Comparative genomics of fungal cytochrome P450s

D. Nelson April 3, 2008 Edinburgh, Eurofungbase workshop

The diversity of fungal P450s is greater than the diversity found in animal or plant P450s and is comparable to bacterial P450 diversity. [slide animal diversity] Animals currently have 2740 named P450 sequences in 110 CYP families. 38% of these families have only one or two members. However, among vertebrates, no new CYP families have been found in 9 years. [slide plant diversity] In plants there are 2867 named P450 sequences in 95 families including green algae and moss. 28% of these families have only 1 or two members and all of these low abundance families are in algae or moss. If we exclude algae and moss, there are 61 seed plant families with from 7 to 470 members each. [slide fungal diversity] Fungi have 1384 named P450s in 329 families. 62% of these families have only one or two members. Note that this is three times the number of P450 families in animals or plants. [slide bacterial diversity] Bacteria have 813 named P450s in 205 families. 61% of these families have one or two members. The summary of P450 diversity shows that fungi like bacteria are far from saturation of P450 sequence space. There will be many more fungal P450 families discovered and it will take dozens to a hundred or more genomes to begin to fill up the known families with multiple sequences as has happened in vertebrates and is beginning to happen in seed plants.

P450 families are loosely defined as being 40% identical or more at the amino acid level. We know that some true orthologs across species in animals are less than 40% identical. There is also some evidence for this in fungi from P450s inside gene clusters. Because the CYP names have been assigned based on sequence relatedness, these orthologs have been placed in different families. As more information is available some of these names can be corrected. Clearly, the family level of P450 nomenclature is not sufficient to discuss higher-level relationships. A higher order level of nomenclature is required to consider groups of related families and possible orthologs that are less than 40% identical. This is called the clan nomenclature in the P450 field. Animals and plants have been sorted into about 10 clans each. Fungal P450s are more diverse so I have determined a rough set of about 15 clans [slide tree] based on the deepest branches on phylogenetic trees. This was done in 2004. [slide, Deng and Dean] In 2007 a very significant paper published by Deng, Carbone and Dean used 70% bootstrap values as the definition of a fungal clan. This was the first time that a numerical value was placed on the clan breaks. The result was 376 P450 sequences sorted into 115 clans. 65 clans were so-called orphan clans with only one member. I think a hybrid approach may be more useful, without a total dependence on a 70% bootstrap value. Deng and Dean noted that 30 of their clusters were in the common ancestor of the four filamentous fungi. After I added *Nectria haematococca* into this analysis, the number of CYP clusters common to filamentous fungi increased to 36. This seems like a useful division among the P450s but it is nearly four times as many clans as in plants and animals, so I would propose some of these to be subclans, with a smaller number of about 15-20 larger clans.

As more P450 sequences are determined in the fungi, these groups may become clearer.

As I mentioned above, fungal P450 diversity is quite high, but if one limits the discussion to a subset of filamentous ascomycetes including Aspergillus and Fusarium the diversity should drop. [slide fungal diversity 17 species A.] This chart shows the P450 family distribution among 17 species, including *Aspergilus niger* that I just annotated. A. niger has 152 P450s with 19 new families. The set of 8 rows beginning with N. crassa are from complete genomes. The upper 9 rows are from incomplete genomes.

The sequences are displayed in CYP name order. The older sequences have the lower CYP numbers and they tend to be found in more species. One peculiar feature of this layout is the blocks of solid red in some rows. This occurs because a whole genome's collection of P450s are usually named together. The block of sequence families without any gaps represents the historical naming order and these were all in new CYP families at the time they were named. In the case of N. crassa, 31/39 P450 families were new when the genome was annotated. Magnaporthe grisea was next with 46/74 families new. This is followed by Fus. Gram. [slide fungal diversity 17 species B.] The second part of this chart continues with five more genomes P450s being named. As we move to higher CYP family numbers the sequences become less abundant in other species, since the most prolific P450 families have already been found. You would think the number of new families would decrease over time, but *Aspergillus niger*, the fourth Aspergillus genome still had 19/87 new families. [slide new vs. old families] This is summarized in this graph with the blue indicating previously named P450 families and red are the new families in each genome.
Asp. fumigatus looks like an exception, but this is due to sharing 47 CYP families with Asp. nidulans.

The next [slide shows the top 52 P450 families] from this data set. The families have been sorted by number of species present per family. Note that only 8 of 252 families are found in all 8 species. Two of these are the ergosterol biosynthesis P450s CYP51 and CYP61. They are required by all fungi. The fission yeast S. pombe has only these two CYPs. The other abundant P450s are carbon source utilization enzymes or general toxin biosynthesis P450s. We do see a distinct falling off of conserved families in these 8 related fungal species. This is shown in the [next slide].

About half of the CYP families in these 8 fungi have only one member. This is very different from what we saw in plants or especially in vertebrates, where every vertebrate has the same set of P450 families and only the subfamilies differ.

I illustrate this by a car analogy. [slide vert. cars] Vertebrates are very similar while [slide fungal cars] fungi are quite novel except for the basic CYP51 and CYP61. Everything else can vary.

As we saw from Geoff Turner's talk on secondary metabolism clusters, there are blocks of genes dedicated to making toxins, pigments, antibiotics and other secondary metabolites. These clusters share some common features like fungal transcription factors, efflux pumps, polyketide synthases, non-ribosomal peptide synthases and P450s. By searching for these components the clusters can be identified. [slide Perrin] More

recently, the global secondary metabolite regulator LaeA has been used to define the boundaries of some of these clusters by expression studies. 22 of these gene clusters have been identified in *Aspergillus fumigatus*. [slide gliotoxin] 9 of these contain P450 genes. As an example I show you the gliotoxin gene cluster. This cluster is related to the sirodesmin gene cluster in *Leptosphaeria maculans* as shown in this figure from a 2005 paper by Gardiner, Waring and Howlett.

I have labelled the cytochrome P450 genes. These genes are color-coded to show orthologs. Note that CYP613B1 gliC is a presumed ortholog to CYP5082B1 sirC but they are only 34% identical. This is an example of a place where the nomenclature for P450s has failed to identify the ortholog when the genes were named. There are 3 CYP5082A genes in Aspergillus fumigatus, Neosartorya fischeri and Aspergillus terreus. There is a CYP613A1 in Magnaporthe grisea and CYP613B2 in Neosartorya fischeri is 95% identical to CYP613B1. These are surely doing the same thing. Then there are CYP613C1 in Aspergillus oryzae, CYP613C2 in Aspergillus clavatus, and CYP613D1 in Aspergillus oryzae. Examination of the surrounding genes may show toxin gene clusters, but the conservation of the genes may indicate a different product is being made.

I was invited here because I am working with Diane Kelly on annotation of the 117 Aspergillus nidulans P450s. To get an idea of what these genes are doing I first sorted them by chromosomal location to see if any were in P450 gene clusters. There were 17 clusters of P450 genes involving 44 P450s (38%). After that I searched for all the polyketide synthase genes and found 29 of these. 15 of 29 PKS genes have P450s near them, and there are 24 P450s near these 15 PKS genes. I did the same for the NRPS genes and I found 14 of those. One gene is a PKS/NRPS hybrid. 8 of the 14 NRPS genes have P450s nearby. 5 NRPS genes are adjacent to a P450 and three NRPS genes are only 3-5 genes distant. 15 P450s are near these 8 NRPS genes, but 7 of these were in already seen in the PKS list. (32/117) 27% of A. nidulans P450s are near a PKS or an NRPS gene.

The next several slides show the 17 A. nidulans gene clusters. I won't go through them all, but I will point out a few interesting ones. The CYP617D1 CYP51F2 pair has a group of genes between them that are used in aromatic amino acid catabolism. These include a 4-hydroxyphenylpyruvate dioxygenase, homogentisate 1,2-dioxygenase, fumarylacetoacetate hydrolase, maleylacetoacetate isomerase. These are four genes in the pathway of tyrosine degradation. The last gene resembles glutathione-S-transferase and this may be a miss-annotation found in some other P450 gene clusters that may be degrading aromatic compounds. The two P450s may be performing hydroxylations of an aromatic ring in preparation for the dioxygenases. It is interesting to note that the CYP51F2 gene is seen in 5 other filamentous fungi. In Aspergillus terreus CYP51F2 is four genes from the four aromatic amino acid catabolic genes and 12 genes from a CYP617D ortholog (64% identical). In A. niger the CYP51F2 is only 2 genes from the catabolic cluster but CYP617D6 is 244kb away. It seems very unlikely that CYP51F2 is a sterol demethylase like its more famous cousin CYP51F1. [slide Nectria 504 cluster] In Nectria haematococca the same catabolic cluster is also surrounded by P450s. The CYP504B gene has been shown in A. nidulans to be involved in phenylacetate

metabolism, though it is not near to the gene cluster. In Nectria it is right next door and the CYP504A gene is not far away, only 60kb downstream.

About 700 genes away on the same chromosome there is a three P450 cluster that includes a geranylgeranyl diphosphate synthase gene. These genes make cyclic diterpenes. Nearby there are a fungal transcription factor, a MFS transporter, and the 3 P450s including CYP566C1.

The geranylgeranyl diphosphate synthase gene is 65% identical to a Neosartorya fischeri gene that is adjacent to CYP566C3 and [slide gliotoxin cluster] another gene in the gliotoxin gene cluster we saw already in A. fumigatus. A. nidulans does not have the gliotoxin gene cluster and there is a transposon between GliT and CYP566C2 in A. fumigatus and Neosartorya. So the gliotoxin cluster may be excised as a module and be independent of the geranylgeranyl toxin cluster. However, their proximity suggests a hybrid toxin may be possible with diterpene and ETP components.

Chromosome VI has a four P450 gene cluster that includes a PKS gene, a fungal transcription factor an MFS transporter and most interesting a non-plant terpene cyclase that is 57% identical to Penicillium roqueforti aristolochene synthase (Ari1). The high sequence conservation suggests this cluster is involved with sesquiterpene toxin biosynthesis. [slide terpene cluster] The same cluster exists in Asp. clavatus with all four P450s. A. oryzae has a similar cluster with two of the four P450s and a new one CYP682B2. This cluster seems to be decaying since it lost the PKS and Zn transcription factor. A. terreus has the terpene cyclase, but the genomic region around this gene has not been sequenced yet.

[slide part 2 of table] Here we see a conidiophore pigment gene cluster and toward the bottom is the well studied sterigmatiocystin gene cluster.

[slide part 3 of table] The last cluster I want to mention is this large one on chr 4. There are 5 P450 genes in a16 gene region. Also included are a PKS gene, an efflux pump, a Zn fungal transcription factor, an NRPS, an FAD monooxygenase and a gene that is like dimethylallyltryptophan synthase, which catalyses the first step in ergot alkaloid biosynthesis. This is certainly a toxin biosynthetic cluster and it may be making a alkaloid toxin.

We have seen a variety of gene clusters in several of these filamentous fungal species. Some of them are conserved between species and others appear to be decaying or changing function. Asp. nidulans has at least 17 of these clusters. Asp. fumigatus had 22. In these 8 completed genomes we can expect about 150 clusters. The next step will be identification of all these clusters and side by side comparison to determine the likely structure of the original cluster and tracking of changes over time as the species have diverged. The cytochrome P450s are an integral part of many of these pathways. They probably account for much of the diversity seen in the products and they might be mixed and matched to exploit their functions to make novel antibiotics or other pharmaceuticals.