FungiDB: Secondary Metabolites and clusters

1. Finding secondary metabolites and gene clusters

There are several main classes of secondary metabolites in fungi: the non-ribosomal peptides and amino acid-derived compounds, polyketides and fatty acid-derived compounds, and terpenes (Hoffmeister and Keller, 2007). The secondary metabolites can be segregated into groups based on the first step of their biosynthesis, more specifically, the "key enzymes" that are required: Non-Ribosomal Peptide Synthetases (NRPS), PolyKetide Synthases (PKS), followed by Terpene synthases (TS). (Read more: https://www.ncbi.nlm.nih.gov/pubmed/24692239)

a) Initiate a text search in *A. nidulans*. Find all genes annotated with *terpene*.

- Use wild cards to cast a wide search for any combinations of the word *terpene*

b) What can you tell about the chromosomal locations of these genes?

- Can you identify regions with multiple genes?

Hint: Shaded areas are highlighting areas with multiple genes in close proximity. These are not programmatically determined clusters but rather a visual representation of gene dense areas

.

c) Next, navigate to *AN3277* gene record page in FungiDB.

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description	Found in	Score
AN3277	AN3277-T	A. nidulans FGSC A4	ChrVI_A_nidulans_FGSC_A4:1,253,459..1,254,638(+)	Has domain(s) with predicted magnesium ion binding, terpene synthase activity	GOTerms, Product, MetabolicPathways	3

- Note that the product description is referring to domains associated with a specific function, however this gene is not directly annotated to terpene synthesis pathway.

- Investigate further by navigating to the *Protein feature and properties* section of the record page, look through associated documentation about domains, *etc.*, and select *STRING: functional protein association networks:*

▼ STRING: functional protein association networks

Please select the organism:

[Aspergillus nidulans ⬍]

[Submit]

- Look though available information and click on the Databases, which provides a summary of annotation from curated resources:

  i. Do you think this gene may be a part of terpenoid synthesis pathway?

DATABASE KNOWLEDGE

Relevant datasets in Aspergillus nidulans:

annotated pathway (KEGG) *Inter-pathway connection between 'Steroid biosynthesis' and 'Terpenoid backbone biosynthesis'*	AN3376.2 AN8143.2 AN0654.2 AN1592.2 AN8012.2 AN2002.2 AN8878.2 AN3867.2 AN5600.2
annotated pathway (KEGG) *Name: Terpenoid backbone biosynthesis*	AN8143.2 AN0654.2 AN1592.2 AN8012.2 AN2002.2 AN8878.2 AN3867.2 AN5600.2 [... and 12 other proteins]
annotated pathway (KEGG) *Name: Biosynthesis of antibiotics*	AN3376.2 AN8143.2 AN0654.2 AN1592.2 AN8012.2 AN2002.2 AN3867.2 [... and 256 other proteins]
annotated pathway (KEGG) *Name: Sesquiterpenoid and triterpenoid biosynthesis*	AN3277.2 AN3376.2
annotated pathway (KEGG) *Inter-pathway connection between 'Terpenoid backbone biosynthesis' and 'Sesquiterpenoid and triterpenoid biosynthesis'*	AN3277.2 AN3376.2 AN8143.2 AN0654.2 AN1592.2 AN8012.2 AN2002.2 AN8878.2 AN3867.2 AN5600.2

Relevant information transferred from other organisms:

curated complex (Gene Ontology) *protein farnesyltransferase complex*	Schizosaccharomyces pombe: cpp1 cwp1

- Look though available information and click on the Databases, which provides a summary of annotation from curated resources:

d) Navigate to *AN3277* gene record page in in JGI MycoCosm.
- Identify protein domains and their function.

- Identify terpene synthesis clusters in other species by navigating to Annotation, and select Terpenes n(TC) from the Cluster type and refresh for all genomes:

- How many genomes have annotated TC cluster/s?

- Which genome has the highest number of genes?