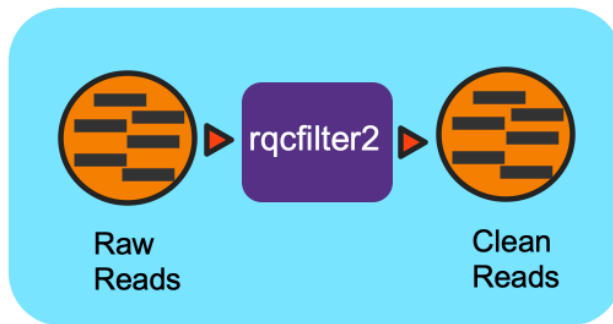


Workflow de Reads QC (v1.0.1)



Descripción General

Este workflow realiza un control de calidad en ‘reads’ ‘crudos’ de Illumina para recortar/filtrar datos de baja calidad y eliminar artefactos, enlazadores, adaptadores, limpieza de ‘reads’ de control de calidad, y contaminaciones de distintos hospederos (i.e., humanos, animales, etc.) y contaminantes de comunes microbianos.

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 ‘Reads QC’ requiere datos ‘paired-end’ de Illumina como un archivo entrelazado o parejas separadas en archivos FASTQ.

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

Detalles

Este workflow utiliza el programa ‘rcqfilter2’ de BBTools para realizar control de calidad de ‘reads’ crudos de Illumina. El workflow realiza cortes a base de calidad, eliminación de artefactos, eliminación de ‘linkers’, recortes de adaptadores y eliminación de picos (usando BBDuk), y realiza la eliminación de humanos/gatos/perros/ratones/microbios (usando BBMap).

Versiones de Software

- rqcfilter2 (BBTools v38.94)
- bbdduk (BBTools v38.94)
- bbmap (BBTools v38.94)

Productos

El resultado principal son los datos ‘limpios’ como un archivo FASTQ entrelazado (.fq.gz). También hay generales estadísticas y estadísticas más detalladas del ‘QC workflow’ en archivos de texto (.txt).

Archivos de Salida Primarios	Descripción
------------------------------	-------------

‘Reads’ de Secuenciación Filtradas	‘Limpios’ datos ‘paired-end’ de forma entrelazados (.fastq.gz)
Estadísticas QC	Resumen Estadístico de ‘Reads QC’

Ejecutando el ‘Reads QC 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 ‘ReadsQC’ desde el menú desplegable debajo de Workflow.

The screenshot shows the 'Run a Single Workflow' page in the NMDC EDGE application. The interface includes a left sidebar with navigation options and a main content area for workflow configuration. Four orange arrows with numbers 1 through 4 point to specific elements:

- Arrow 1:** Points to the 'Run a Single Workflow' option in the 'WORKFLOWS' section of the left sidebar.
- Arrow 2:** Points to the 'Project/Run Name' input field, which has a warning icon and a note: '(required, at 3 but less than 30 characters)'.
- Arrow 3:** Points to the 'Description' input field, which has a note: '(optional)'.
- Arrow 4:** Points to the 'ReadsQC' option in the expanded 'Workflow' dropdown menu.

The 'Workflow' dropdown menu lists the following options: ReadsQC, Read-based Taxonomy Classification, Metagenome Assembly, Metagenome Annotation, and Metagenome MAGs.

Input

Este workflow requiere datos ‘paired-end’ de Illumina en forma FASTQ como input; el archivo puede ser entrelazado y puede ser 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).

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
ReadsQC	On	Done	03:03:03	2021-10-20 20:33:14	2021-10-20 23:36:17

▶ "Project Configuration" : {...}

La sección ‘ReadsQC Result’ (Resultado de ReadsQC) muestra el aporte de datos y proporciona una variedad de métricas incluyendo el número de ‘reads’ y bases antes y después de recortar y filtrar.

ReadsQC Result	
Input	SRR7877884-int
Reads	Status
inputReads	44,943,418
inputBases	6,741,512,700
qtrimmedReads	8,583
qtrimmedBases	8,690
qfilteredReads	200,626
qfilteredBases	29,786,796
ktrimmedReads	6,186,690
ktrimmedBases	354,478,706
kfilteredReads	100,360
kfilteredBases	14,684,762
outputReads	33,510,668
outputBases	4,868,925,674
gcPolymerRatio	0.42

La sección ‘Browser/Download Output’ (Navegador/Descargar Productos) proporciona archivos de resultados disponibles para descargar. Los datos ‘limpios’ estarán en un archivo .fq.gz entrelazado. Las estadísticas generales de control de calidad se encuentran en el archivo filterStats.txt.

Browser/Download Outputs		
File	Size	Last Modified
ReadsQC		
SRR7877884-int		
filterStats2.txt	706 B	5 days ago
filterStats.json	337 B	5 days ago
filterStats.txt	287 B	5 days ago
SRR7877884-int.anqpht.fq.gz	1713.58 MB	5 days ago