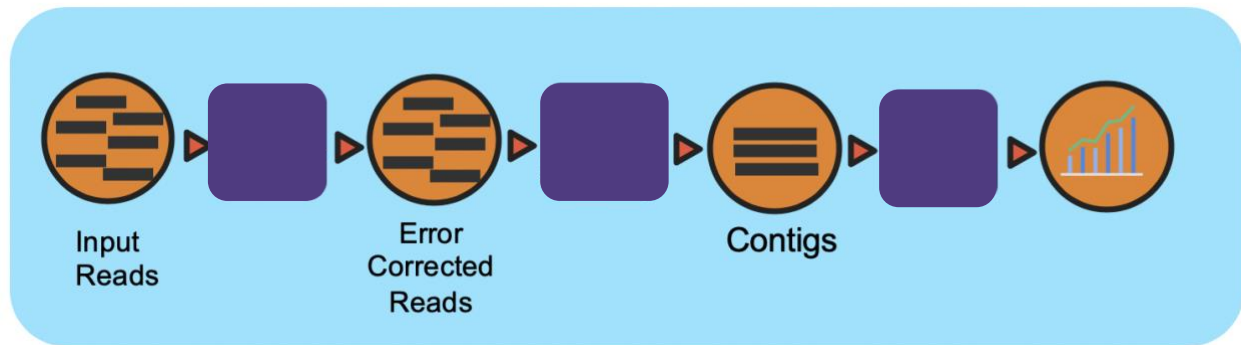


## Workflow Assemblage de Métagénome (v1.0.1)



### Aperçu

Ce workflow exécute une correction des erreurs des reads Illumina appariées, un assemblage de ces reads, et une validation de l'assemblage.

### 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

L'assemblage de métagénome nécessite des reads Illumina appariés sous forme de fichier entrelacé ou sous forme de deux fichiers FASTQ distincts. L'entrée recommandée est la sortie du workflow ReadsQC NMDC.

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

### Instructions détaillées

Ce workflow utilise les séquences Illumina appariées et effectue une correction d'erreur à l'aide de bbcms (BBTools). Ensuite, les séquences corrigées sont assemblées à l'aide de metaSPAdes. Après l'assemblage, les reads sont mappées sur les contigs par bbmap (BBTools) pour estimer la couverture de chaque contig.

### Versions des outils

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

### Fichiers de sortie

La sortie principale est le fichier des contigs assemblés (assembly\_contigs.fna).

Fichiers de sortie principaux	Description
Assembly Contigs	Fichier final des contigs assemblés (assembly.contigs.fna)
Assembly Scaffolds	Fichier final des scaffolds assemblés (assembly_scaffolds.fna)

Assembly AGP	Un fichier au format AGP qui décrit l'assemblage
Assembly Coverage BAM	Fichier au format bam trié de séquences mappées à l'assemblage final
Assembly Coverage Stats	Statistiques de couverture pour les contigs assemblés

## Exécution du workflow d'assemblage de métagénome 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 « Metagenome Assembly » dans le menu déroulant sous 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

### Fichiers d'entrée

Ce workflow accepte des données de séquençage Illumina au format FASTQ en entrée; le fichier peut être compressé. Ces fichiers d'entrée peuvent être les fichiers de sortie du workflow ReadsQC, qui sont par défaut au format recommandé. **Formats de fichier acceptés:** .fastq, .fq, .fastq.gz, .fq.gz

5. 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 ».

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 « Select a File » 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

Input interleaved fastq Add interleaved fastq

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

Submit

## Fichiers de sortie

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

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 section « Metagenome Assembly Result » contient toutes les statistiques de l'assemblage.

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 section « Browser/Download » fournit des fichiers de sortie disponibles au téléchargement. Le fichier de sortie principal est assembly\_contigs.fna qui peut également être l'entrée du workflow d'annotation de Métagénome. Le fichier pairedMapped\_sorted.bam ainsi que le fichier de contigs assemblés peuvent être l'entrée du workflow de génération de MAGs.

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