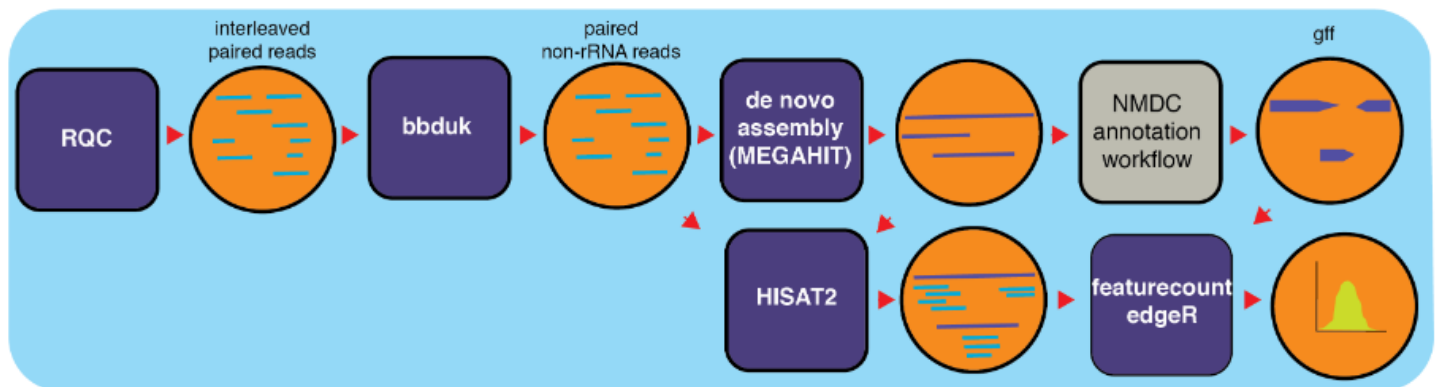


## Workflow Metatranscriptómica (v0.0.2)



### Descripción General

El workflow del metatranscriptoma ('metaT') toma datos 'crudos' del metatranscriptoma, filtra la calidad de los datos, elimina las 'reads' de rRNA y luego ensambla y anota el ensamblaje de metatranscriptoma. Los datos se asignan a características genómicas del ensamblaje y los RPKM (lecturas por kilobase de transcripción por millón de lecturas asignadas) se calculan para cada característica en el archivo de anotación funcional.

### 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

La metatranscriptómica requiere datos de Illumina de extremos emparejados ('paired-end') como un archivo entrelazado o como pares separados de archivos FASTQ.

- **Formatos de archivo aceptables:** .fastq, .fq, .fastq.gz, .fq.gz

### Detalles

MetaT es un workflow diseñado para analizar metatranscriptomas, y este workflow se basa en otros workflows de NMDC para procesar datos de secuenciación. El workflow de metatranscriptómica toma datos de secuenciación de ARN 'crudos' y filtra la calidad de las 'reads' utilizando el workflow 'ReadsQC'. Luego, el workflow MetaT filtra las 'reads' de ARN ribosómico (utilizando la base de datos de rRNA SILVA) y separa los archivos entrelazados en pares de archivos separados utilizando bbduk (BBTools). Después de los pasos de filtrado, los 'reads' se ensamblan usando MEGAHIT y se anotan usando el '[Metagenome Annotation NMDC Workflow](#)' (Anotación de Metagenoma) que produce archivos de anotación funcional GFF. Las características se cuentan con '[Subread's featureCounts](#)', que asigna 'reads' asignados a características genómicas y genera RPKM para cada característica en un archivo GFF para 'reads' con 'sense' y 'antisense'.

### Versions de Software

- BBTools v38.44
- hisat2 v2.1
- Python v3.7.6.
- featureCounts v2.0.1
- R v3.6.0
- edgeR v3.28.1

- pandas v1.0.5

- gffutils v0.10.1

## Productos

La siguiente tabla enumera los archivos de salida principales. Los resultados principales son las transcripciones recopiladas y el archivo de características anotadas. También se pueden descargar varios archivos de anotaciones.

Archivos de Salida Primarios	Descripción
INPUT_NAME.contigs.fa	Ensamblaje metatranscriptómico
rpkmsorted_features.tsv	Tabla de funciones ordenada por RPKM

## Ejecutando el 'Metatranscriptomics Workflow' en NMDC EDGE

### Elige un workflow

1. Desde la categoría 'Metatranscriptomics' (Metatranscriptómica) 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. Elige 'Metatranscriptome' desde el menú desplegable debajo de Workflow.

### Aporte

El workflow del metatranscriptoma requiere datos de Illumina de extremos emparejados ('paired-end') en formato FASTQ como input; el archivo puede ser entrelazado y/o comprimido. **Formatos de archivo**

**aceptables:** .fastq, .fq, .fastq.gz, .fq.gz

5. La configuración predeterminada es que los datos 'crudos' estén en formato entrelazado ('reads' emparejadas entrelazado en un archivo). Si los datos 'crudos' son 'reads' emparejadas en archivos separados (hacia adelante y hacia atrás), haga clic en 'No'.
6. Se pueden agregar archivos de datos adicionales (del mismo tipo, entrelazado o separados) con el botón abajo.
7. Haga clic en el botón a la derecha del espacio en blanco de entrada de datos para seleccionar el archivo de datos para el análisis. (Si tiene archivos separados, hay dos espacios en blanco para ingresar sus datos). Se abrirá un cuadro llamado 'Select a File' (Seleccionar un Archivo) para permitir al usuario encontrar el archivo deseado, archivos de proyectos ejecutados anteriormente, la carpeta de datos públicos o los archivos cargados 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 'Metatranscriptome Result' (Resultados del Metatranscriptoma) incluye una tabla de los 100 resultados más abundantes a base de RPKM ordenados por RPKM. Al seleccionar el título de cada columna, se ordenarán estos datos según esa columna. Esta sección también incluye un botón para descargar rápidamente un archivo tsv de todas las características detectadas en el conjunto de datos de entrada para su posterior análisis.

La sección 'Browser/Download Output' (Navegador/Descargar Productos) proporciona todos los archivos de salida disponibles para descargar. Los 'contigs' de salida se encontrará en la carpeta de ensamblaje y el archivo tsv de todas las características detectadas ordenadas por RPKM está disponible en carpeta metat\_output.