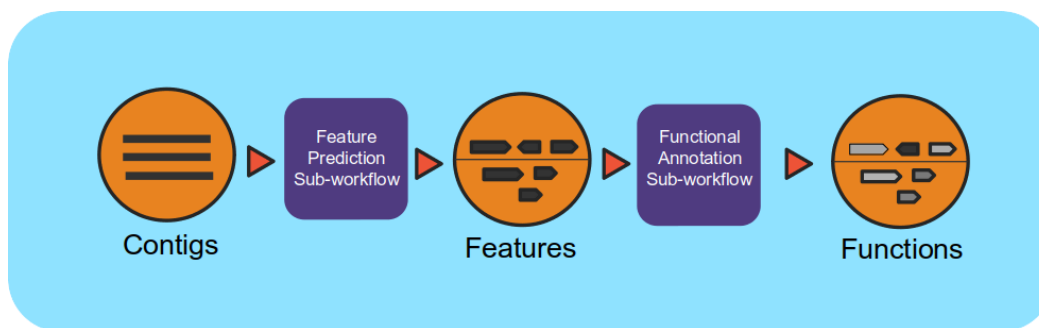


Workflow de Anotación de Metagenoma (v1.0.0)



Descripción General

Este workflow toma metagenomas ensamblados y genera anotaciones estructurales y funcionales.

Ejecutando el Workflow

Actualmente, este workflow se puede ejecutar en [NMDC EDGE](#) o desde la línea de comandos. (Las instrucciones y requisitos de CLI se encuentran [aquí](#)).

[Aquí](#) se encuentran videos tutoriales sobre cómo ejecutar cada workflow en NMDC EDGE.

Input

La anotación de metagenoma requiere ‘contigs’ ensamblados en un archivo FASTA. Es recomendado utilizar la salida del workflow del ‘Metagenome Assembly’ (Ensamblaje Metagenómica).

- **Formatos de archivo aceptables:** .fasta, .fa, .fna, .fasta.gz, .fa.gz, .fna.gz

Detalles

El workflow utiliza una serie de herramientas y bases de datos de código abierto para generar anotaciones estructurales y funcionales. El ensamblaje de entrada se divide primero en divisiones de 10 MB para procesarse en paralelo. Dependiendo de la configuración del motor del workflow, la división se puede procesar en paralelo. Cada división primero se anota estructuralmente y luego esos resultados se utilizan para la anotación funcional. La anotación estructural utiliza tRNAscan-se, RFAM, CRT, Prodigal y GeneMarkS. Estos resultados se unían para crear una anotación estructural de consenso. El GFF resultante es la entrada para la anotación funcional que utiliza múltiples bases de datos de familias de proteínas (SMART, COG, TIGRFAM, SUPERFAMILY, Pfam y Cath-FunFam) junto con modelos HMM personalizados. Las predicciones funcionales se crean utilizando Last y HMM. Estas anotaciones también se unían en un archivo GFF de consenso. Finalmente, las respectivas anotaciones divididas se unían para generar un único archivo de anotación estructural y un único archivo de anotación funcional. Además, se generan varios archivos de resumen en formato TSV.

Versiones de Software

- Conda
- tRNAscan-SE >= 2.0
- Infernal 1.1.2
- CRT-CLI 1.8
- Prodigal 2.6.3
- GeneMarkS-2 >= 1.07
- Last >= 983
- HMMER 3.1b2
- TMHMM 2.0

Productos

Los resultados principales son el archivo de anotación estructural y el archivo de anotación funcional. El archivo de anotaciones funcionales puede ser un input para el workflow ‘MAGs Generation’ (Workflow de Generación de MAGs).

Archivos de Salida Primarios	Descripción
Anotación Estructural	Archivo de anotación estructural de consenso de múltiples herramientas (.gff)
Anotación Funcional	Archivo de anotación funcional de consenso de múltiples herramientas (.gff)
Resumen KEGG	Resumen tabular de la función del gen KEGG (.tsv)
Resumen EC	Resumen tabular de la Comisión de Enzimas (.tsv)
Resumen de filogenia genética	Resumen tabular de filogenia genética (.tsv)

Ejecutando el ‘Metagenome Annotation Workflow’ en NMDC EDGE

Elige un workflow

1. Desde la categoría Metagenómica en la barra de menú de la izquierda, elige ‘Run a Single Workflow’ (Ejecuta un workflow individual).
2. Un nombre de proyecto/ejecución única sin espacios (los guiones bajos están permitidos).
3. Una descripción es opcional, pero recomendada.
4. Elige ‘Metagenome Annotation’ (Anotación Metagenoma) desde el menú desplegable debajo de Workflow.

Input

Este workflow acepta datos de Illumina en formato FASTA como input, y el archivo puede ser comprimido. Se recomienda ingresar los 'contigs' ensamblados desde el workflow de 'Metagenome Assembly' (Workflow de Ensamblaje Metagenomica). **Formatos de archivo aceptables:** .fasta, .fa, .fasta.gz, .fa.gz, fna.gz.

1. Haga clic en el botón a la derecha del espacio en blanco de 'entrada de datos' para seleccionar el archivo de datos para el análisis (Si tiene archivos separados, hay dos espacios en blanco para ingresar sus datos). Se abrirá un cuadro llamado 'Select a File' (Seleccionar un Archivo) para permitir al usuario encontrar los archivos deseados de proyectos ejecutados anteriormente, la carpeta de datos públicos o los archivos cargados por el usuario.
5. Haga clic en 'Submit' (Someter).

Productos

La sección General del resultado muestra cual workflow se ejecutó y la información del tiempo de ejecución.

General					
Workflow	Run	Status	Running Time	Start	End
Metagenome Annotation	On	Done	01:22:05	2021-10-14 15:07:49	2021-10-14 16:29:54
"Project Configuration" : { ... }					

La sección 'Metagenome Annotation Result' (Resultado de la Anotación del Metagenoma) tiene estadísticas para secuencias procesadas, genes previstos e información de calidad general del workflow.

Metagenome Annotation Result

Processed Sequences Statistics

Data type	Number of seqs	Number of bps	Median length	Average length	Length shortest seq	Length longest seq	Standard deviation
final_fasta	25,726	52,201,077	818.5	2,029.118	200	859,644	16,939.403
sequences_with_genes	24,248	51,497,305	865	2,123.775	200	859,644	17,443.493
sequences_without_genes	1,478	703,772	404	476.165	203	1,918	217.554

Predicted Genes Statistics

Feature type	Prediction method	Number of seqs	Number of bps	Median length	Average length	Length shortest seq	Length longest seq	Standard deviation	Number of predicted features
CDS	Prodigal v2.6.3	12,478	3,694,932	180	228.831	75	1,935	156.372	16,147
CDS	GeneMark.hmm-2 v1.05	18,576	35,352,681	480	669.267	90	16,545	616.622	52,823
tRNA	tRNAscan-SE v.2.0.7 (Oct 2020)	451	67,404	76	79.486	56	146	10.062	848
misc_feature	INFERNAL 1.1.3 (Nov 2019)	4	1,454	366.5	363.5	349	372	10.408	4
regulatory	INFERNAL 1.1.3 (Nov 2019)	4	1,454	366.5	363.5	349	372	10.408	4
ncRNA	INFERNAL 1.1.3 (Nov 2019)	4	1,454	366.5	363.5	349	372	10.408	4
rRNA	INFERNAL 1.1.3 (Nov 2019)	4	1,454	366.5	363.5	349	372	10.408	4
tmRNA	INFERNAL 1.1.3 (Nov 2019)	4	1,454	366.5	363.5	349	372	10.408	4
CRISPR	CRT 1.8.2	11	7,170	456	551.538	155	1,168	341.877	13

General Quality Info

Name	Status
Coding density	74.88%
Genes per 1M bp	1,353.8
Seqs per 1M bp	492.83

La sección ‘Browser/Download Output’ (Navegador/Descargar Productos) proporciona archivos de resultados disponibles para descargar. Los resultados principales son los archivos de anotación funcional y de anotación estructural (.gff). El archivo de anotación funcional es necesario para el workflow ‘MAGs Generation’ (Generación de MAGs) junto con los ‘contigs’ ensamblados.

Browser/Download Outputs		
File	Size	Last Modified
MetagenomeAnnotation		
Annotation_Test.faa	20.53 MB	20 days ago
Annotation_Test_cath_funfam.gff	11.89 MB	20 days ago
Annotation_Test_cog.gff	7.92 MB	20 days ago
Annotation_Test_contigs.fna	51.30 MB	20 days ago
Annotation_Test_crt.crisprs	11 kB	20 days ago
Annotation_Test_ec.tsv	1.27 MB	20 days ago
Annotation_Test_functional_annotation.gff	17.43 MB	20 days ago
Annotation_Test_gene_phylogeny.tsv	10.45 MB	20 days ago
Annotation_Test_ko.tsv	2.36 MB	20 days ago
Annotation_Test_ko_ec.gff	44.29 MB	20 days ago
Annotation_Test_pfam.gff	9.71 MB	20 days ago
Annotation_Test_product_names.tsv	5.21 MB	20 days ago
Annotation_Test_proteins.cath_funfam.domtblout	151.86 MB	20 days ago
Annotation_Test_proteins.cog.domtblout	51.46 MB	20 days ago
Annotation_Test_proteins.pfam.domtblout	15.08 MB	20 days ago
Annotation_Test_proteins.smart.domtblout	7.59 MB	20 days ago
Annotation_Test_proteins.supfam.domtblout	339.68 MB	20 days ago
Annotation_Test_proteins.tigrfam.domtblout	3.00 MB	20 days ago
Annotation_Test_smart.gff	3.33 MB	20 days ago
Annotation_Test_structural_annotation.gff	9.99 MB	20 days ago
Annotation_Test_structural_annotation_stats.json	6 kB	20 days ago
Annotation_Test_structural_annotation_stats.tsv	3 kB	20 days ago
Annotation_Test_supfam.gff	12.60 MB	20 days ago
Annotation_Test_tigrfam.gff	1.79 MB	20 days ago
rc	2 B	20 days ago
script	35 kB	20 days ago