

Running workflow "genome annotation"

Step 1: Input dataset

genome fasta file

1: Aspergillus_sequence.fasta
2: uniprot_sprot.fasta

type to filter, [enter] to select all

Step 2: Input dataset

protien database

1: Aspergillus_sequence.fasta
2: uniprot_sprot.fasta

type to filter, [enter] to select all

Step 3: Augustus(version 3.1.0)

Genome Sequence

Output dataset 'output' from step 1

Don't report transcripts with in-frame stop codons (--noInFrameStop)

Not available.

Predict genes independently on each strand

Not available.

Predict the untranslated regions in addition to the coding sequence

Not available.

Model Organism

Aspergillus fumigatus

Predict genes on specific strands

both

Gene Model

complete

GFF formatted output

True

Output options

predicted protein sequences (--protein) coding sequence as comment in the output file (--codingseq) predicted intron sequences (--introns) predicted start codons (--start) predicted stop codons (--stop) CDS region (--cds)

Step 4: NCBI BLAST+ makeblastdb(version 0.2.00)

Molecule type of input

protein

Input FASTA files(s)

Output dataset 'output' from step 2

Title for BLAST database

unipro_sprot_database

Parse the sequence identifiers

Not available.

Enable the creation of sequence hash values

True

Optional ASN.1 file(s) containing masking data

Selection is Optional

Taxonomy options

Do not assign a Taxonomy ID to the sequences

Step 5: Interproscan functional predictions of ORFs(version 5.0.0)**Protein Fasta File**

Output dataset 'protein_output' from step 3

Type of the input sequences

Protein

Applications to run

TIGRFAM: protein families based on Hidden Markov Models or HMMs PIRSF: non-overlapping clustering of UniProtKB sequences into a hierarchical order (evolutionary relationships) ProDom: set of protein domain families generated from the UniProtKB Panther: Protein ANalysis THrough Evolutionary Relationships SMART: identification and analysis of domain architectures based on Hidden Markov Models or HMMs PROSITE Profiles: protein domains, families and functional sites as well as associated profiles to identify them PROSITE Pattern: protein domains, families and functional sites as well as associated patterns to identify them HAMAP: High-quality Automated Annotation of Microbial Proteomes PfamA: protein families, each represented by multiple sequence alignments and hidden Markov models PRINTS: group of conserved motifs (fingerprints) used to characterise a protein family SUPERFAMILY: database of structural and functional annotation Coils: Prediction of Coiled Coil Regions in Proteins Gene3d: Structural assignment for whole genes and genomes using the CATH domain structure database SignalP Eukaryotic Bacteria Phobius: combined transmembrane topology and signal peptide predictor TMHMM: Prediction of transmembrane helices in proteins

Include pathway information

True

Include Gene Ontology (GO) mappings

True

Provide additional mappings

Not available.

Output format

Tab-separated values format (TSV)

Step 6: NCBI BLAST+ blastp(version 0.2.00)**Protein query sequence(s)**

Output dataset 'protein_output' from step 3

Subject database/sequences

BLAST database from your history

None

Not available.

Protein BLAST database

Output dataset 'outfile' from step 4

None

Not available.

Type of BLAST

blastp – Traditional BLASTP to compare a protein query to a protein database

Set expectation value cutoff

0.001

Output format

Tabular (standard 12 columns)

Advanced Options

Hide Advanced Options

☐ Send results to a new history

tools used in this workflow:

```
"tool_id": null,  
"tool_id": null,  
"tool_id": "toolshed.g2.bx.psu.edu/repos/bgruening/augustus/augustus/3.1.0",  
"tool_id": "toolshed.g2.bx.psu.edu/repos/devteam/ncbi_blast_plus/ncbi_makeblastdb/0.2.00",  
"tool_id": "toolshed.g2.bx.psu.edu/repos/bgruening/interproscan5/interproscan/5.0.0",  
"tool_id": "toolshed.g2.bx.psu.edu/repos/devteam/ncbi_blast_plus/ncbi_blastp_wrapper/0.2.00",
```