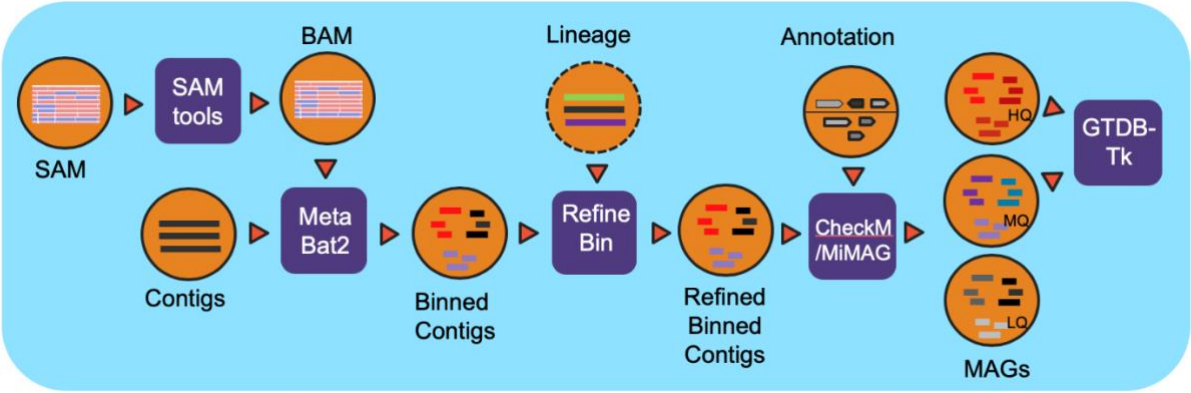
Workflow Assemblage de Génomes à partir de Métagenomes (MAGs) (v1.0.2)



Aperçu

Ce workflow groupe les contigs au sein de bins et les bins sont elle-même affinées à l'aide du fichier d'annotation fonctionnelle. Les bins sont évaluées quant à leur exhaustivité et leur contamination. La qualité des bins est déterminée, et une affiliation taxonomique est attribuée à chaque bin de qualité élevée ou moyenne.

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 worfklow dans NMDC EDGE sont disponibles ici.

Fichiers d'entrée

Ce workflow nécessite des contigs assemblés dans un fichier FASTA, le fichier d'alignement des reads sur les contigs assemblés (SAM ou BAM), une annotation fonctionnelle de l'assemblage dans un fichier GFF.

- **Formats de fichier acceptés :** contigs assemblés (.fasta, .fa, ou .fna); alignement des reads sur l'assemblage (.sam.gz ou .bam); annotation fonctionnelle (.gff)

Instructions détaillées

Le workflow est basé sur le pipeline de binning des métagénomes d'IMG et a été modifié spécifiquement pour le projet NMDC. Pour tous les métagénomes traités, il groupe les contigs dans des bins à l'aide de MetaBat2. Ensuite, les bins sont affinées à l'aide du fichier d'annotation fonctionnelle (GFF) du workflow d'annotation de métagénomes et, si disponibles, des informations d'affiliation taxonomique des contigs. L'exhaustivité et la contamination des bins sont évaluées par CheckM et les bins se voient attribuer un niveau de qualité (Haute Qualité (HQ), Qualité Moyenne (MQ), Faible Qualité (LQ)) sur la base des normes MiMAG. En dernière étape, GTDB-Tk est utilisé pour attribuer une affiliation taxonomique aux bins HQ et MQ. La base de données GTDB-Tk requise est intégrée à NMDC EDGE.

Versions des outils

- Biopython v1.74
- Sqlite
- Pymysql
- requests
- samtools > v1.9 (License: MIT License)
- Metabat2 v2.15
- CheckM v1.1.2
- GTDB-TK v1.2.0

LA-UR-21-21661

- FastANI v1.3
- FastTree v2.1.10

Fichiers de sortie

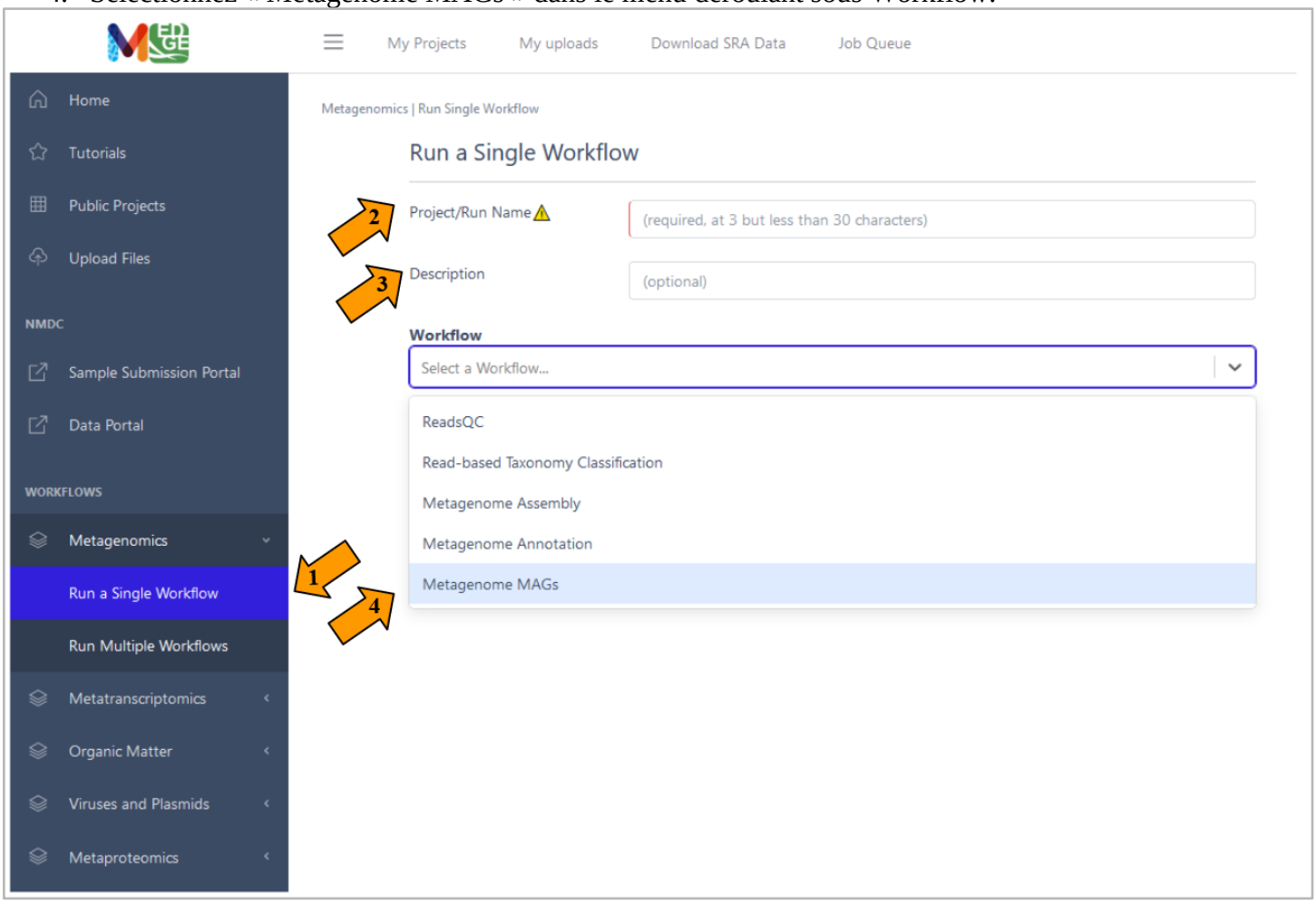
Le fichier de sortie principal est l'archive zip des bins de haute qualité (HQ) et de qualité moyenne (MQ). Les contigs non inclus dans une bin ou inclus dans une bin de faible qualité sont aussi disponibles.

Fichiers de sortie principaux	Description
hqmq-metabat-bins.zip	Bins de haute qualité ou de qualité moyenne.

Exécution du workflow « Assemblage de Génomes à partir de Métagenomes » (MAG) dans NMDC EDGE

Sélectionner un workflow

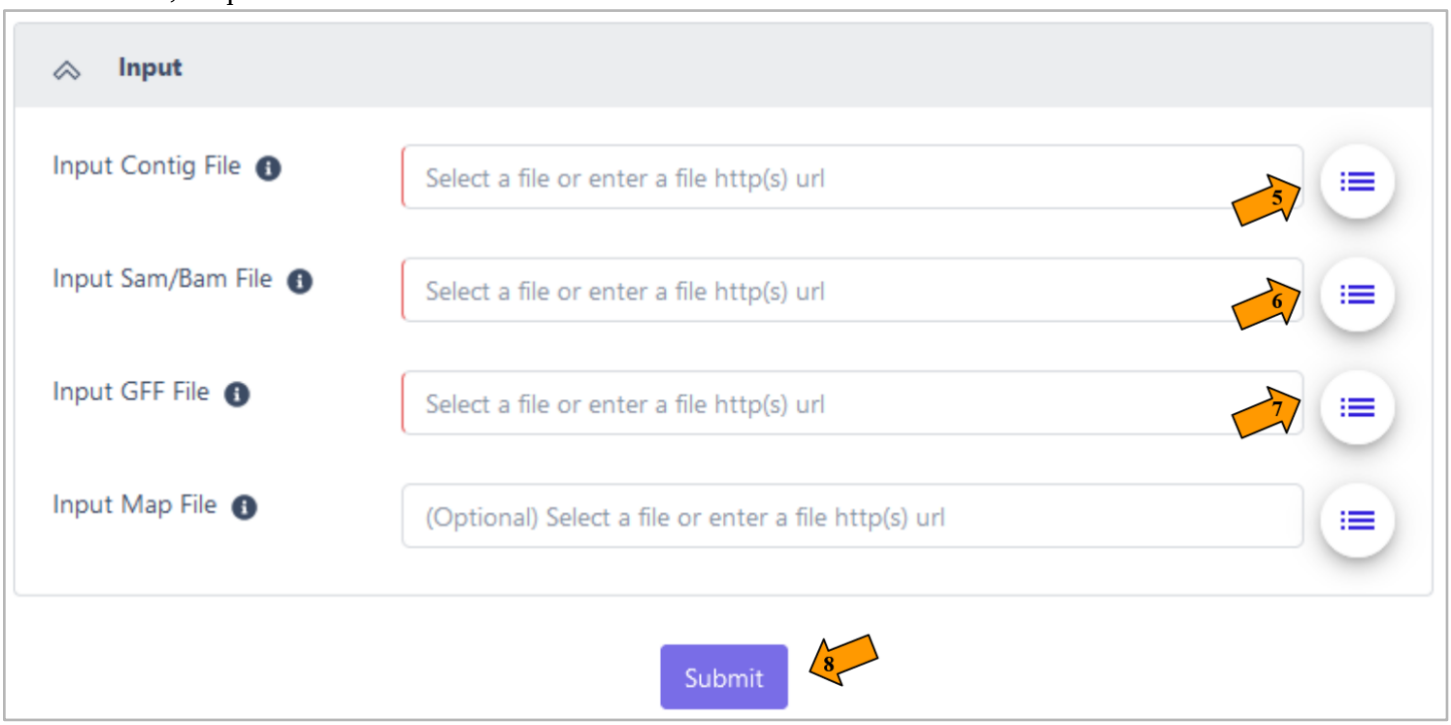
1. Dans la catégorie « Metagenomics » dans la barre de menu de gauche, sélectionnez « Exécuter un workflow ».
2. Entrez un nom de projet ***unique*** sans espaces (les traits de soulignement sont possibles).
3. Une description est facultative, mais utile.
4. Sélectionnez « Metagenome MAGs » dans le menu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



Fichiers d'entrée

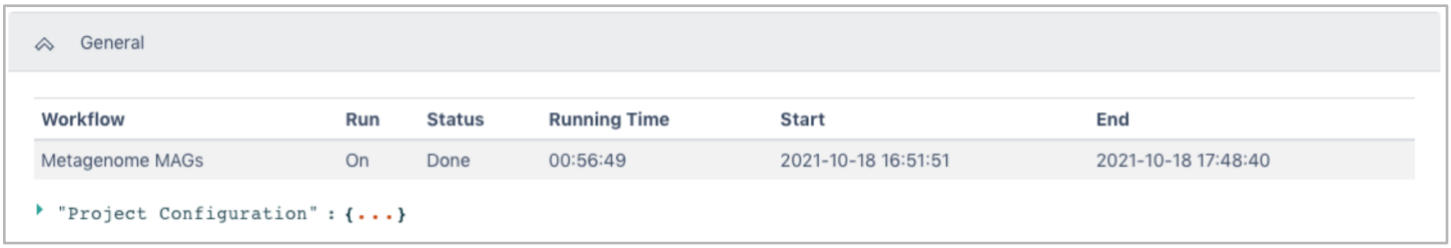
Le workflow Metagenome MAG nécessite des contigs assemblés, le fichier d'alignement des reads sur les contigs assemblés et un fichier d'annotation fonctionnelle. L'entrée recommandée correspond aux fichiers de sortie des workflows d'assemblage et d'annotation NMDC. **Formats de fichiers acceptables:** contigs assemblés (.fasta, .fa ou .fna) ; alignement des reads sur l'assemblage (.sam.gz ou .bam) ; annotation fonctionnelle (.gff).

5. Cliquez sur le bouton à droite de l'espace vide pour l'entrée « Input Contig File ». Une boîte de dialogue appelée « Select a File » s'ouvrira pour permettre à l'utilisateur de trouver le fichier souhaité à partir d'un workflow d'assemblage précédemment exécuté, du dossier de données publiques, ou d'un fichier téléchargé par l'utilisateur
6. Cliquez sur le bouton à droite de l'espace vide pour le fichier d'entrée Sam/Bam. Une boîte appelée « Select a File » s'ouvrira pour permettre à l'utilisateur de trouver le fichier d'alignement des reads à partir d'un workflow d'assemblage précédemment exécuté, du dossier de données publiques, ou d'un fichier téléchargé par l'utilisateur.
7. Cliquez sur le bouton à droite de l'espace vide pour le fichier GFF d'entrée. Une boîte appelée « Select a File » s'ouvrira pour permettre à l'utilisateur de trouver le(s) fichier(s) souhaité(s) à partir d'un workflow d'annotation précédemment exécuté, du dossier de données publiques, ou d'un fichier téléchargé par l'utilisateur.
8. 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 « Metagenome MAGs Result » fournit une section « Summary » avec des informations sur les contigs regroupés au sein de bins et les contigs non regroupés. La section « MAG » fournit des informations telles que l'exhaustivité, le degré de contamination et le nombre de gènes présents pour toutes les bins déterminés comme étant de qualité élevée ou moyenne.

Metagenome MAGs Result

Summary

Name	Status
input_contig_num	25,726
too_short_contig_num	15,158
lowDepth_contig_num	0
unbinned_contig_num	9,334
binned_contig_num	1,234

MAGs

bin_name	number_of_contig	completeness	contamination	gene_count	bin_quality	num_16s	num_5s	num_23s	num_tRNA	gtdbtk_domain	gtdbtk_phy
bins.1	63	99.48	0.16	4,826	HQ	1	4	1	76	Bacteria	Proteobacte
bins.4	35	99.68	0.61	6,653	MQ	0	0	0	63	Bacteria	Proteobacte
bins.6	17	99.45	5.19	3,575	MQ	1	3	0	71	Bacteria	Firmicutes
bins.3	29	78.23	0	1,248	MQ	0	1	0	28	Bacteria	Firmicutes
bins.5	7	68.97	0	1,819	MQ	0	1	0	24	Bacteria	Firmicutes
bins.7	18	62.07	0	3,161	MQ	2	3	2	54	Bacteria	Firmicutes

La section « Browser/Download » fournit des fichiers de sortie disponibles au téléchargement. Le fichier de sortie principal est le fichier compressé avec toutes les bins déterminées comme étant de qualité élevée ou moyenne (hqmq-metabat-bins.zip).

Browser/Download Outputs

File	Size	Last Modified
MetagenomeMAGs		
activity.json	7 kB	16 days ago
bins.lowDepth.fa	0 B	16 days ago
bins.tooShort.fa	8.68 MB	16 days ago
bins.unbinned.fa	16.82 MB	16 days ago
checkm_qa.out	3 kB	16 days ago
data_objects.json	2 kB	16 days ago
gtdbtk.bac120.summary.tsv	11 kB	16 days ago
hqmq-metabat-bins.zip	6.04 MB	16 days ago
MAGs_stats.json	7 kB	15 days ago
metabat-bins.zip	1.81 MB	16 days ago