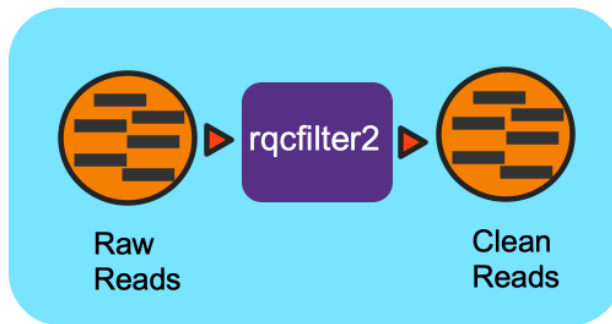


Workflow de contrôle qualité des reads (v1.0.1)



Aperçu

Ce workflow effectue un contrôle qualité sur les reads Illumina pour retirer/filtrer les données de faible qualité et pour supprimer les artefacts, les adaptateurs, les reads artificiels et les reads correspondant à plusieurs hôtes et contaminants microbiens courants.

Exécution du Workflow

Actuellement, ce workflow peut être exécuté via [NMDC EDGE](#) ou sur des ressources de calcul locales (les instructions et conditions d'installation se trouvent [ici](#) et [ici](#).)

Des didacticiels vidéo sur la façon d'exécuter chaque workflow dans NMDC EDGE sont disponibles [ici](#).

Fichiers d'entrée

Le workflow « Metagenome ReadsQC » nécessite des données Illumina appariées sous forme d'un fichier entrelacé FASTQ unique ou sous forme de reads appariés dans deux fichiers FASTQ séparés.

- **Formats de fichiers acceptés:** .fastq, .fq, .fastq.gz, .fq.gz

Instructions détaillées

Ce workflow utilise le programme « rqcfilter2 » de BBTools pour effectuer un contrôle qualité sur les reads brutes Illumina. Le workflow supprime les bases de mauvaise qualité, les reads artefactuels ou correspondant à des adaptateurs, et les reads correspondant aux génomes contaminants humain/chat/chien/souris/microbe (à l'aide de BMap).

Versions des outils

- rqcfilter2 (BBTools v38.94)
- bbduk (BBTools v38.94)
- bmap (BBTools v38.94)

Fichiers de sortie

La sortie principale est le jeu de données nettoyé dans un fichier FASTQ compressé entrelacé (.fq.gz). Il existe également des statistiques générales et des statistiques plus détaillées du workflow QC dans des fichiers texte (.txt).

Fichiers de sortie principaux	Description
Filtered Sequencing Reads	Données nettoyées au format entrelacées (.fastq.gz)
QC statistics	Statistiques récapitulatives du workflow QC

Exécution du workflow de contrôle qualité des reads dans NMDC EDGE

Sélectionner un workflow

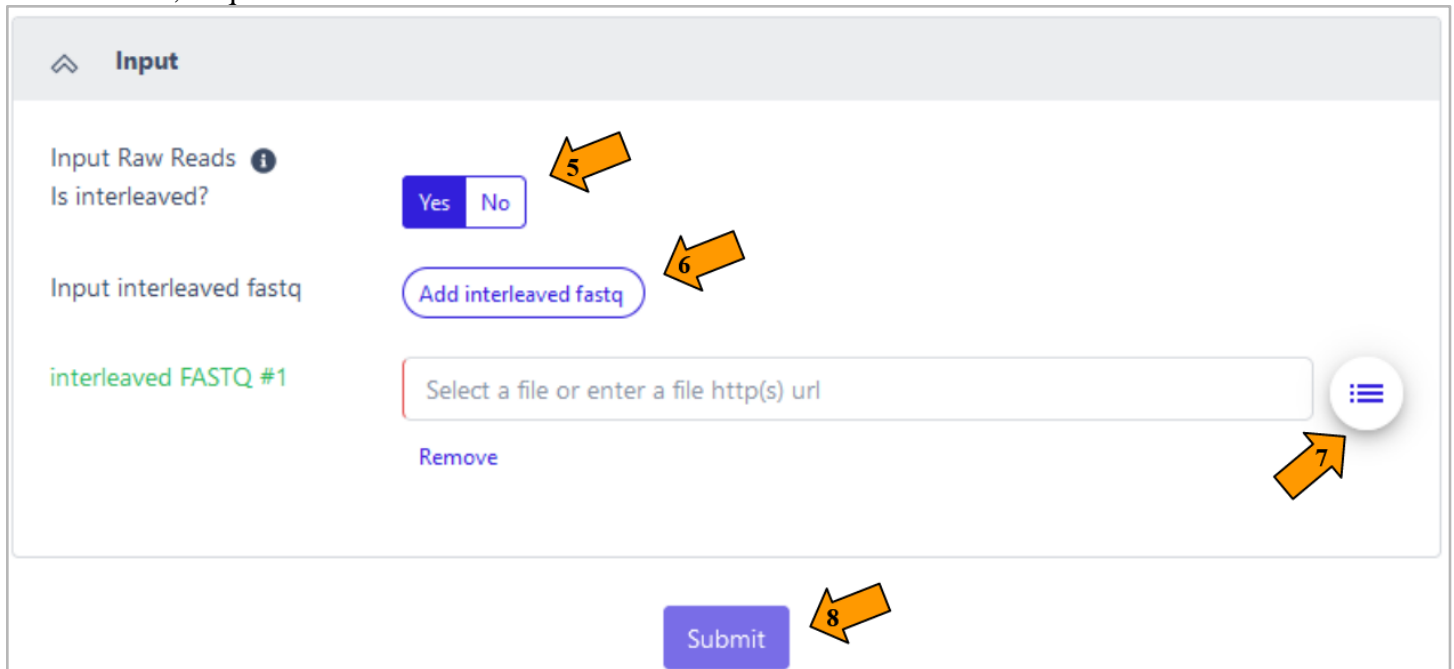
1. Dans la catégorie Metagenomics dans la barre de menu de gauche, sélectionnez 'Run a Single Workflow'.
2. Entrez un nom de projet unique sans espaces (les traits de soulignement sont possibles).
3. Une description est facultative, mais utile.
4. Sélectionnez « ReadsQC » dans le menu déroulant sous Workflow.

The screenshot displays the NMDC EDGE web interface for running a workflow. On the left sidebar, under the 'WORKFLOWS' section, 'Metagenomics' is expanded, and 'Run a Single Workflow' is highlighted with an orange arrow labeled '1'. The main content area is titled 'Metagenomics | Run Single Workflow' and 'Run a Single Workflow'. It contains three input fields: 'Project/Run Name' with a warning icon and a note '(required, at 3 but less than 30 characters)' (arrow 2), 'Description' with a note '(optional)' (arrow 3), and a 'Workflow' dropdown menu (arrow 4) showing a list of workflows with 'ReadsQC' selected at the top.

Fichiers d'entrée

Le workflow ReadsQC nécessite des données Illumina appariées au format FASTQ comme entrée ; le fichier peut être entrelacé et compressé. **Formats de fichiers acceptés** : .fastq, .fq, .fastq.gz, .fq.gz

5. Par défaut, les données d'entrée sont attendues au format entrelacé (reads appariés entrelacés dans un seul fichier). Si les reads appariés sont dans des fichiers séparés (avant et arrière), cliquez sur « No ».
6. Des fichiers de données supplémentaires (du même type – entrelacés ou séparés) peuvent être ajoutés avec le bouton ci-dessous.
7. Cliquez sur le bouton à droite du champ de saisie des données pour sélectionner le fichier de données pour l'analyse. (S'il y a des fichiers séparés, il y aura deux espaces de saisie.) Une boîte de dialogue appelée « Sélectionner un fichier » s'ouvrira pour permettre à l'utilisateur de trouver le(s) fichier(s) souhaité(s) à partir de projets précédemment exécutés, du dossier de données public ou des fichiers téléchargés. par l'utilisateur.
8. Enfin, cliquez sur « Submit ».



Input

Input Raw Reads ⓘ
Is interleaved? Yes No 5

Input interleaved fastq Add interleaved fastq 6

interleaved FASTQ #1 Select a file or enter a file http(s) url 7
Remove

Submit 8

Fichiers de sortie

La section « General » indique quel workflow a été exécuté, ainsi que les informations d'exécution.

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 section « ReadsQC » contient les informations sur les données d'entrée et un ensemble de statistiques tels que le nombre de reads et de paires de bases avant et après nettoyage et filtration.

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 section « Browser/Download » fournit des fichiers de sortie disponibles au téléchargement. Les données nettoyées seront disponibles dans un fichier entrelacé .fq.gz. Les statistiques de qualité sont disponibles dans le fichier 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