

Galaxy Workflow ' genome–functional–annotation'

Step

Annotation

Step 1: Input dataset

Genome sequence in fasta format
select at runtime

Step 2: Input dataset

Protein database in fasta format
select at runtime

Step 3: Augustus

Genome Sequence

Output dataset 'output' from step 1

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

False

Predict genes independently on each strand

False

Predict the untranslated regions in addition to the coding sequence

False

Model Organism

Arabidopsis thaliana

Predict genes on specific strands

both

Gene Model

complete

GFF formatted output

False

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+ blastp

Protein query sequence(s)

Output dataset 'protein_output' from step 3

Subject database/sequences

FASTA file from your history (see warning note below)

None

Empty.

None

Empty.

Protein FASTA file to use as database

Output dataset 'output' from step 2

Type of BLAST

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

Set expectation value cutoff

0.001

Output format

Tabular (extended 25 columns)

Advanced Options

Hide Advanced Options