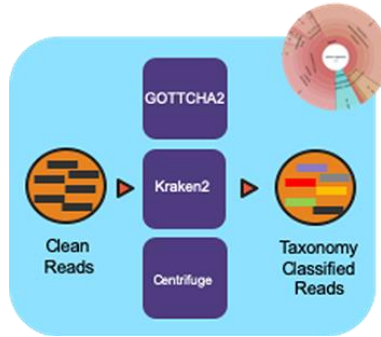


Clasificación de Taxonomía Basada en ‘reads’ Metagenómicos (v1.0.1)



Descripción General

Este workflow toma archivos de secuenciación de Illumina (de un solo extremo o de pares) y perfila los ‘reads’ utilizando múltiples herramientas de clasificación taxonómica.

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 Read-based Taxonomy Classification’ (Clasificación de Taxonomía Basada en ‘reads’ Metagenómicos) requiere datos de Illumina y puede aceptar datos como un archivo entrelazado o como pares separados de archivos FASTQ. Los datos entrelazados se tratarán como ‘reads’ de un solo extremo (single-end). Se recomienda ingresar datos ‘limpios’ del workflow ‘ReadsQC’.

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

Detalles

Para crear un perfil de comunidad, este workflow utiliza tres herramientas de clasificación de taxonomía: GOTTCHA2, Kraken2 y Centrifuge. Estas herramientas varían en niveles de especificidad y sensibilidad. Cada herramienta tiene un separado base de datos de referencia. Estas bases de datos (152 GB) están integradas en NMDC EDGE. Los usuarios pueden seleccionar uno, dos o todos tres de las herramientas de clasificación para ejecutar en el workflow.

Versions de Software

- GOTTCHA2 v2.1.6
- Kraken2 v2.0.8
- Centrifuge v1.0.4

Productos

Los resultados completos están disponibles para cada herramienta en tres niveles taxonómicos (especie, género y familia) en el archivo .tsv más grande para cada herramienta, y los mejores resultados en el archivo .tsv más pequeño. También se genera un gráfico Krona interactivo para los resultados de cada herramienta.

Archivos de Salida Primarios

Descripción

Resultados de creación de perfiles para cada herramienta	Resultados tabulares del perfil para cada herramienta. (.tsv)
Gráficos Kronas para cada herramienta	Archivo gráfico interactivo (.html)

Ejecutando el 'Read-based Taxonomy Classification Workflow' en NMDC EDGE

Elige un workflow

1. Desde 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 'Reads-based Taxonomy Classification' desde el menú desplegable debajo de Workflow.

Metagenomics | Run Single Workflow

Run a Single Workflow

Project/Run Name (required, at 3 but less than 30 characters)

Description (optional)

Workflow

Select a Workflow...

- ReadsQC
- Read-based Taxonomy Classification**
- Metagenome Assembly
- Metagenome Annotation
- Metagenome MAGs

Input

Este workflow requiere datos 'paired-end' de Illumina en forma FASTQ como input; el archivo puede ser entrelazado o comprimido. **Formatos de archivo aceptables:** .fastq, .fq, .fastq.gz, .fq.gz

5. Selecciona su herramienta de análisis usando el menú desplegable. Los usuarios pueden seleccionar una, dos o las tres herramientas de clasificación para ejecutarlas en el workflow.


- La configuración predeterminada es que los datos ‘crudos’ estén en formato entrelazado (‘reads’ emparejadas entrelazado en un archivo). Si los datos ‘crudos’ son ‘reads’ emparejados en archivos separados (hacia adelante y hacia atrás), haga clic en ‘No’.
- Se pueden agregar archivos de datos adicionales (del mismo tipo, entrelazado o separados) con el botón abajo.
- 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.
- Haga clic en ‘Submit’ (Someter).

Input

Select Analysis Tool(s) ⚠ Select tool(s)... ▾

Input Raw Reads ⓘ
Is single-end? **Yes** No

Input single-end fastq **Add single-end fastq**

single-end 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
Read-based Taxonomy Classification	On	Done	00:49:28	2021-10-14 15:03:38	2021-10-14 15:53:06
-- gottcha2					
-- kraken2					
-- centrifuge					
▶ "Project Configuration" : { ... }					

La sección 'Read-based Taxonomy Classification Results' (Resultados de la Clasificación de Taxonomía Basada en 'Reads') tiene una sección de resumen en la parte superior y resultados para cada herramienta en tres niveles de taxonomía en la sección 'Taxonomy Top 10' (10 Taxonomías Principales). La sección 'Details'

(Detalles) ha clasificado los resultados de ‘reads’ y los resultados de abundancia relativa para cada herramienta en tres niveles de taxonomía.

Read-based Taxonomy Classification Result

Summary

Tool	Classified Reads	Species Reads	Species
gottcha2	89,222,937	89,222,937	9
centrifuge	14,874,315	14,485,925	5,127
kraken2	30,854,417	29,421,033	2,791

Taxonomy Top 10

Species

Genus

Family

Tool	Level	Top1	Top2	Top3	Top4	Top5	Top6	Top7	Top8	Top9	Top10
gottcha2	species	Pseudomonas aeruginosa	Salmonella enterica	Listeria monocytogenes	Enterococcus faecalis	Lactobacillus fermentum	Bacillus subtilis	Escherichia coli	Staphylococcus aureus	Listeria phage A500	
centrifuge	species	Pseudomonas aeruginosa	Enterococcus faecalis	Bacillus subtilis	Bacillus intestinalis	Listeria monocytogenes	Lactobacillus fermentum	Pseudomonas fluorescens	Pseudomonas sp. AK6U	Salmonella enterica	Enterococcus faecalis
kraken2	species	Pseudomonas aeruginosa	Salmonella enterica	Bacillus subtilis	Listeria monocytogenes	Enterococcus faecalis	Lactobacillus fermentum	Escherichia coli	Staphylococcus aureus	Homo sapiens	Bacillus subtilis

Detail

centrifuge

gottcha2

kraken2

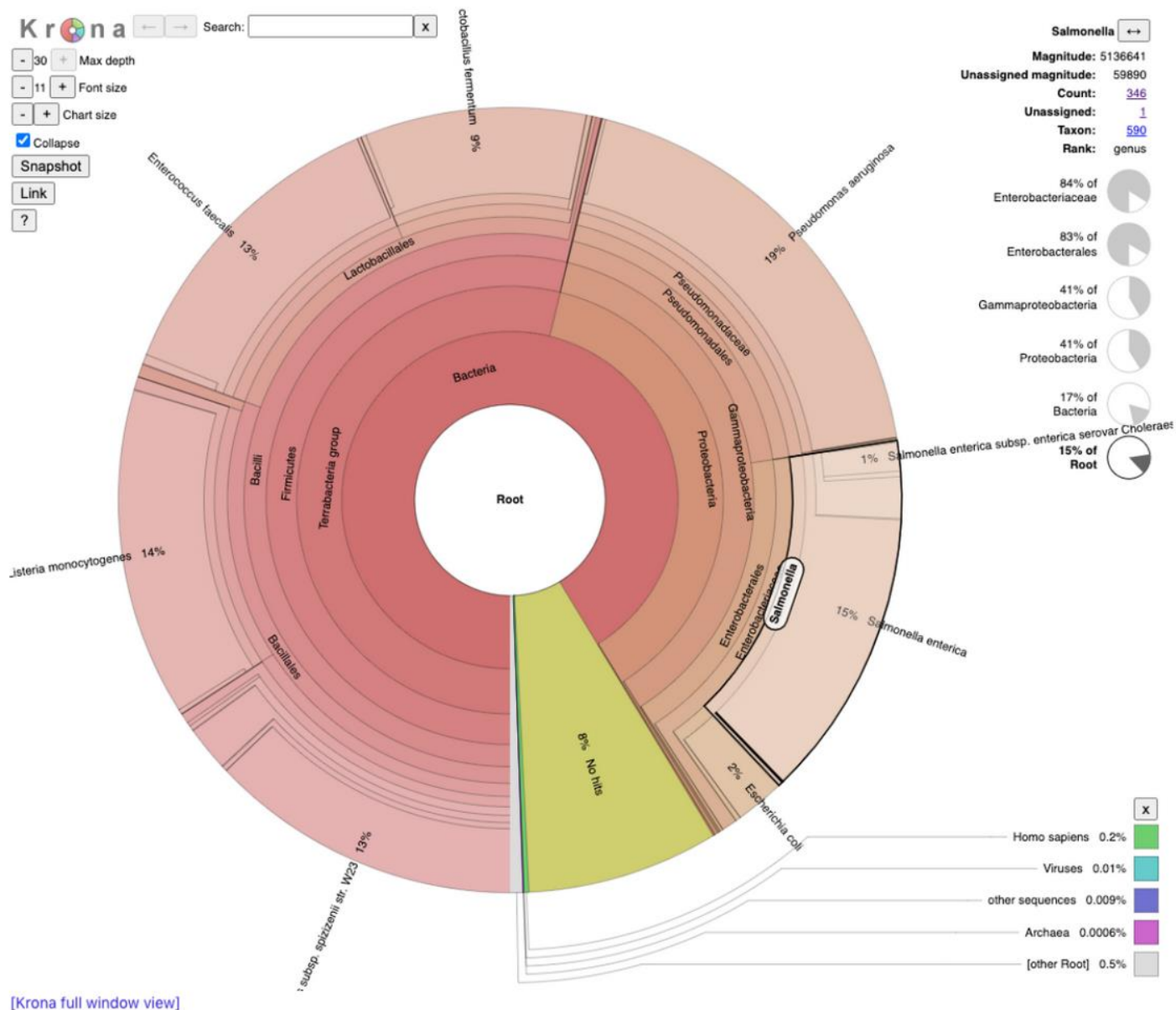
Species

Genus

Family

Level	Taxonomy	Reads	Abundance
species	Pseudomonas aeruginosa	7,026,567	0.001
species	Enterococcus faecalis	5,840,658	0.002
species	Bacillus subtilis	5,224,145	0
species	Bacillus intestinalis	5,080,010	0.002
species	Listeria monocytogenes	4,963,265	0.002
species	Lactobacillus fermentum	4,223,501	0.002
species	Pseudomonas fluorescens	2,735,891	0
species	Pseudomonas sp. AK6U	2,685,240	0
species	Salmonella enterica	2,353,819	0
species	Escherichia coli	776,322	0

La sección ‘Detail’ (Detalle) también proporciona un gráfico Krona interactivo para cada herramienta.



La sección 'Browser/Download Output' (Navegador/Descargar Productos) proporciona archivos de resultados disponibles para descargar. Cada herramienta tiene una carpeta separada para los resultados de esa herramienta. Los resultados tabulares completos se encuentran en el archivo .tsv más grande y los gráficos interactivos Krona (archivos .html) se abren en una ventana separada del navegador.

File	Size	Last Modified
ReadbasedAnalysis		
centrifuge		
Taxonomy_NMDC_test.classification.tsv	3723.93 MB	20 days ago
Taxonomy_NMDC_test.krona.html	4.78 MB	20 days ago
Taxonomy_NMDC_test.report.tsv	553 kB	20 days ago
gottcha2		
Taxonomy_NMDC_test.full.tsv	552 kB	20 days ago
Taxonomy_NMDC_test.krona.html	232 kB	20 days ago
Taxonomy_NMDC_test.tsv	4 kB	20 days ago
kraken2		
Taxonomy_NMDC_test.classification.tsv	2464.25 MB	20 days ago
Taxonomy_NMDC_test.krona.html	2.59 MB	20 days ago
Taxonomy_NMDC_test.report.tsv	412 kB	20 days ago
Taxonomy_NMDC_test.json	2.27 MB	20 days ago