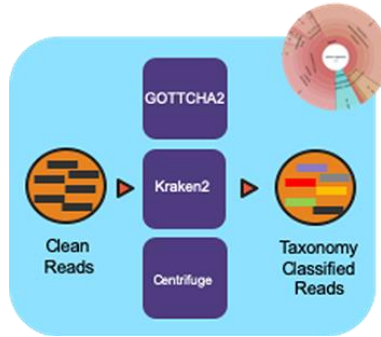


Workflow de classification taxonomique basée sur les reads (v1.0.1)



Aperçu

Ce workflow prend en entrée les fichiers de séquençage Illumina (appariés ou non) et profile les reads à l'aide de plusieurs outils de classification taxonomique.

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 de classification taxonomique basée sur les reads nécessite des données Illumina et peut accepter les données sous forme de fichier entrelacé ou de paires distinctes de fichiers FASTQ. Les données entrelacées seront traitées comme un seul read. Il est fortement recommandé d'utiliser des données filtrées obtenues via le workflow NMDC ReadsQC.

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

Instructions détaillées

Pour créer un profil taxonomique de communauté, ce workflow utilise trois outils de classification taxonomique : GOTTCHA2, Kraken2 et Centrifuge. Ces outils varient en terme de spécificité et de sensibilité. Chaque outil dispose d'une base de données de référence distincte. Ces bases de données (152 Go) sont intégrées à NMDC EDGE. Les utilisateurs peuvent sélectionner un, deux ou les trois outils de classification à exécuter dans le workflow.

Versions des outils

- GOTTCHA2 v2.1.6
- Kraken2 v2.0.8
- Centrifuge v1.0.4

Fichiers de sortie

Les résultats complets sont disponibles pour chaque outil à trois niveaux taxonomiques (Espèce, Genre et Famille) dans les fichiers .tsv complets pour chaque outil et les meilleurs résultats dans le fichier .tsv réduit. Un graphique Krona interactif est également généré pour les résultats de chaque outil.

Fichiers de sortie principaux

Description

| | |
|-----------------------------------|--|
| Résultats pour chaque outil | Table de résultat pour chaque outil de classification (.tsv) |
| Graphique Krona pour chaque outil | Fichier contenant la visualisation graphique interactive (.html) |

Exécution du workflow « Read-based Taxonomy Classification » 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 « Read-based Taxonomy Classification » dans le menu déroulant sous Workflow.

The screenshot shows the NMDC EDGE web interface. On the left is a dark sidebar menu with categories like Home, Tutorials, Public Projects, Upload Files, NMDC, Sample Submission Portal, Data Portal, WORKFLOWS, and Metagenomics. Under Metagenomics, 'Run a Single Workflow' is highlighted. On the right, the 'Run a Single Workflow' form is displayed. It includes a 'Project/Run Name' field (required, 3-30 characters), a 'Description' field (optional), and a 'Workflow' dropdown menu. The dropdown menu is open, showing options: ReadsQC, Read-based Taxonomy Classification (highlighted), Metagenome Assembly, Metagenome Annotation, and Metagenome MAGs. Four orange arrows with numbers 1 through 4 point to the corresponding steps in the interface.

Fichiers d'entrée

Ce workflow accepte des données de séquençage Illumina au format FASTQ en entrée; le fichier peut être entrelacé et peut être compressé. Ces fichiers d'entrée peuvent être les fichiers de sortie du workflow ReadsQC (fichiers d'entrée recommandés). **Formats de fichier acceptés:** .fastq, .fq, .fastq.gz, .fq.gz

- Sélectionnez l'outil désiré en utilisant le menu déroulant. Les utilisateurs peuvent sélectionner un, deux ou les trois outils de classification, qui seront utilisés dans le workflow.
- 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 ».
- Des fichiers de données supplémentaires (du même type – entrelacés ou séparés) peuvent être ajoutés avec le bouton ci-dessous.
- 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.
- Enfin, cliquez sur « Submit ».

The screenshot shows the 'Input' configuration panel. It includes a dropdown menu for 'Select Analysis Tool(s)' (arrow 5), a 'Yes/No' toggle for 'Input Raw Reads' (arrow 6), an 'Add single-end fastq' button (arrow 7), a file selection input field with a menu icon (arrow 8), and a 'Submit' button (arrow 9).

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 |
| Read-based Taxonomy Classification | On | Done | 00:49:28 | 2021-10-14 15:03:38 | 2021-10-14 15:53:06 |
| -- gotcha2 | | | | | |
| -- kraken2 | | | | | |
| -- centrifuge | | | | | |
| "Project Configuration" : { ... } | | | | | |

La section « Read-based Taxonomy Classification Results » comporte une section récapitulative en haut et les résultats détaillés pour chaque outil à trois niveaux de taxonomie dans la section « Taxonomy Top 10 ». La section « Details » contient les résultats des reads classés et les résultats d'abondance relative par taxon pour chaque outil à trois niveaux de taxonomie.

Read-based Taxonomy Classification Result

Summary

| Tool | Classified Reads | Species Reads | Species |
|------------|------------------|---------------|---------|
| gottcha2 | 89,222,937 | 89,222,937 | 9 |
| centrifuge | 14,874,315 | 14,485,925 | 5,127 |
| kraken2 | 30,854,417 | 29,421,033 | 2,791 |

Taxonomy Top 10

Species

Genus

Family

| Tool | Level | Top1 | Top2 | Top3 | Top4 | Top5 | Top6 | Top7 | Top8 | Top9 | Top10 |
|------------|---------|------------------------|-----------------------|------------------------|------------------------|-------------------------|-------------------------|-------------------------|-----------------------|---------------------|-------|
| gottcha2 | species | Pseudomonas aeruginosa | Salmonella enterica | Listeria monocytogenes | Enterococcus faecalis | Lactobacillus fermentum | Bacillus subtilis | Escherichia coli | Staphylococcus aureus | Listeria phage A500 | |
| centrifuge | species | Pseudomonas aeruginosa | Enterococcus faecalis | Bacillus subtilis | Bacillus intestinalis | Listeria monocytogenes | Lactobacillus fermentum | Pseudomonas fluorescens | Pseudomonas sp. AK6U | Salmonella enterica | |
| kraken2 | species | Pseudomonas aeruginosa | Salmonella enterica | Bacillus subtilis | Listeria monocytogenes | Enterococcus faecalis | Lactobacillus fermentum | Escherichia coli | Staphylococcus aureus | Homo sapiens | |

Detail

centrifuge

gottcha2

kraken2

Species

Genus

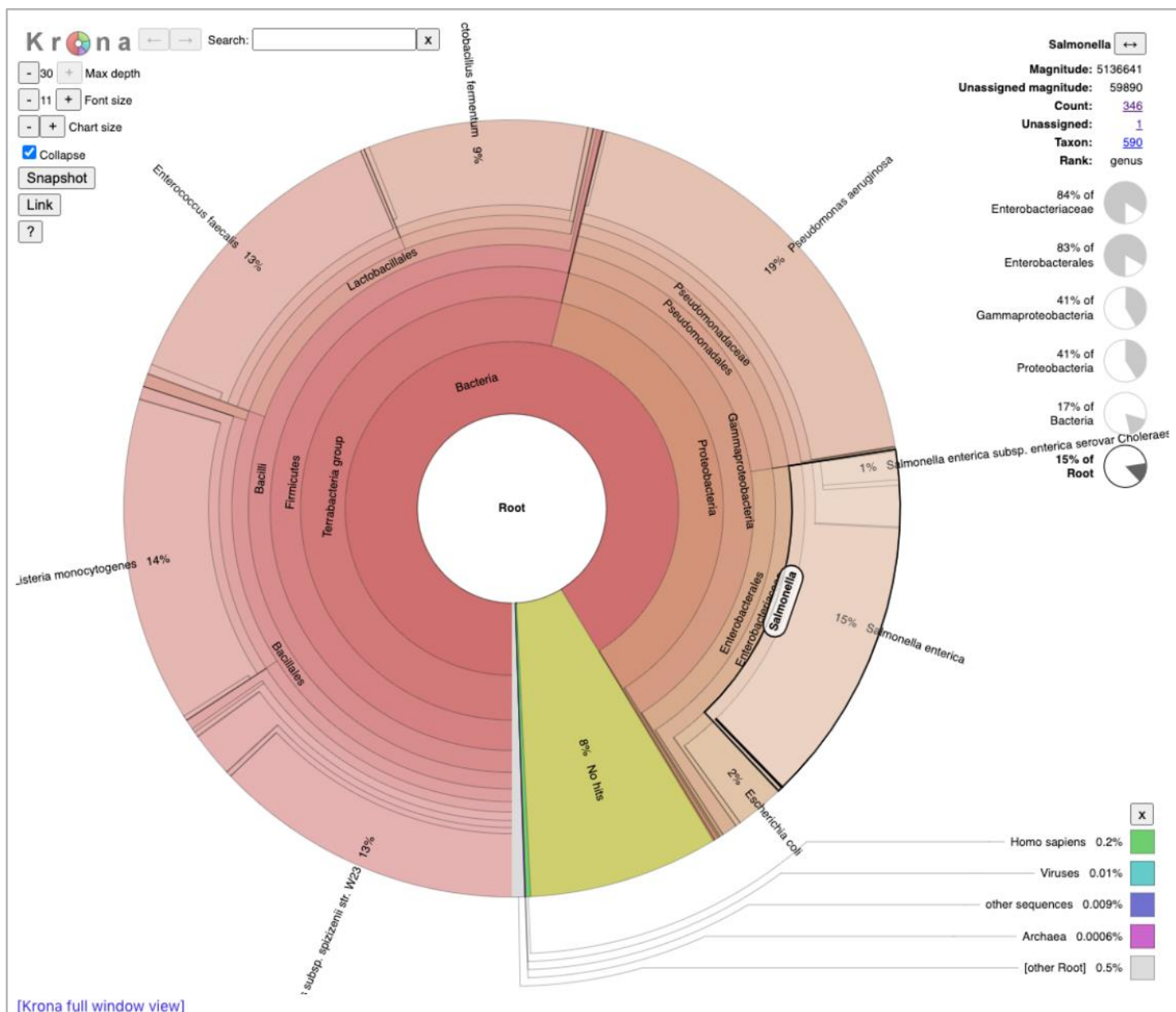
Family

| Level | Taxonomy | Reads | Abundance |
|---------|-------------------------|-----------|-----------|
| species | Pseudomonas aeruginosa | 7,026,567 | 0.001 |
| species | Enterococcus faecalis | 5,840,658 | 0.002 |
| species | Bacillus subtilis | 5,224,145 | 0 |
| species | Bacillus intestinalis | 5,080,010 | 0.002 |
| species | Listeria monocytogenes | 4,963,265 | 0.002 |
| species | Lactobacillus fermentum | 4,223,501 | 0.002 |
| species | Pseudomonas fluorescens | 2,735,891 | 0 |
| species | Pseudomonas sp. AK6U | 2,685,240 | 0 |
| species | Salmonella enterica | 2,353,819 | 0 |
| species | Escherichia coli | 776,322 | 0 |

La section « Detail » propose aussi un graphique interactif Krona pour chaque outil.

LA-UR-21-21661

4



La section « Browser/Download Output » fournit les fichiers de sortie disponibles au téléchargement. Les fichiers pour chaque outil sont organisés dans des dossiers différents. Les tableaux de résultat complets se trouvent dans le plus grand fichier .tsv et les graphiques interactifs Krona (fichiers .html) s'ouvrent dans une fenêtre de navigateur séparée.

| File | Size | Last Modified |
|---------------------------------------|------------|---------------|
| ReadbasedAnalysis | | |
| centrifuge | | |
| Taxonomy_NMDC_test.classification.tsv | 3723.93 MB | 20 days ago |
| Taxonomy_NMDC_test.krona.html | 4.78 MB | 20 days ago |
| Taxonomy_NMDC_test.report.tsv | 553 kB | 20 days ago |
| gottcha2 | | |
| Taxonomy_NMDC_test.full.tsv | 552 kB | 20 days ago |
| Taxonomy_NMDC_test.krona.html | 232 kB | 20 days ago |
| Taxonomy_NMDC_test.tsv | 4 kB | 20 days ago |
| kraken2 | | |
| Taxonomy_NMDC_test.classification.tsv | 2464.25 MB | 20 days ago |
| Taxonomy_NMDC_test.krona.html | 2.59 MB | 20 days ago |
| Taxonomy_NMDC_test.report.tsv | 412 kB | 20 days ago |
| Taxonomy_NMDC_test.json | 2.27 MB | 20 days ago |