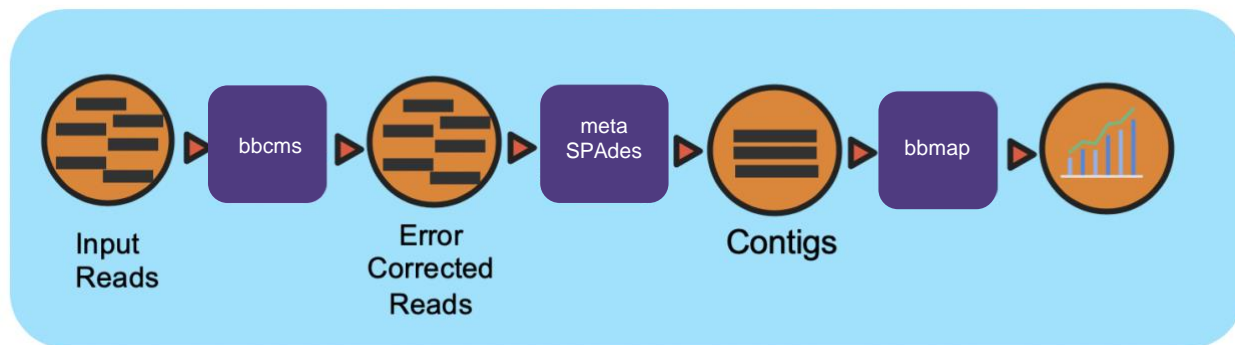


Workflow de Ensamblaje Metagenómica (v1.0.1)



Descripción General

Este workflow toma datos de Illumina emparejados (paired-end), ejecuta corrección de errores, ensamblaje y validación de ensamblaje.

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

El workflow 'Metagenome Assembly' (Ensamblaje Metagenómica) requiere datos de Illumina emparejados (paired-end) y puede aceptar datos como un archivo entrelazado o como pares separados de archivos FASTQ. Se recomienda ingresar datos del workflow 'ReadsQC'.

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

Detalles

Este workflow toma 'reads' de Illumina emparejadas y realiza corrección de errores mediante bbcms (BBTools). Los 'reads' corregidas se ensamblan utilizando metaSPAdes. Después del ensamblaje, bbmap (BBTools) asigna los 'reads' nuevamente a los 'contigs' para obtener información de cobertura.

Versiones de Software

- bbcms (BBTools:38.94)
- metaSpades (v3.15.0)
- bbmap (BBTools:38.94)

Productos

El resultado principal es el archivo 'contigs' ensamblado (assembly_contigs.fna).

Archivos de Salida Primarios	Descripción
'Contigs' Ensamblado	Ensamblaje Final de 'Contigs' (assembly.contigs.fna)

‘Scaffolds’ Ensamblado	Ensamblaje Final de ‘Scaffolds’ (assembly_scaffolds.fna)
AGP Ensamblado	Un archivo en formato AGP que describe el ensamblaje
Cobertura del Ensamblado BAM	Archivo BAM ordenado de ‘reads’ asignadas al ensamblaje final
Estadísticas de Cobertura del Ensamblado	Información de cobertura de ‘contigs’ ensamblados.

Ejecutando el ‘Metagenome Assembly Workflow’ en NMDC EDGE

Elige un workflow

1. Des de la categoría 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. Elige ‘Metagenome Assembly’ desde el menú desplegable debajo de Workflow.

The screenshot shows the NMDC EDGE web interface. On the left is a dark sidebar menu with categories like Home, Tutorials, Public Projects, Upload Files, NMDC, Sample Submission Portal, Data Portal, WORKFLOWS, Metagenomics, Run a Single Workflow (highlighted in blue), Run Multiple Workflows, Metatranscriptomics, Organic Matter, Viruses and Plasmids, and Metaproteomics. The main content area is titled 'Metagenomics | Run Single Workflow' and contains a form titled 'Run a Single Workflow'. The form has three input fields: 'Project/Run Name' (required, 3-30 characters), 'Description' (optional), and a 'Workflow' dropdown menu. The dropdown menu is open, showing a list of workflows: ReadsQC, Read-based Taxonomy Classification, Metagenome Assembly (highlighted in blue), Metagenome Annotation, and Metagenome MAGs. Four orange arrows with numbers 1-4 point to the following elements: 1 points to 'Run a Single Workflow' in the sidebar; 2 points to the 'Project/Run Name' field; 3 points to the 'Description' field; and 4 points to the 'Metagenome Assembly' option in the workflow dropdown.

Input

Este workflow acepta datos de Illumina en forma FASTQ como input; el archivo puede ser entrelazado y puede ser comprimido. Esta entrada puede ser la salida del workflow 'ReadsQC' y esto se recomienda. **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).

Input

Input Raw Reads ⓘ
Is interleaved? Yes No

Input interleaved fastq Add interleaved fastq

interleaved FASTQ #1 Select a file or enter a file http(s) url ⋮

Remove

Submit

Productos

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

General					
Workflow	Run	Status	Running Time	Start	End
Metagenome Assembly	On	Done	02:02:22	2021-10-13 22:51:35	2021-10-14 00:53:57
▶ "Project Configuration" : { ... }					

La sección 'Metagenome Assembly Result' (Resultado del Asamblea Metagenómica) tiene todas las estadísticas de la asamblea.

Name	Status
scaffolds	25,324
contigs	25,726
scaf_bp	52,206,897
contig_bp	52,201,077
gap_pct	0.011
scaf_N50	691
scaf_L50	4,103
ctg_N50	724
ctg_L50	3,971
scaf_N90	14,186
scaf_L90	726
ctg_N90	14,473
ctg_L90	716
scaf_logsum	645,093
scaf_powsum	120,098
ctg_logsum	638,015
ctg_powsum	116,432
asm_score	33.765
scaf_max	1,491,105
ctg_max	859,644
scaf_n_gt50K	96
scaf_l_gt50K	20,678,937
scaf_pct_gt50K	39.61
gc_avg	0.473
gc_std	0.062
filename	assembly_scaffolds.fna

La sección ‘Browser/Download Output’ (Navegador/Descargar Productos) proporciona archivos de resultados disponibles para descargar. El resultado primario es el archivo ‘assembly_contigs.fna’ (ensamblador_contigs.fna) que también puede ser la entrada para el workflow de ‘Metagenome Annotation’ (Anotación Metagenómica). El archivo ‘pairedMapped_sorted.bam’ junto con el archivo ‘contigs’ ensamblado pueden ser la entrada para el workflow de ‘MAGs Generation’ (Generación de MAG).

Browser/Download Outputs		
File	Size	Last Modified
MetagenomeAssembly		
assembly.agp	1.72 MB	21 days ago
assembly_contigs.fna	51.30 MB	21 days ago
assembly_scaffolds.fna	51.22 MB	21 days ago
covstats.txt	1.92 MB	21 days ago
pairedMapped.sam.gz	2338.54 MB	21 days ago
pairedMapped_sorted.bam	2130.75 MB	21 days ago
stats.json	619 B	21 days ago