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.

٩	≑ Gene ID	Transcript ID	韋 Organism 🕹	≑ Genomic Location (Gene) 🥸	Product Description Q	韋 Found in 🥹	韋 Score 🥝 🔒
	AN3277	AN3277-T	A. nidulans FGSC A4	ChrVI_A_nidulans_FGSC_A4:1,253,4591,254,638(+)	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?

DATAB	DATABASE KNOWLEDGE					
Relevant datasets in Aspergillus nidulans:	sets in Aspergillus nidulans:					
annotated pathway (KEGG) Inter-pathway connection between 'Steroid biosynthesis' and 'Terpeno backbone biosynthesis'						
annotated pathway (KEGG) Name: Terpenoid backbone biosynthesis	AN8143.2 O AN0654.2 O AN1592.2 O AN8012.2 O AN2002.2 O AN8878.2 AN3867.2 O 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 organi	vant information transferred from other organisms:					
curated complex (Gene Ontology) protein farnesyltransferase complex	Schizosaccharomyces pombe:					

- 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.

Name:		AN3277			_			
Protein ID		8825						
Location:		ChrVI_A_nid	ulans_FGSC_A4:1253459-1254638					
Strand:		+						
Number of	f exons	c 4						
Description	ec.	Linking AMIN	The Free DB					
AsnGD:		Link to AN33	77 in the Ascentilus Genome Datatu	ana (A				
Best Hit:		ani:AN3277.	2 hypothetical protein (model%; 99, h	15:10	00. 80	one: 17	77. %id: 1	001 [Assergillus nidulans FGSC A4]
total hits(s	hown)	60 (10)						
ASPECT		60 H	GO Desc				Interr	rold Interno Desc
	-	0016829	lyase activity				IPR0056	30 Tercene synthese, metal-binding
Molecular	Punctic	0000287	magnesium ion binding				IPR0056	30 Terpene synthase, metal-binding
View nucl	octide a	and 3-frame tran	slation To Genome Browser					
NCBI blas	e P	redicted number	of transmembrane domains: 0					
	HN3277		To Genove Brows	er				
	1			1180				
		292 67	220 427					
		56	19 59					
	1 .	35 69 104	138 172 206 240 275 309	,342				
				- 0510	ans - 1		meth C	1490/ as 1
	_							
	_			- 5574	1576:	Terpen	old synth	ses [superFakily]
÷								
Flin Star				End	Len	ze z	ID Score	Description (taxihan)
•				- 341	341	1002 1	002 1777	HECG_ani DNS277.2 ani DNS277.2 hypothetical protein
-				_				Exemption montain root with
•				341	341	1002 1	00Z 1698	nr_b_b_07525639 TPH: conserved hypothetical protein
-								CAspergillus modulans FGSC 043
•				330	330	1002 9	37 1561	nr_b_b_121712471 terpene synthase metal binding domain protein
								CAspergillus clavatus NAPL 11
• 3	, -			029	343	87Z 5	7Z 906	nr_b_b_121700682 conserved hypothetical protein
								CAspergillus clavatus NSSL 11
• •				341	342	87I 6	1X 938	Swissprot_416643 ARIS_PENRO RecName: Full=#ristolochene mynthase: Short=#65 AltName: Full=Sequiterpene cyclase
								[Penicillium roqueforti]
• •				345	342	87Z 6	1Z 917	nr_b_b_426643 aristolochene muthase
								[Penicillium roqueforti]
• •	. =			341	342	877 6	07 911	re h h 25501811 0-12406110
								[Penicillium chrysogenum Misconsin 54-1255]
•				100	200		17 007	er h h 1515/025 8 Chain B delatalochene Suthane Encenn) Comley
						0		[Penicillius roasefort]]
• •				222	224	N7 5	97 805	or h h 20456098 EED our lans
						0		[Ponisilius evanual]
				-	164			to be apprending and an end of a second
	-			102	~^		on 1074	The particular second second produces between the particular second se
								cusho 81100 of 2000 to

• 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:

ANNOTATIONS -	MCL CLUSTERS			
GENE ONTOLOGY PFAM DOMAINS		Genomes	Cluster Type	
KOG		Aspergillus aculeatus ATCC16872 v1.1 Aspergillus carbonarius ITEM 5010 v3	NRPS NRPS-Like	
SECONDARY META	BOLISM CLUSTERS	Aspergillus nidulans	PKS DKS Like	
CAZYMES		Neurospora crassa OR74A v1.0	TC	
PEPTIDASES			[
TRANSPORTERS				
TRANSCRIPTION FA	ACTORS			

- How many genomes have annotated TC cluster/s?
- Which genome has the highest number of genes?