORGANISM ABUNDANCE

Representative Hit Classification

» Best Hit Classification

Lowest Common Ancestor

FUNCTIONAL ABUNDANCE

Hierarchical Classification

All Annotations

2 Data Selection

Metagenomes 4441593.3

Annotation Sources

Max. e-Value Cutoff 1e-5

Min. % Identity Cutoff 60 %

Min. Alignment Length Cutoff 15

Workbench

☐ use features from workbench

3 Data Visualization

☐ barchart ☐ tree ☒ table ☐ heatmap ☐ PCoA ☐ rarefaction

generate

Workbench (0 Features) Getting Started

To create a visualization, first select an analysis view from the **Analysis Views** box. The default is 'Organism Classification'. Then choose the data and cutoffs you wish to use in the **Data Selection** box. Depending on the type of data, you might have a set of possible visualizations. Pick one of them and click the **generate** button.

You will see the generated visualizations created in separate tabs in this tab-view. In addition to the visualization, the created tabs will display the settings used to create them. Generating a new visualization will preserve the previous visualizations you created.

You can rename the tabs by double clicking the tab-header and entering a new title. You can remove a tab by clicking the 'x' symbol in the top right corner of the tab-header. You can switch between tabs by single clicking the tab-header.

- Asegúrate de que el e-value esté configurado en -10 y selecciona la visualización de gráfico de barras (la primera).
- Haz clic en "Generate"
- Cuando tu análisis termine, deberías ver una nueva pestaña que dice "Organism barchart 1"
- Una vez que aparezca el gráfico de barras, haz clic en Bacteria. La distribución de Fila (Phylum) será visible

Responde las siguientes preguntas asociadas al análisis:

¿Cuántas secuencias mapearon en contra de Proteobacteria? (Usa la opción Redraw para usar reads en vez de proporciones (raw))

¿Cuántas secuencias de Salmonella (Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae) fueron identificadas?

Trabajo de laboratorio para la próxima semana

El trabajo de laboratorio para la próxima semana consta de dos partes. La primera parte ya la tienes lista. Simplemente tienes que responder las preguntas que aparecen a través de esta guía y enviar un informe a bioinformatica.unab2016@gmail.com NO al profesor! Envíen a este correo los informes hasta el jueves de la semana siguiente a la realización del práctico. La hora límite es las 23:59. En el Asunto del correo pongan Informe de Laboratorio 0x así nosotros podemos clasificar automáticamente los informes. Para la próxima semana el Asunto del correo debería ser Informe de Laboratorio 04.

Profesor	Nombre	Correo electrónico
Profesor responsable	Dr. Eduardo Castro Nallar	eduardo.castro@unab.cl
Profesor ayudante sección 1	Ingrid Araya Durán	ingrid.araya.duran@gmail.com
Profesor ayudante sección 1	Sandro Valenzuela	sandrolvalenzuelad@gmail.com
Profesor ayudante sección 2	Javier Cáceres	ja.caceresmolina@gmail.com
Profesor ayudante sección 2	Consuelo Bello	consuelobelloz@gmail.com

La segunda parte tiene que ver con leer el artículo Jansson J, Field D, Fierer N, et al. ***Unlocking the potential of metagenomics through replicated experimental design***. Nature Publishing Group. 2012;30(6):513-520. doi:10.1038/nbt.2235.

Puedes acceder al artículo [aquí](#). Recuerda que el contenido de este artículo es el material para el próximo control de entrada.