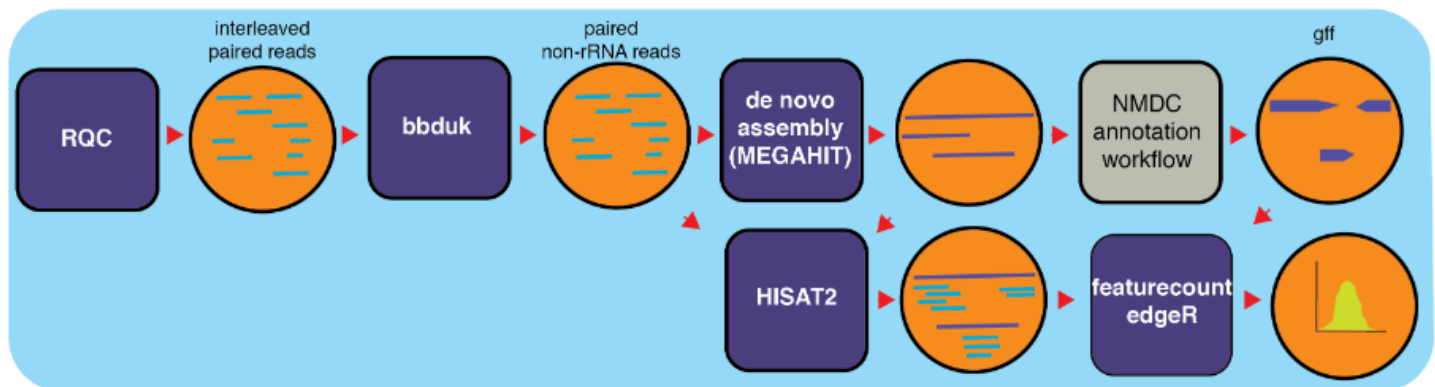


Workflow d'analyse Métatranscriptomique (v0.0.2)



Aperçu

Le workflow métatranscriptomique (metaT) filtre les données brutes de métatranscriptome pour en vérifier la qualité, supprime les séquences d'ARNr (ARN ribosomal), puis assemble et annote les séquences transcrites. Les reads sont mappées sur les gènes prédits et les RPKM (nombre de reads par kilobase de transcrits par million de reads mappées) sont calculés pour chaque gène présent dans le fichier d'annotation fonctionnelle.

Exécution du Workflow

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

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

Fichiers d'entrée

Métatranscriptomique nécessite des séquences Illumina appariées sous forme de fichier entrelacé ou sous forme de deux fichiers FASTQ distincts.

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

Instructions détaillées

MetaT est un workflow conçu pour analyser les métatranscriptomes, et ce workflow s'appuie sur d'autres workflows NMDC pour traiter les données de séquençage d'entrée. Le workflow métatranscriptomique prend en entrée les données brutes de séquençage d'ARN et filtre la qualité des séquences à l'aide du workflow ReadsQC. Ensuite, le workflow MetaT filtre les séquences d'ARN ribosomal (à l'aide de la base de données SILVA rRNA) et sépare les fichiers de reads entrelacés en paires de fichiers distincts à l'aide de bbduk (BBTools). Après les étapes de filtration, les séquences sont assemblées en transcrits à l'aide de MEGAHIT et les transcrits sont annotés à l'aide du workflow [Metagenome Annotation NMDC](#) qui produit un fichier d'annotation fonctionnelle GFF. Les gènes sont comptés avec la fonction [featureCounts de Subread](#) qui génère deux RPKM pour chaque gène dans un fichier GFF, pour les reads sens et antisens.

Versions des outils

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

Fichiers de sortie

Le tableau ci-dessous répertorie les principaux fichiers de sortie. Les principaux résultats sont les transcrits assemblés et le fichier de gènes annotés. Plusieurs fichiers d'annotations sont également disponibles au téléchargement.

Fichiers de sortie principaux	Description
INPUT_NAME.contigs.fa	Transcrits assemblés
rpkm_sorted_features.tsv	Table de gènes triés par RPKM

Exécution du workflow d'analyse Métatranscriptomique « Metatranscriptomics » dans NMDC EDGE

Sélectionner un workflow

1. Dans la catégorie Metatranscriptomics 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 « Metatranscriptome » dans le menu déroulant sous Workflow.

Fichiers d'entrée

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

1. Par défaut, les données d'entrée sont attendues au format entrelacé (séquences appariées entrelacées dans un seul fichier). Si les reads appariés sont dans des fichiers séparés, cliquez sur « No ».
2. Des fichiers de données supplémentaires (du même type – entrelacés ou séparés) peuvent être ajoutés avec le bouton ci-dessous.
3. 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.
4. Enfin, cliquez sur « Submit ».

Fichiers de sortie

La section « General » indique quel workflow et quels outils ont été exécutés, ainsi que les informations d'exécution.

La section « Metatranscriptome Result » contient une table des 100 gènes avec le RPKM le plus élevé à partir du fichier de resultat global trie par RPKM. Cliquer sur l'en-tête de chaque colonne triera ces données en fonction de cette colonne. Cette section comprend également un bouton permettant de télécharger rapidement un fichier TSV de tous les gènes détectés dans l'ensemble de données d'entrée pour une analyse plus approfondie.

La section Browser/Download Output liste tous les fichiers de sortie disponibles au téléchargement. Les contigs obtenus sont disponibles dans le dossier « assembly » et le fichier tsv de tous les gènes détectés triés par RPKM est disponible dans le dossier « métat_output ».