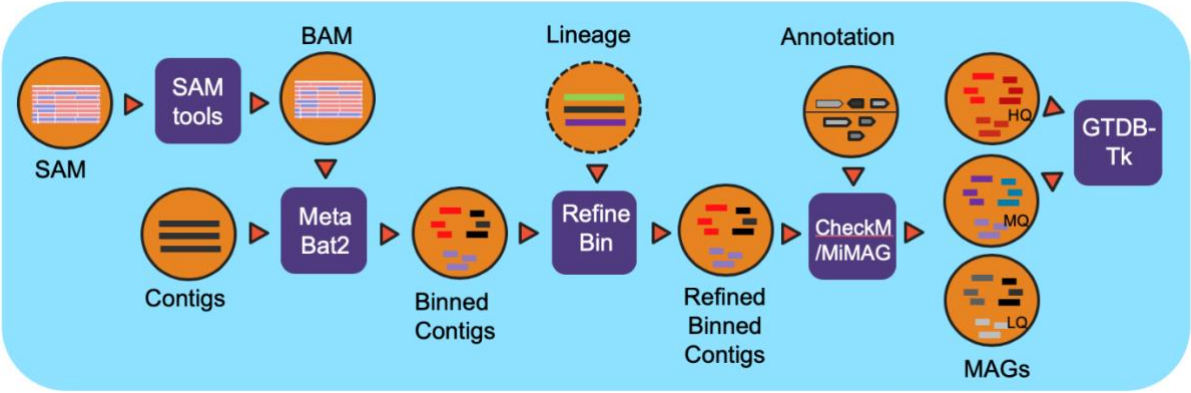
Workflow de Genomas Ensamblados Metagenoma (MAGs) (v1.0.2)



Descripción General

Este workflow clasifica los 'contigs' en 'MAG' o genomas ensamblados de metagenomas (MAGs), y con los MAG resultantes se refina utilizando el archivo de anotación funcional. Los MAGs se evalúan para determinar su integridad y contaminación. Además, se determina la calidad de los MAGs y se asigna un linaje a cada bin de calidad alta o media.

Ejecutando el Workflow

Actualmente, este workflow se puede ejecutar en NMDC EDGE o desde la línea de comandos. (Las instrucciones y requisitos de CLI se encuentran aquí).

Aquí se encuentran videos tutoriales sobre cómo ejecutar cada workflow en NMDC EDGE.

Input

Este workflow requiere 'contigs' ensamblados en un archivo FASTA, el archivo de mapeo leído del ensamblaje (SAM o BAM), y una anotación funcional del ensamblaje en un archivo GFF.

- **Formatos de archivo aceptables:** 'contigs' ensamblados (.fasta, .fa o .fna); leer el mapeo al ensamblaje (.sam.gz o .bam); Anotación funcional (.gff)

Detalles

El workflow se basa en el proceso de agrupación de metagenomas IMG y se ha modificado específicamente para el proyecto NMDC. Para todos los metagenomas procesados, clasifica los 'contigs' en MAGs utilizando MetaBat2. A continuación, los MAG se refinan utilizando el archivo de anotación funcional (GFF) del workflow de 'Metagenome Annotation' (Anotación de Metagenoma) y la información de linaje 'contig' opcional. CheckM evalúa la integridad y la contaminación presente en los MAGs y les asigna un nivel de calidad (alta calidad (HQ), calidad media (MQ), baja calidad (LQ)) según los estándares MiMAG. Al final, GTDB-Tk se utiliza para asignar el linaje de los MAGs HQ y MQ. La base de datos GTDB-Tk requerida está incorporada en NMDC EDGE.

Versiones de Software

- Biopython v1.74
- Sqlite
- Pymysql
- requests
- samtools > v1.9 (License: MIT License)
- Metabat2 v2.15

LA-UR-21-21661

- CheckM v1.1.2
- GTDB-TK v1.2.0
- FastANI v1.3
- FastTree v2.1.10

Productos

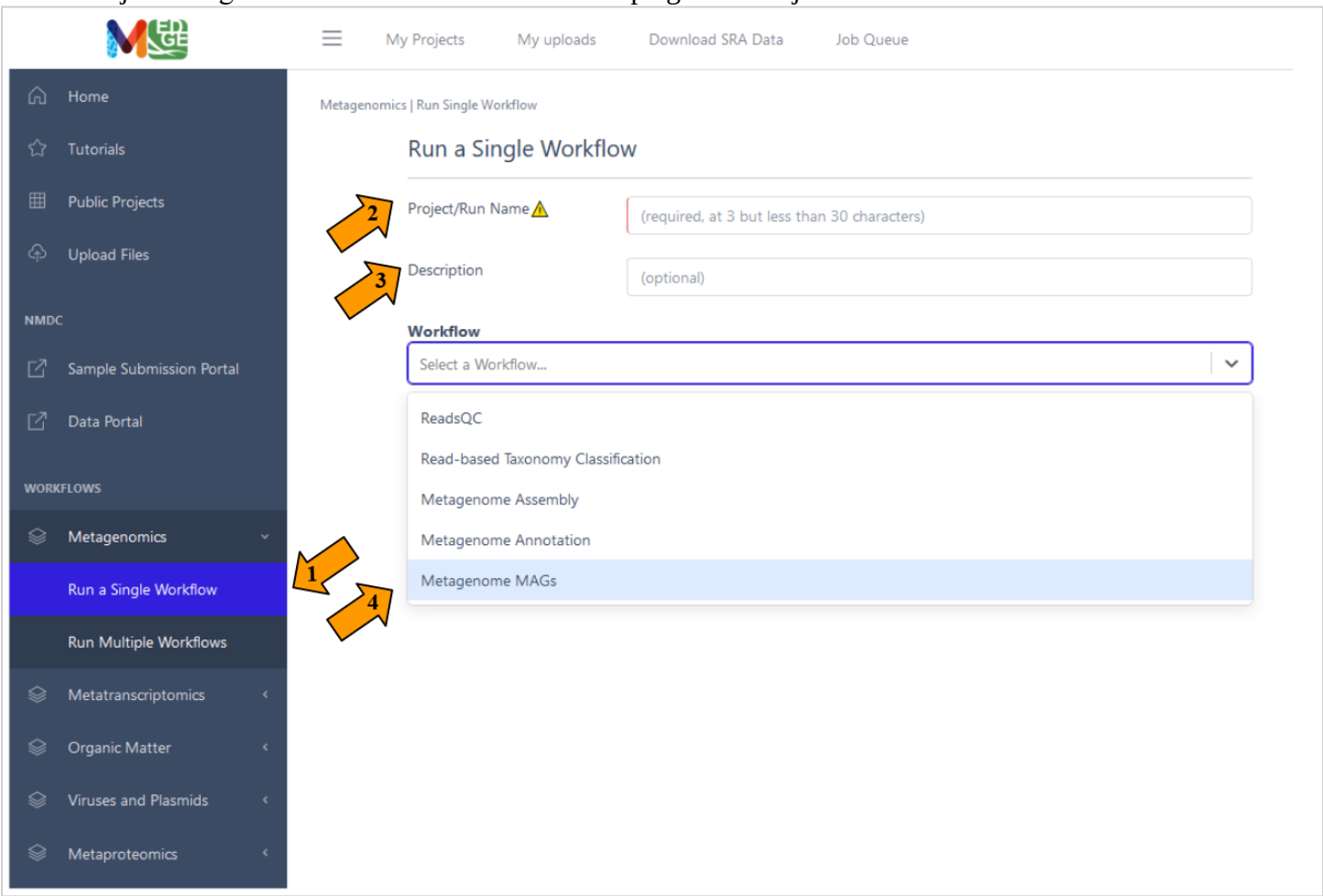
El resultado principal es el archivo comprimido de MAG de alta calidad (HQ) y calidad media (MQ). También están disponibles 'contigs' sin agrupar y MAG de baja calidad.

Archivos de Salida Primarios	Descripción
hqmq-metabat-MAG.zip	MAGs clasificados como calidad alta o media

Ejecutando el 'Metagenome Assembled Genomes (MAGs) Workflow' en NMDC EDGE

Elije un workflow

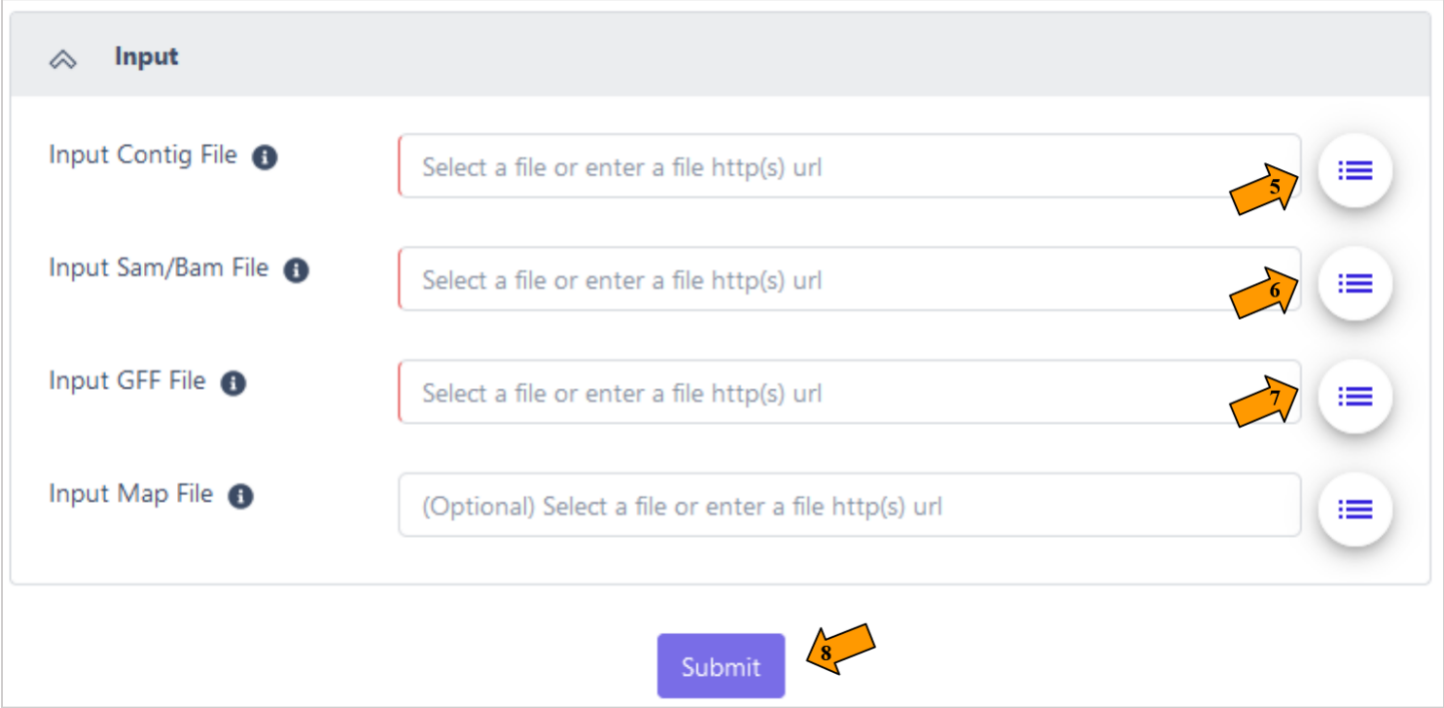
1. Des de la categoría 'Metagenomics' (Metagenómica) en la barra de menú de la izquierda, elige 'Run a Single Workflow' (Ejecuta un Workflow Individual).
2. Un nombre de proyecto/ejecución ***única*** sin espacios (los guiones bajos están permitidos).
3. Una descripción es opcional, pero recomendada.
4. Elije 'Metagenome MAGs' desde el menú desplegable debajo de Workflow.



Input

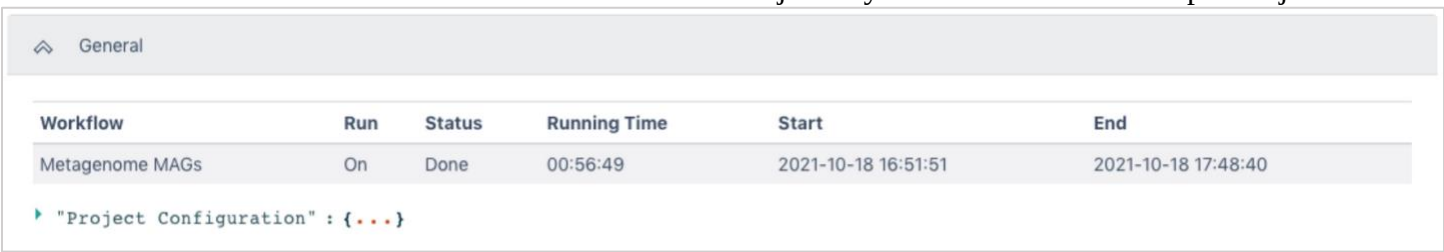
Los Genomas Ensamblados de Metagenomas (MAGs) requieren 'contigs' ensamblados, el archivo de mapeo de ‘reads’ a conigs ensamblados y un archivo de anotación funcional. La entrada recomendada sería la de los workflows de ensamblaje y anotación del NMDC ('Metagenome Assembly' y 'Metagenome Annotation').
Formatos de archivo aceptables: 'contigs' ensamblados (.fasta, .fa o .fna); leer el mapeo al ensamblaje (.sam.gz o .bam); anotación funcional (.gff).

5. Haga clic en el botón a la derecha del espacio en blanco para 'Input Contig File' (Entrada de Archivo 'Contig'). Se abrirá un cuadro llamado ‘Select a File’ (Seleccionar un Archivo) para permitir al usuario encontrar el archivo deseado de un proyecto de ensamblaje ejecutado previamente, la carpeta de datos públicos o un archivo cargado por el usuario.
6. Haga clic en el botón a la derecha del espacio en blanco para 'Input Sam/Bam file' (Entrada de Archivo Sam/Bam). Se abrirá un cuadro llamado 'Select a File' (Seleccionar un Archivo) para permitir al usuario encontrar el archivo de mapeo leído de un proyecto de ensamblaje ejecutado previamente, la carpeta de datos públicos o un archivo cargado por el usuario.
7. Haga clic en el botón a la derecha del espacio en blanco para 'Input GFF File' (Entrada de Archivo GFF). Se abrirá un cuadro llamado 'Select a File' (Seleccionar un Archivo) para permitir al usuario encontrar los archivos deseados de un proyecto de anotación ejecutado anteriormente, la carpeta de datos públicos o un archivo cargado por el usuario.
8. Haga clic en 'Submit' (Someter).



Productos

La sección General del resultado muestra cual workflow se ejecutó y la información del tiempo de ejecución.



La sección ‘Metagenome MAGs Result’ (Resultado de Metagenome MAG) proporciona una sección de ‘Summary’ (Resumen) con información sobre ‘contigs’ agrupados y no agrupados. La sección MAGs proporciona información como la integridad del genoma, la cantidad de contaminación y la cantidad de genes presentes en todos los MAG que se consideran de calidad alta o media.

Metagenome MAGs Result

Summary

Name	Status
input_contig_num	25,726
too_short_contig_num	15,158
lowDepth_contig_num	0
unbinned_contig_num	9,334
binned_contig_num	1,234

MAGs

bin_name	number_of_contig	completeness	contamination	gene_count	bin_quality	num_16s	num_5s	num_23s	num_tRNA	gtdbtk_domain	gtdbtk_phy
bins.1	63	99.48	0.16	4,826	HQ	1	4	1	76	Bacteria	Proteobacte
bins.4	35	99.68	0.61	6,653	MQ	0	0	0	63	Bacteria	Proteobacte
bins.6	17	99.45	5.19	3,575	MQ	1	3	0	71	Bacteria	Firmicutes
bins.3	29	78.23	0	1,248	MQ	0	1	0	28	Bacteria	Firmicutes
bins.5	7	68.97	0	1,819	MQ	0	1	0	24	Bacteria	Firmicutes
bins.7	18	62.07	0	3,161	MQ	2	3	2	54	Bacteria	Firmicutes

La sección ‘Browser/Download Output’ (Navegador/Descargar Productos) resultados proporciona archivos de resultados disponibles para descargar. El archivo de salida principal es el archivo comprimido con todos los MAG determinados como de calidad alta o media (hqmq-metabat-MAG.zip).

Browser/Download Outputs

File	Size	Last Modified
MetagenomeMAGs		
activity.json	7 kB	16 days ago
bins.lowDepth.fa	0 B	16 days ago
bins.tooShort.fa	8.68 MB	16 days ago
bins.unbinned.fa	16.82 MB	16 days ago
checkm_qa.out	3 kB	16 days ago
data_objects.json	2 kB	16 days ago
gtdbtk.bac120.summary.tsv	11 kB	16 days ago
hqmq-metabat-bins.zip	6.04 MB	16 days ago
MAGs_stats.json	7 kB	15 days ago
metabat-bins.zip	1.81 MB	16 days ago