

FungiDB: Secondary Metabolites and clusters

Learning objectives:

- Explore InterPro search in FungiDB
- Cross-reference the results with MycoCosm data

- **Finding secondary metabolites and gene clusters**

Fungi produce a plethora of secondary metabolites. 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 (NRPSs), NRPS-like, Polyketide synthases (PKSs), PKS-like, Hybrid PKS – NRPS, Prenyltransferases (DMAT), Terpene cyclases/synthase (TC).

1. Use the InterPro search to identify NRPS genes in all *Aspergilli*.

NRPS genes have at least the three domains:

- AMP-binding (PF00501)
- PP-binding (PF00550)
- Condensation (PF00668)

The screenshot displays the FungiDB InterPro search interface. On the left, a list titled "Identify Genes based on InterPro Domain" shows a comprehensive list of *Aspergillus* species, each with a checkbox and a reference link. On the right, a "Search for..." box contains the term "inter". Below this, a "Genes" section lists various databases and features. A "Domain Database" section shows a dropdown menu set to "PFAM". A "Specific Domain(s)" section contains a search bar with the text "PF00501 : AMP-binding AMP-dependent synthetase/ligase". Below this, three orange arrows point down to the domain names: "PF00550 : PP-binding Phosphopantetheine binding ACP domain", "PF00668 : Condensation Condensation domain", and "NRPS". At the bottom, a diagram of the NRPS biosynthetic pathway is shown, with three steps: Step 1 (PF00501, 2,639 Genes), Step 2 (PF00550, 1,273 Genes), and Step 3 (PF00668, 766 Genes). A button labeled "Add a step" is also present.

Strategy URL:

<https://fungidb.org/fungidb/app/workspace/strategies/import/85a1e3a5a603efc6>

- How many genes were identified in *Aspergillus fumigatus* Af293?

Hint: use the organism filter on the left to limit your search results to Af293 genes only.

The image shows two parts of the MycoCosm interface. On the left is the 'Organism Filter' panel. It has a search bar with 'af293' entered. Below it is a tree view of taxonomic levels: Fungi (766), Ascomycota (766), Eurotiomycetes (766), Eurotiales (766), Aspergillaceae (766), Aspergillus (766), Aspergillus fumigatus (31), and Aspergillus fumigatus Af293 (15). The 'Aspergillus fumigatus Af293' option is selected. On the right is a diagram of an NRPS cluster. It shows three steps: Step 1 (PF00501, 2,639 Genes), Step 2 (PF00550, 2,164 Genes), and Step 3 (PF00668, 858 Genes). The genes in Step 3 are highlighted in yellow and labeled '15 Genes'. An 'Add a step' button is also visible.

- Create a search for NRPS genes in MycoCosm. Access the *A. fumigatus* Af293 portal (<https://mycocosm.jgi.doe.gov/Aspfu1>) and navigate to the Secondary Metabolism Clusters page (under the 'Annotations' tab). How many genes did you get?

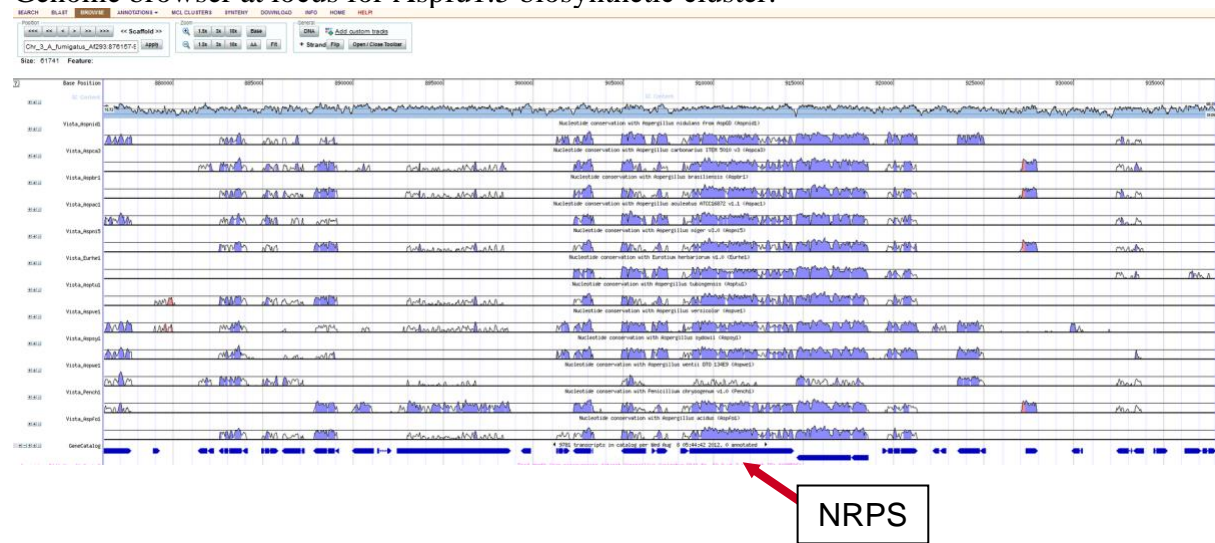
The image shows the 'Secondary Metabolism Clusters' page in MycoCosm. The 'Annotations' tab is selected. The 'Genomes' dropdown is set to 'Aspergillus fumigatus Af293 from AspGD'. The 'Cluster Type' dropdown is set to 'NRPS'. The 'Scale' is set to 'Per Cluster' and 'Clusters Per Page' is set to '50'. A 'Refresh' button is present. Below the filters, a table lists 9 clusters found. The table has columns for Cluster Id, Cluster Type, Scaffold, Size (bp), and Genes. The clusters are Aspfu1.5, Aspfu1.7, Aspfu1.10, Aspfu1.15, Aspfu1.16, Aspfu1.18, Aspfu1.28, Aspfu1.31, and Aspfu1.32. Each cluster is associated with a scaffold and a size in base pairs. The 'Genes' column shows a visual representation of the gene structure with arrows indicating the direction of transcription.

Cluster Id	Cluster Type	Scaffold	Size (bp)	Genes
Aspfu1.5	NRPS	Chr_3_A_fumigatus_Af293:876157-937897	61,740	[Gene structure diagram]
Aspfu1.7	NRPS	Chr_3_A_fumigatus_Af293:3423866-3446129	22,263	[Gene structure diagram]
Aspfu1.10	NRPS	Chr_3_A_fumigatus_Af293:4007787-4023468	15,681	[Gene structure diagram]
Aspfu1.15	NRPS	Chr_1_A_fumigatus_Af293:2655644-2694887	39,243	[Gene structure diagram]
Aspfu1.16	NRPS	Chr_1_A_fumigatus_Af293:4662924-4713331	50,407	[Gene structure diagram]
Aspfu1.18	NRPS	Chr_8_A_fumigatus_Af293:20854-49410	28,556	[Gene structure diagram]
Aspfu1.28	NRPS	Chr_5_A_fumigatus_Af293:3307809-3342792	34,983	[Gene structure diagram]
Aspfu1.31	NRPS	Chr_6_A_fumigatus_Af293:2334637-2372302	37,665	[Gene structure diagram]
Aspfu1.32	NRPS	Chr_6_A_fumigatus_Af293:3004871-3035305	30,434	[Gene structure diagram]

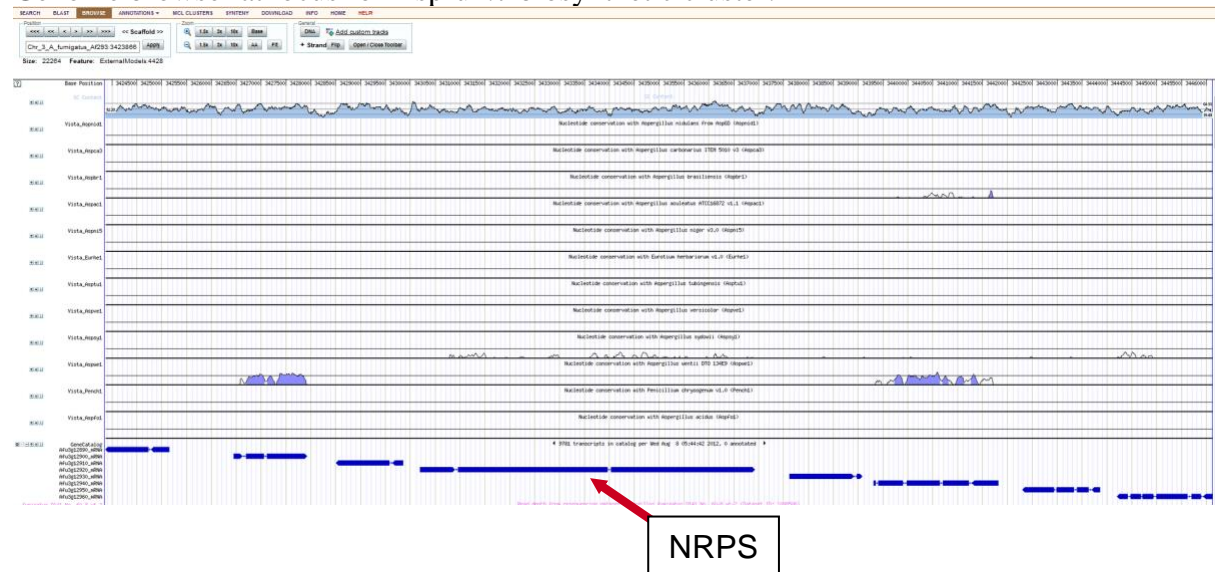
- What do you think may be causing the difference in the predicted gene number?
- This view on MycoCosm allows you to analyze backbone and auxiliary proteins across the entire predicted secondary metabolism cluster. How conserved are these secondary metabolite clusters across related Aspergilli? Click on the scaffold coordinates for Aspfu1.5 and analyze the Vista curve tracks in the genome browser. How many related Aspergilli show some synteny with this region? Repeat this exercise for the next cluster, Aspfu1.7.
 - Answer: Synteny is observed across most Aspergilli for Aspfu1.5, raising the possibility that this SM cluster is widespread across the genus. However,

Aspfu1.7 shows no synteny except for at a couple auxiliary genes in *Aspergillus wentii*, suggesting that it is possibly lineage specific.

Genome browser at locus for Aspfu1.5 biosynthetic cluster:



Genome browser at locus for Aspfu1.7 biosynthetic cluster:



Reference: PMID:24692239