

How to blast a genome assembly for particular genes

Inputs

- assembled genome as contigs, e.g. `genome_assembly.fasta`
- genes of interest to look for in the assembly as a single or multi-fasta file, as amino acids, e.g. `Psilocybin_synthase.fasta`

Input Genome Assembly

- Sample data for this tutorial: Genome assembly of *Psilocybe subaeruginosa*
- Import this history to get test data:
<https://usegalaxy.org.au/u/anna/h/genome-for-blast>
- The history information describes where this data is from.

How can I find sequences for genes of interest?

- Here, we will search for a protein
- go to NCBI: <https://www.ncbi.nlm.nih.gov/>
- In "All Databases", select "Protein"
- Type in protein of interest, e.g. "psilocybin"
- In the left hand column, set up any required filters
 - e.g. under "Species", select "Fungi"
 - e.g. under "Source databases", select "UniProtKB"
- Click on e.g. the first match
- Click on "Identical Proteins"
- Click on "FASTA"
- Highlight the header name and sequence (everything from > onwards) and copy
- In Galaxy, go to "Upload Data", "Paste/Fetch data", paste in the sequence
- Re-name file, e.g. `Psilocybin_synthase.fasta`

Create a blast database from the genome assembly

- tool: **NCBI BLAST+ makeblastdb**
 - tool settings:
 - molecule type: nucleotide
 - Select input:
 - Input is a: Dataset in history
 - FASTA input: `genome_assembly.fasta` [Your assembly may have a different filename]
 - all other settings: leave as default
 - run tool
 - output: nucleotide BLAST database

Search the database with a query

- tool: **NCBI BLAST+ tblastn**
- this takes the nucleotide blast database created with "NCBI BLAST + makeblastdb", and translates it into amino acids
- then, it uses the protein query to search this database
 - tool settings:
 - protein query sequence: e.g., `Psilocybin_synthase.fasta`
 - subject database: BLAST database from your history

- Nucleotide BLAST database: the nucleotide blast database that we created from the genome assembly
 - Type of BLAST: tblastn
 - Set expectation value: leave as default or change
 - Output format: pairwise HTML or other setting if preferred
 - Other settings: leave as default
 - run tool
 - output: see: the `html` file
-
- Note: if you want to see the results in different output formats, you can re-run the tool with the same settings except for "output format", and select a different output each time.

More information

- More about BLAST inputs, parameters, and result interpretation: <https://blast.ncbi.nlm.nih.gov/doc/blast-topics/>
- What other things can you do with BLAST in Galaxy: Peter J. A. Cock, John M. Chilton, Björn Grüning, James E. Johnson, Nicola Soranzo (2015). NCBI BLAST+ integrated into Galaxy. *GigaScience* 4:39 <https://doi.org/10.1186/s13742-015-0080-7>
- BLAST FAQ: <https://blast.ncbi.nlm.nih.gov/doc/blast-help/FAQ.html> (e.g. What is the E value?)