Bioinformation by Biomedical Informatics Publishing Group open access
www.bioinformation.net Hypothesis

Phylogenetic analysis of homologous fatty acid synthase and polyketide synthase involved in aflatoxin biosynthesis

Marina Marcet-Houben[1], Maria Cabré[2], José L. Paternáin[2] and Antoni Romeu[2,] *

[1]Department of Bioinformatics, Centro de Investigación Principe felipe, E-46013 Valencia, Spain;
[2]Department of Biochemistry and Biotechnology, University Rovira i Virgili, E-43007 Tarragona, Spain;
Antoni Romeu* - E-mail: antoni.romeu@urv.cat; * Corresponding author

received May 24, 2008; accepted August 31, 2008; published September 20, 2008

Abstract:
The first two steps of aflatoxin biosynthesis are catalyzed by the HexA/B and by the Pks protein. The phylogenetic analysis clearly distinguished fungal HexA/B from FAS subunits and from other homologous proteins. The phylogenetic trees of the HexA and HexB set of proteins share the same clustering. Proteins involved in the synthesis of fatty acids or in the aflatoxin or sterigmatocystin biosynthesis cluster separately. The Pks phylogenetic tree also differentiates the aflatoxin-related polypeptide sequences from those of other kinds of secondary metabolism. The function of some of the *A. flavus* Pks homologues may be deduced from the phylogenetic analysis. The conserved sequence motifs of protein domains shared by HexA/B and Pks - namely, β-polyketide synthase (KS), acetyl transferase (AT) and acyl carrier protein (ACP) - have been identified, and the HexA/B and Pks involved in aflatoxin biosynthesis have been distinguished from those involved in primary metabolism or other kinds of secondary metabolism.

Keywords: aflatoxin; aflatoxin biosynthesis; HexA/B multienzymatic complex; polyketide synthase; Aspergillus

Background:

Aflatoxins, a group of polyketide-derived compounds **[1]**, are toxic and carcinogenic secondary metabolites synthesised only by such Aspergillus species as *Aspergillus flavus*, *Aspergillus parasiticus* and *Aspergillus nominus* **[2, 3]**. However, sterigmatocystin, which is the penultimate precursor of aflatoxin, is produced by several species of Aspergillus **[4]**. Most genes encoding the aflatoxin biosynthetic enzymes have been identified. These genes make up a large gene cluster in the fungal genome of about 90 kb **[5]**. The genomic DNA regions that represent the complete sequence of the well-organized aflatoxin pathway cluster are available for each species, *A. flavus*, *A. parasiticus* and *A. nomius* (EMBL ID: AF391094, AY510451, AY510454, respectively). The aflatoxin biosynthetic pathway starts with the reception of acetyl-CoA and malonyl-CoA substrates, which produce hexanoic acid by a repetitive reaction sequence (Mahanti and colleagues 1996). Aspergillus species use the HexA/B multienzymatic complex, a close homologue to the fatty acid synthase (FAS) system, to catalyze the first step of the aflatoxin biosynthetic pathway **[6]**. Thus, FAS is responsible for fatty acid synthesis in primary metabolism, while HexA/B synthesises the carbon acyl chain in secondary metabolism. In the aflatoxin gene cluster, *fas1* and *fas2,* large genes that encode for HexB and HexA respectively, are located side by side. These genes are also named *hexA* and *hexB* because of the hexanoate synthase α and β subunits, respectively. The same reaction that occurs in fatty acid synthesis is considered to be involved in the formation of hexanoate in aflatoxin synthesis **[1]**. The FAS from bacteria and plants is a complex of at least seven different polypeptides **[7]**. In fungi, all seven activities reside in two polypeptides: 210-kD α(Fas-2, HexA) and 230-kD β(Fas-1, HexB) **[8]**. In vertebrates, they are located in a single, large, 270-kD polypeptide **[9]**. As far as the structure of fungal FAS is concerned, the acyl carrier protein (ACP), the ketoacylreductase (KR) (E.C. 1.1.1.100) and the ketoacyl synthase (KS) (E.C. 2.3.1.41) domains are distributed in the α-chain; and the acetyl transferase domain (AT) (E.C. 2.1.3.39), the enoylreductase (ER) (E.C. 1.3.1.9), the dehydratase (DH) (E.C. 4.2.1.61) and the malonyl-ACP transferase (MPT) (E.C. 2.1.3.39) domains are distributed in the β-chain **[8]**.

In the aflatoxin gene cluster, the *pksA*, which is also a large gene encoding for a polyketide synthase (Pks), is located 3987 bp upstream from the *fas2* gene position. Fungal polyketide synthases involved in the aflatoxin biosynthesis are defined as iterative type I synthases which are multidomain enzymes that are similar to the FAS system **[10]**. Polyketide synthases catalyze the second step of the aflatoxin biosynthetic pathway, in which hexanoyltethrahydroxyanthrone is formed from the hexanoate resulting from step one **[4]**. As far as the structure of these proteins is concerned, the KS, AT, and ACP domains are essential for both FAS (HexA/B) and Pks, whereas the KR, DH, and ER domains are present in FAS (HexA/B), but are absent in aflatoxin-related Pks. This absence causes the formation of a β-polyketone which, with a subsequent cyclation, produces an anthrone that has a hexanoyl and four hydroxyls as the final reaction product.

In this study we have phylogenetically analysed the amino acid sequences of fungal FAS (HexA/B) and Pks enzymes. Some protein domains are essential for both FAS (HexA/B) and Pks, but their biological function is very different so evolutionary relationships and protein-level features should be indicated.

Methodology:

Sequences of Fas-1 (HexB), Fas-2 (HexA) and Pks were extracted from the UniProt and Aspergillus flavus database (http://www.aspergillusflavus.org). To construct the FAS phylogenetic tree, 36 sequences were used from 22 fungal organisms. Most organisms belonged to the Ascomycota phylum and only two sequences belonged to Basidiomycota. The UniProt sequences were retrieved using the BlastP search **[11]** on default settings against all organisms found in UniProt. The sequences used as queries during the blast were the *A. flavus* FAS aflatoxin sequences (Q5VDA2, Q5VDA1). From the original BlastP results, 32 sequences were selected. These sequences had an E-value of 0.0, there were no duplicates and they appeared both in HexA and HexB. For the Pks phylogenetic tree, 39 sequences were used, mostly belonging to the Aspergillus species. The remaining sequences were obtained from other Pezizomycotina organisms such as *Gibberella zeae* or *Emericella nidulans*. Sequences were extracted from the UniProt database using the BlastP program, with default parameters, and the sequence used as the query was the *A. flavus* aflatoxin Pks (Q5VDF2). From the group of similar sequences that resulted from the blast, we selected 35 in accordance with the species they belonged to. Thus all the results belonging to the Aspergillus species were selected as were the sequences that belonged to organisms that had previously appeared in our analysis of the FAS proteins (i.e. *Neurospora Crassa*) or which were taxonomically related to them (*Bipolaris oryzae*).

To select the *A. flavus* sequences, we initially ran a blast of the Aspergillus flavus database using the same sequences we had used as queries in the initial blast search of UniProt, and obtained four different sequences for each protein. Some of these were not complete so we took the various segments and made an initial study. ClustalW was used for the multialignment and the subsequent phylogenetic tree construction in its default settings. Once the first version of the phylogenetic tree had been constructed we used the sequences that appeared closely clustered to the genes extracted from the Aspergillus flavus database to run a second BlastP. Once the complete sequences of the protein homologues of *A. flavus* had been added to the initial sequences obtained from UniProt, the final phylogenetic tree construction was carried out using ClustalW **[12]**. The tree was built by neighbour-joining (NJ) and the parameter values were on default settings. Bootstrap values were generated by 1000 replications of the bootstrap procedure. Trees were represented using the program MEGA 3.1 **[13]**.

Results and discussion:

Phylogenetic analysis

Figure 1 show the phylogenetic tree based on HexA and HexB sequence multialignment. The fact that these two proteins are involved in different kinds of metabolism means that within a single species there can be several homologues. We made two phylogenetic trees, one for each protein, which contained sequences belonging to 22 fungal species. Of these species, seven had at least two homologues, and the maximum was five (*A. Oryzae*). The *A. flavus* species contain four homologues for both the HexA and the HexB proteins.

The two trees share many of their features and show that in most cases the evolution of the two proteins is very similar. In the case of HexA (Figure 1a), we find four main clusters strongly supported by bootstrap values. Each one of them has a homologue of *A. flavus*. These sequences are always closely related to a matching sequence in *A. oryzae*, according to the similarity between the two species **[14]**. The cluster coloured in blue contains the proteins that belong to the aflatoxin gene cluster. Closely related to them is the gene product belonging to the sterigmatocystin cluster and a hypothetical protein from *Coccidioides immitis* (Q1DM99). The fact that *C. immitis* has two apparent homologues of HexA makes us think that it is related to some kind of secondary metabolism.

The second cluster coloured in green, gathers proteins involved in the synthesis of fatty acids. In this cluster, we find many of the organisms that have only one gene encoding for Fas-2, such as *Saccharomyces cerevisiae* or *Candida albicans*. We also find the HexA homologues involved in the primary metabolism of *E. nidulans* and *C. immitis*, as well as the *A. flavus* sequence AFSG001384 and the *A. oryzae* sequence Q2U734. From the data obtained, we can deduce that these *A. flavus* and *A. oryzae* sequences are also involved in primary metabolism. Proteins from other species are also described as hypothetical proteins whose function can be deduced by the same reasoning, so they are probably Fas-2 proteins (i.e. Q41B77 *G. zeae*). *A. parasiticus* and *A. nomius* do not appear in this group but that is probably because their enzymes have not yet been sequenced or introduced into the Uniprot database. The remaining homologues are distributed in two clusters, coloured in black. Their function has not been described but they may belong to some kind of secondary metabolism.

The phylogenetic tree belonging to the HexB (Figure 1b) protein is nearly a duplicate of the HexA tree. The main clusters are supported by high bootstrap values. The aflatoxin cluster remains completely unchanged, and even the branching pattern is the same. The main difference in the tree is found in the hypothetical protein clusters (coloured in black). Instead of two hypothetical protein clusters there are three, but the low value of the bootstrap that situates *E. nidulans* (Q5AV07) closer to *P. nodorum* than to the two Aspergillus indicates that that this grouping is not statistically supported. The other difference between the two trees is the position of *Schizosaccharomyces pombe* within the fatty acid cluster. Although it appears close to the Pezizomycotina in the tree representing HexA evolution, in the HexB tree it appears to have separated from the ancestral organism before the Pezizomycotina and the Saccharomycotina branched into two different groups.

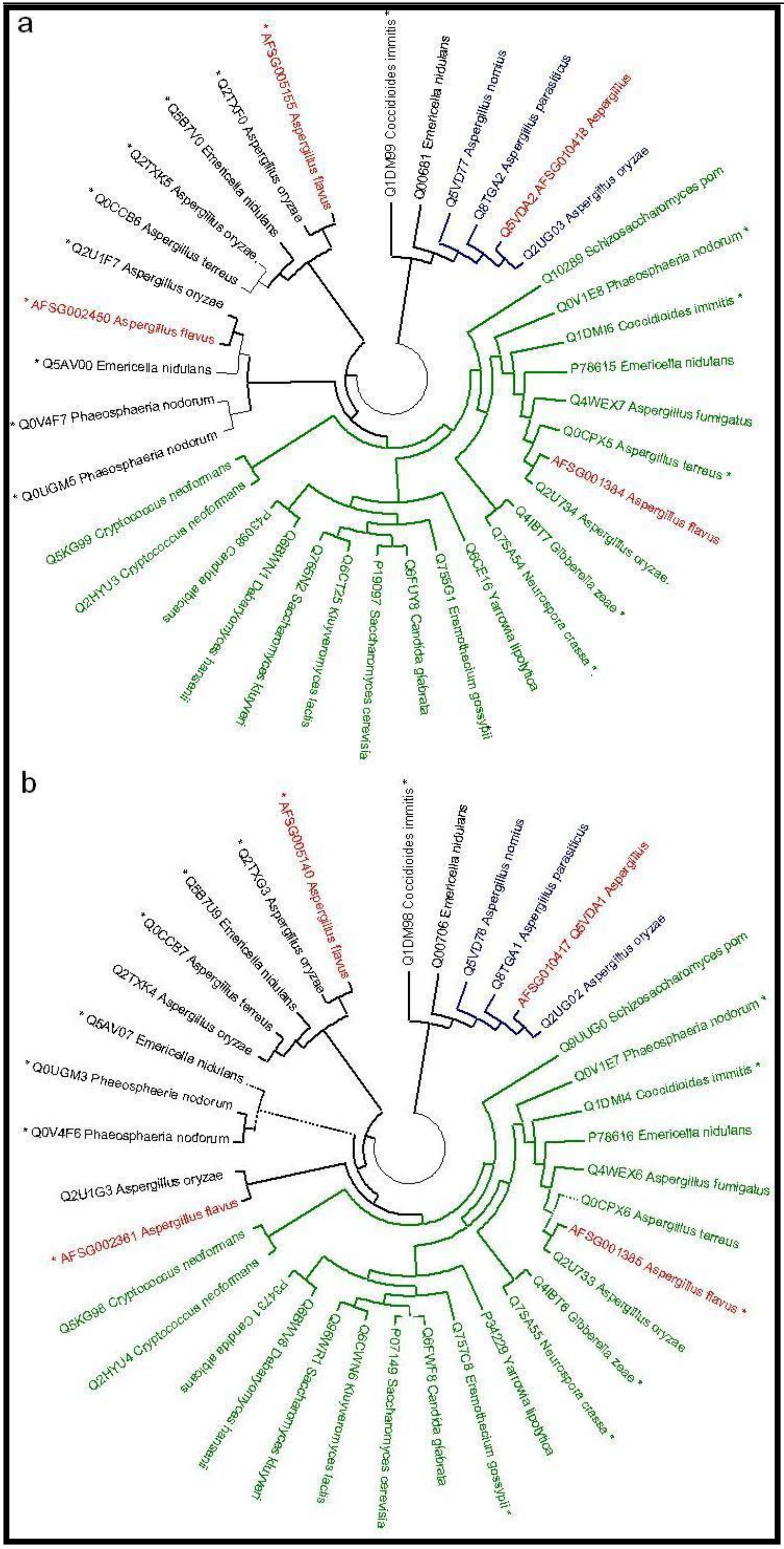

a
Q1DM99 Coccidioides immitis *
Q00681 Emericella nidulans
Q5VD77 Aspergillus nomius
Q8TGA2 Aspergillus parasiticus
Q5VDA2 AFSG010418 Aspergillus
Q2UG03 Aspergillus oryzae
Q10289 Schizosaccharomyces pom
Q0V1E8 Phaeosphaeria nodorum *
Q1DMI6 Coccidioides immitis *
P78615 Emericella nidulans
Q4WEX7 Aspergillus fumigatus
Q0CPX5 Aspergillus terreus *
AFSG001384 Aspergillus flavus
Q2U734 Aspergillus oryzae
Q4IBT7 Gibberella zeae *
Q7SA54 Neurospora crassa *
Q6CE16 Yarrowia lipolytica
Q735G1 Eremothecium gossypii
Q6FUY8 Candida glabrata
P19097 Saccharomyces cerevisia
Q6CT25 Kluyveromyces lactis
Q765N2 Saccharomyces kluyveri
Q6BWN1 Debaryomyces hansenii
P43098 Candida albicans
Q2HYU3 Cryptococcus neoformans
Q5KG99 Cryptococcus neoformans
* Q0UGM5 Phaeosphaeria nodorum
* Q0V4F7 Phaeosphaeria nodorum
* Q5AV00 Emericella nidulans
* AFSG002450 Aspergillus flavus
* Q2U1F7 Aspergillus oryzae
* Q0CCB6 Aspergillus terreus
* Q2TXK5 Aspergillus oryzae
* Q5B7V0 Emericella nidulans
* Q2TXF0 Aspergillus oryzae
* AFSG005155 Aspergillus flavus
b
Q1DM98 Coccidioides immitis *
Q00706 Emericella nidulans
Q5VD78 Aspergillus nomius
Q8TGA1 Aspergillus parasiticus
AFSG010417 Q5VDA1 Aspergillus
Q2UG02 Aspergillus oryzae
Q9UUG0 Schizosaccharomyces pom
Q0V1E7 Phaeosphaeria nodorum *
Q1DMI4 Coccidioides immitis *
P78616 Emericella nidulans
Q4WEX6 Aspergillus fumigatus
Q0CPX6 Aspergillus terreus
AFSG001385 Aspergillus flavus *
Q2U733 Aspergillus oryzae
Q4IBT6 Gibberella zeae *
Q7SA55 Neurospora crassa *
P34229 Yarrowia lipolytica
Q757C8 Eremothecium gossypii *
Q6FVF8 Candida glabrata
P07149 Saccharomyces cerevisia
Q6CWN6 Kluyveromyces lactis
Q96WR1 Saccharomyces kluyveri
Q6BWV8 Debaryomyces hansenii
P34731 Candida albicans
Q2HYU4 Cryptococcus neoformans
Q5KG98 Cryptococcus neoformans
* AFSG002361 Aspergillus flavus
Q2U1G3 Aspergillus oryzae
* Q0V4F6 Phaeosphaeria nodorum
* Q0UGM3 Phaeosphaeria nodorum
* Q5AV07 Emericella nidulans
Q2TXK4 Aspergillus oryzae
* Q0CCB7 Aspergillus terreus
* Q5B7U9 Emericella nidulans
* Q2TXG3 Aspergillus oryzae
* AFSG005140 Aspergillus flavus

Figure 1: Phylogenetic tree of fungal HexA, HexB based on protein sequence multialignment. Tree branch lengths are drawn proportional to the amount of sequence changes. Thick branches indicate bootstraps of over 900, thin branches indicate bootstraps between 900 and 750, and lesser bootstraps are indicated by dots. Sequences are indicated by UniProt code and species name (*A. flavus* gene name). Branches corresponding to proteins involved in the aflatoxin gene cluster are denoted in blue, and those involved in primary metabolism are in green. *A. flavus* sequences are marked in red. Sequences described as hypothetical proteins are marked with an asterisk (*).

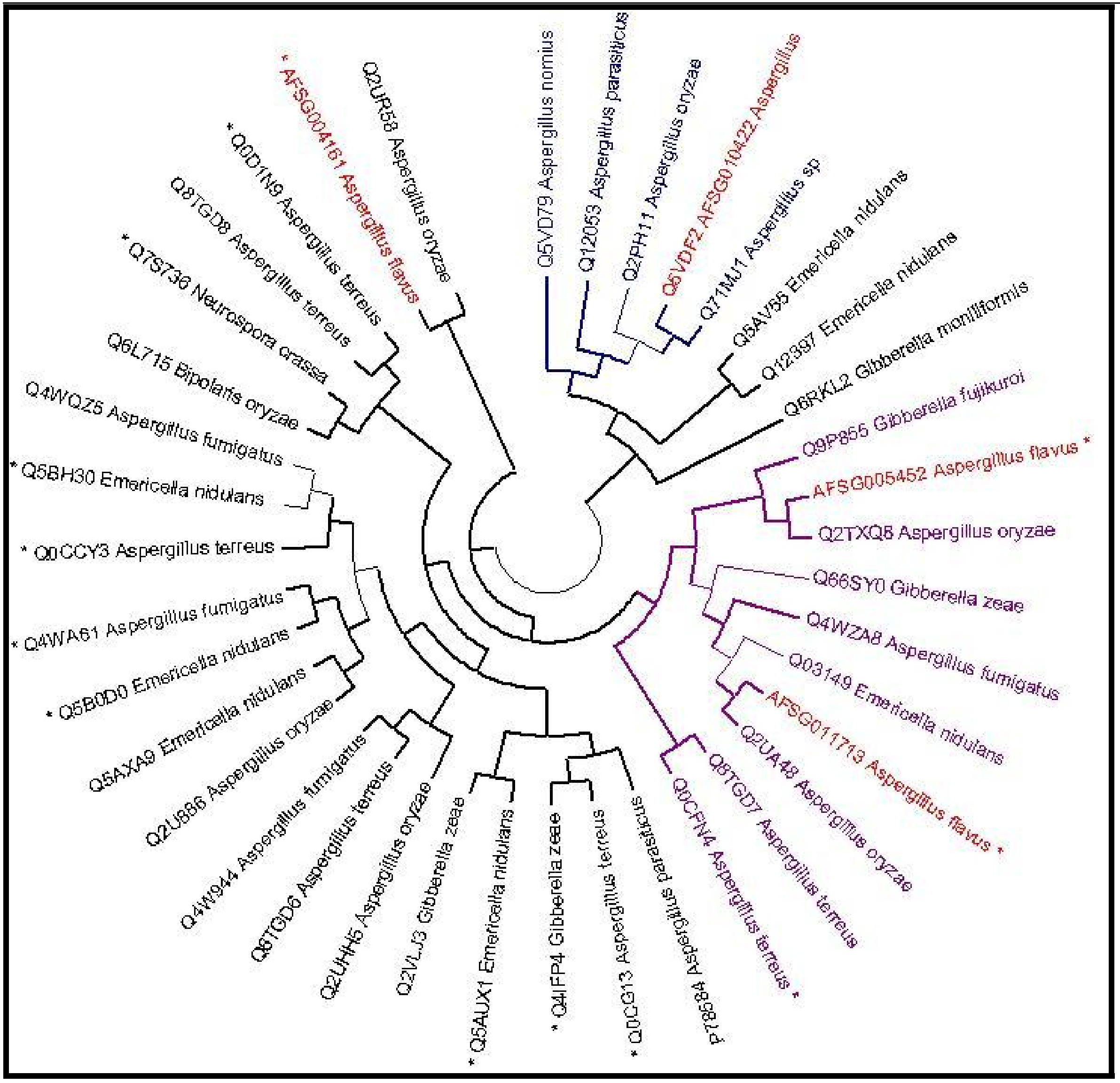

* AFSG004161 Aspergillus flavus
Q2UR58 Aspergillus oryzae
Q5VD79 Aspergillus nomius
Q12053 Aspergillus parasiticus
Q2PH11 Aspergillus oryzae
Q5VDF2 AFSG010422 Aspergillus
Q71MJ1 Aspergillus sp
Q5AV55 Emericella nidulans
Q12397 Emericella nidulans
Q6RKL2 Gibberella moniliformis
Q9P855 Gibberella fujikuroi
AFSG005452 Aspergillus flavus *
Q2TXQ8 Aspergillus oryzae
Q66SY0 Gibberella zeae
Q4WZA8 Aspergillus fumigatus
Q03149 Emericella nidulans
AFSG011713 Aspergillus flavus *
Q2UA48 Aspergillus oryzae
Q8TGD7 Aspergillus terreus
Q0CFN4 Aspergillus terreus *
P78584 Aspergillus parasiticus
* Q0CG13 Aspergillus terreus
* Q4IFP4 Gibberella zeae
* Q5AUX1 Emericella nidulans
Q2VLJ3 Gibberella zeae
Q2UHH5 Aspergillus oryzae
Q8TGD6 Aspergillus terreus
Q4W944 Aspergillus fumigatus
Q2U886 Aspergillus oryzae
Q5AXA9 Emericella nidulans
* Q5B0D0 Emericella nidulans
* Q4WA61 Aspergillus fumigatus
* Q0CCY3 Aspergillus terreus
* Q5BH30 Emericella nidulans
Q4WQZ5 Aspergillus fumigatus
Q6L715 Bipolaris oryzae
* Q7S736 Neurospora crassa
Q8TGD8 Aspergillus terreus
* Q0D1N9 Aspergillus terreus

Figure 2: Phylogenetic tree of fungal Pks based on protein sequence multialignment. Tree branch lengths are drawn proportional to the amount of sequence changes. Thick branches indicate bootstraps of over 900, thin branches indicate bootstraps between 900 and 750, and lesser bootstraps are indicated by dots. Sequences are indicated by UniProt code and species name (*A. flavus* gene name). Branches corresponding to proteins involved in the aflatoxin gene cluster are denoted in blue, those involved in the synthesis of non-melanin pigments are marked in purple. *A. flavus* sequences are written in red. Sequences described as hypothetical proteins are marked with an asterisk (*).

The function of the other clusters is difficult to determine. Most of them are composed of hypothetical proteins and Pks proteins of unknown functions. However, there are cases in which we may be able to computationally determine the function of a Pks protein because of its similarity with other proteins which have a known function. This is the case of one of the *A. flavus* Pks sequences (AFSG011713). This protein is clustered in a group of proteins that are involved in the synthesis of several pigments (Bikaverin, Aurofusarin and Conidial yellow pigment) (coloured in purple). This particular sequence of *A. flavus* is 72% similar to the protein belonging to *E. nidulans* (Q03149) and 71% similar to the one belonging to *A. fumigatus* (Q4WZA8). Both these proteins participate in the synthesis of Conidial yellow pigment so it is very likely that AFSG011713 also synthesises this same compound **[6]**.

Figure 2 shows the phylogenetic tree based on the Pks sequence multialigment. The phylogenetic tree is distributed in five highly bootstrap supported clusters. The sequences that belong to the aflatoxin biosynthetic cluster are clustered together, as expected, and close to them we find the genes that participate in the sterigmatocystin biosynthesis. The other sequence in this cluster belongs to *Gibberella moniliformis* and, while its exact function remains unclear, it has previously been found to be closely related to the aflatoxin Pks genes **[15]**.

ISSN 0973-2063 (online) 0973-2063 (print)
Bioinformation 3(1): 33-40 (2008)

Other clusters have sequences with known functions, but there is not enough evidence to derive these functions to the other proteins in the cluster. For example, in this same cluster, *Gibberella fujikuroi* (Q9P855) is involved in the synthesis of the red pigment, bikaverin. Closely clustered with this sequence we find another homologue of the Pks of aflatoxin biosynthesis for *A. flavus* (AFSG005452) and its matching *A. oryzae* sequence (Q2TXQ8). These two sequences may have the same function as the sequence that belongs to *Gibberella fujikori*, and yet they only share about 50% of similarity. Further studies should be made to safely assign this function to these sequences.

The phylogenetic trees are consistent with the great similarity between *A. flavus* and *A. oryzae*. The former is a plant, animal and human pathogen that can produce aflatoxins. *A. Oryzae*, on the other hand, is not recognized as a pathogen and, although it has the aflatoxin gene cluster, it is unable to synthesise this mycotoxin. As the clusters drawn in the tree show, for each sequence of *A. flavus* there is a corresponding sequence of *A. oryzae.*

ACP domain: comparison between primary and secondary metabolism

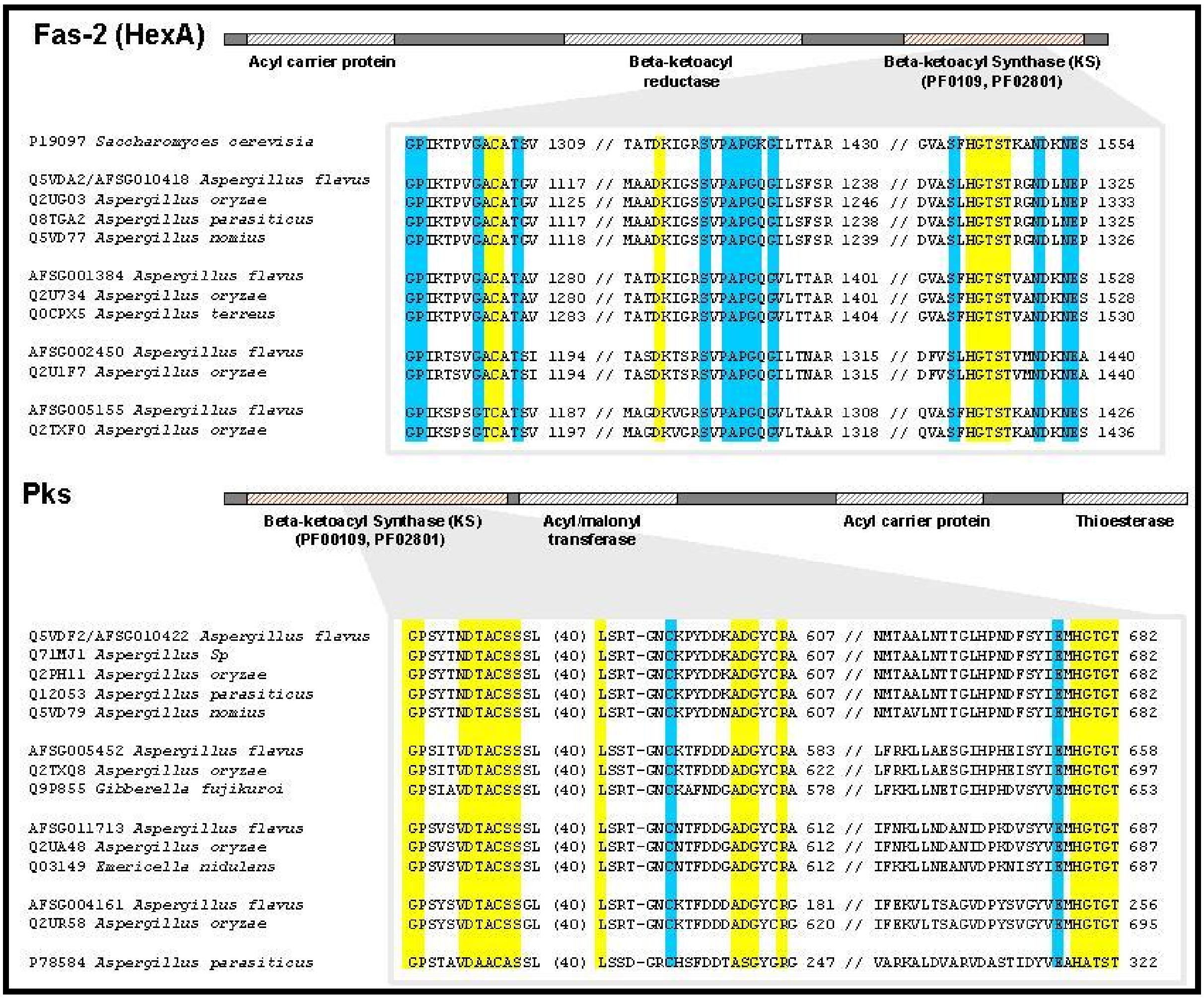
Aspergillus species have several homologous FAS proteins which are believed to participate in different kinds of metabolism. Two of these proteins are involved in the synthesis of fatty acids and aflatoxins. The differences between the two kinds of proteins can be found on the sequence level. A conserved pattern has been described in the ACP domain of Fas-2, which holds the binding site (prosite: PS00012 Phosphopantetheine attachment site). This prosite is built around the active Ser which unites the phosphopantetheine prostetic group so that it can attach activated fatty acids. If we look for this conserved pattern in the multialignment, we find that it has been aligned together in all the fungal sequences. Comparison of the use of amino acids around this Ser position shows that there are two kinds of differences (results not shown). The first kind is easily explained by the different origins of the sequences. So in the second position downstream from the active Ser we find that sequences of the organisms that belong to the Eurotiomycetes class opt for a Leu while the sequences that belong to the Saccharomycotina class have a Val. The chemical proprieties of these two amino acids are not very different so it is not surprising to see that they can be interchanged in organisms as closely related as the ones we have been comparing. Other differences between the sequences are easily attributed to the different functions that the two proteins have. So we find that in the fourth position upstream from the active Ser we have a Val for the sequences belonging to primary metabolism, a Ser for the aflatoxin sequences and a Cys for the sequence of *E. nidulans* that participates in sterigmatocystin biosynthesis. The chemical properties of these amino acids are considerably different. They are not as easily interchanged as the Val and Leu amino acids we compared before and could introduce some differences into the protein structure, especially as they are situated so closely to the active site. In addition, the Arg located one position upstream from the ACP binding Ser site in sequences involved in aflatoxin biosynthesis means that the prosite is not noticed.

The ACP domain is a shared domain between Pks and HexA and, although they have the same function, there are differences between the two domains. While the ACP domain found in the HexA protein is located at the N-terminal end of the protein, the one found in the Pks is located at the C-terminal end. This domain also has a fragment of the sequence surrounding the binding site that nearly complies with the consensus pattern of the prosite found in the HexA protein of primary metabolism. Comparison between the three conserved parts of the phosphopantetheine attachment site (results not shown) demonstrates that while the binding site remains conserved there are substantial differences between the surrounding amino acids. Only one other amino acid (Gly) remains unchanged: it is in the third position upstream from the active Ser and it is already established in the consensus pattern that few other amino acids can occupy its place.

Conserved patterns in the KS and AT protein domains

While HexA, HexB and Pks have different functions, they share several enzymatic activities which are found in specific domains in all the polypeptides: namely, the KS, AT and ACP. In order to find patterns that have been conserved in the different domains through evolution and then use them to identify other homologous sequences, we compared each of the domains to a group of bacterial sequences. To this end, each domain from the *A. flavus* aflatoxin proteins was used to run a BlastP search of all bacterial sequences (results not shown). A multialignment was done so we could find the pattern signatures that would define each domain. The motives shared by both fungi and bacteria from some of the domains are shown in Figure 3 and Figure 4.

Figure 3 shows a comparative analysis of the HexA- Pks-KS domain. The sequences selected as examples in Figure 3 belong to several clusters. The first sequences were taken from the phylogenetic representation of the aflatoxin cluster. The remaining sequences belong to *A. flavus* and its matching *A. oryzae* sequences. A few more sequences were added for each cluster so that they would properly represent the cluster. In the HexA group of sequences, the *Saccharomyces cerevisiae* sequence for primary metabolism was added as a reference. On the other hand, in the Pks group, the sequence belonging to *A. parasiticus,* whose function has yet to be determined, was added because of the singular localization of the ACP domain, which appears much farther to the C-terminal side of the sequence than the ACP of the other sequences of the multialignment. This domain is located on opposite sides in the two protein sequences. One of the motives we located that is maintained by the bacterial and fungal organisms is the prosite detected for KS (prosite: PS00606, Beta-ketoacyl synthase active site). This domain catalyzes the condensation of malonyl-ACP with the growing fatty acid chain and its active site is on a Cys. Of the whole prosite, only the active site and the amino acid one position upstream from it (Ala) are conserved. The remaining amino acids differ mostly when the proteins are hypothetical. Figure 4 shows a similar analysis for the AT domain. This domain in Pks is in a central region while in the HexB it is in the C-terminal region. In this case, there were not as many regions located which were conserved in HexB, Pks and the homologous bacterial sequences. However, quite a large region of conserved amino acids surrounds the malonyl binding site and it is conserved in all the sequences we have studied. Curiously, this conserved region is larger than the one found in the KS prosite.



```
P19097 Saccharomyces cerevisia             GPIKTPVGACATSV 1309 // TATDKIGRSVPAPGKGILTTAR 1430 // GVASFHGTSTKANDKNES 1554

Q5VDA2/AFSG010418 Aspergillus flavus       GPIKTPVGACATGV 1117 // MAADKIGSSVPAPGQGILSFSR 1238 // DVASLHGTSTRGNDLNEP 1325
Q2UG03 Aspergillus oryzae                  GPIKTPVGACATGV 1125 // MAADKIGSSVPAPGQGILSFSR 1246 // DVASLHGTSTRGNDLNEP 1333
Q8TGA2 Aspergillus parasiticus             GPIKTPVGACATGV 1117 // MAADKIGSSVPAPGQGILSFSR 1238 // DVASLHGTSTRGNDLNEP 1325
Q5VD77 Aspergillus nomius                  GPIKTPVGACATGV 1118 // MAADKIGSSVPAPGQGILSFSR 1239 // DVASLHGTSTRGNDLNEP 1326

AFSG001384 Aspergillus flavus              GPIKTPVGACATAV 1280 // TATDKIGRSVPAPGQGVLTTAR 1401 // GVASFHGTSTVANDKNES 1528
Q2U734 Aspergillus oryzae                  GPIKTPVGACATAV 1280 // TATDKIGRSVPAPGQGVLTTAR 1401 // GVASFHGTSTVANDKNES 1528
Q0CPX5 Aspergillus terreus                 GPIKTPVGACATAV 1283 // TATDKIGRSVPAPGQGVLTTAR 1404 // GVASFHGTSTVANDKNES 1530

AFSG002450 Aspergillus flavus              GPIRTSVGACATSI 1194 // TASDKTSRSVPAPGQGILTNAR 1315 // DFVSLHGTSTVMNDKNEA 1440
Q2U1F7 Aspergillus oryzae                  GPIRTSVGACATSI 1194 // TASDKTSRSVPAPGQGILTNAR 1315 // DFVSLHGTSTVMNDKNEA 1440

AFSG005155 Aspergillus flavus              GPIKSPSGTCATSV 1187 // MAGDKVGRSVPAPGQGVLTAAR 1308 // QVASFHGTSTKANDKNES 1426
Q2TXF0 Aspergillus oryzae                  GPIKSPSGTCATSV 1197 // MAGDKVGRSVPAPGQGVLTAAR 1318 // QVASFHGTSTKANDKNES 1436
```

```
Q5VDF2/AFSG010422 Aspergillus flavus       GPSYTNDTACSSSL (40) LSRT-GNCKPYDDKADGYCRA 607 // NMTAALNTTGLHPNDFSYIEMHGTGT 682
Q71MJ1 Aspergillus Sp                      GPSYTNDTACSSSL (40) LSRT-GNCKPYDDKADGYCRA 607 // NMTAALNTTGLHPNDFSYIEMHGTGT 682
Q2PH11 Aspergillus oryzae                  GPSYTNDTACSSSL (40) LSRT-GNCKPYDDKADGYCRA 607 // NMTAALNTTGLHPNDFSYIEMHGTGT 682
Q12053 Aspergillus parasiticus             GPSYTNDTACSSSL (40) LSRT-GNCKPYDDKADGYCRA 607 // NMTAALNTTGLHPNDFSYIEMHGTGT 682
Q5VD79 Aspergillus nomius                  GPSYTNDTACSSSL (40) LSRT-GNCKPYDDNADGYCRA 607 // NMTAVLNTTGLHPNDFSYIEMHGTGT 682

AFSG005452 Aspergillus flavus              GPSITVDTACSSSL (40) LSST-GNCKTFDDDADGYCRA 583 // LFRKLLAESGIHPHEISYIEMHGTGT 658
Q2TXQ8 Aspergillus oryzae                  GPSITVDTACSSSL (40) LSST-GNCKTFDDDADGYCRA 622 // LFRKLLAESGIHPHEISYIEMHGTGT 697
Q9P855 Gibberella fujikuroi                GPSIAVDTACSSSL (40) LSRT-GNCKAFNDGADGYCRA 578 // LFKKLLNETGIHPHDVSYVEMHGTGT 653

AFSG011713 Aspergillus flavus              GPSVSVDTACSSSL (40) LSRT-GNCNTFDDGADGYCRA 612 // IFNKLLNDANIDPKDVSYVEMHGTGT 687
Q2UA48 Aspergillus oryzae                  GPSVSVDTACSSSL (40) LSRT-GNCNTFDDGADGYCRA 612 // IFNKLLNDANIDPKDVSYVEMHGTGT 687
Q03149 Emericella nidulans                 GPSVSVDTACSSSL (40) LSRT-GNCNTFDDGADGYCRA 612 // IFKKLLNEANVDPKNISYIEMHGTGT 687

AFSG004161 Aspergillus flavus              GPSYSVDTACSSGL (40) LSRT-GNCKTFDDDADGYCRG 181 // IFEKVLTSAGVDPYSVGYVEMHGTGT 256
Q2UR58 Aspergillus oryzae                  GPSYSVDTACSSGL (40) LSRT-GNCKTFDDDADGYCRG 620 // IFEKVLTSAGVDPYSVGYVEMHGTGT 695

P78584 Aspergillus parasiticus             GPSTAVDAACASSL (40) LSSD-GRCHSFDDTASGYGRG 247 // VARKALDVARVDASTIDYVEAHATST 322
```

Figure 3: Extract of the HexA and Pks multialigments from selected sequences corresponding to the predicted KS domain (Pfam references denoted). Diagrams of sequence domain organization are shown. Highlighted in blue are the residues that are strictly conserved in this fragment of the fungal multialigment. Highlighted in yellow are the residues that are also conserved in bacterial sequences. Protein sequences are named by their UniProt code and species name, gaps are denoted by a dash. Numbers on the right side represent the residue position on the N-terminus. Numbers between brackets denote the length of the omitted fragment.

Aspergillus species

Aspergillus species are closely related fungi which, despite their similarities, also have important differences. One very clear example of this is their capacity to synthesise aflatoxins **[2, 3]**. A few Aspergillus species can synthesise aflatoxins using the aflatoxin biosynthetic cluster (*A. flavus*, *A. parasiticus*, *A. nomius*). Other organisms cannot catalyze the final steps of the aflatoxin pathway and produce instead one of its intermediate metabolits (*E. nidulans*). Despite having the complete aflatoxin cluster, some species cannot synthesise this mycotoxin (*A. oryzae*). Finally, some species do not even have the aflatoxin cluster (*A. fumigatus*). Of all these species only the genome of *A. oryzae*, *E. nidulans* and *A. fumigatus* is completely sequenced while the whole genome sequencing project for Aspergillus flavus, funded by the USDA/NRI Microbial Genome Sequencing Project and the USDA/ARS, is complete. Therefore, in order to obtain information about the sequences extracted from the *A. flavus* genome project, we will mostly rely on the information provided by their three organisms.

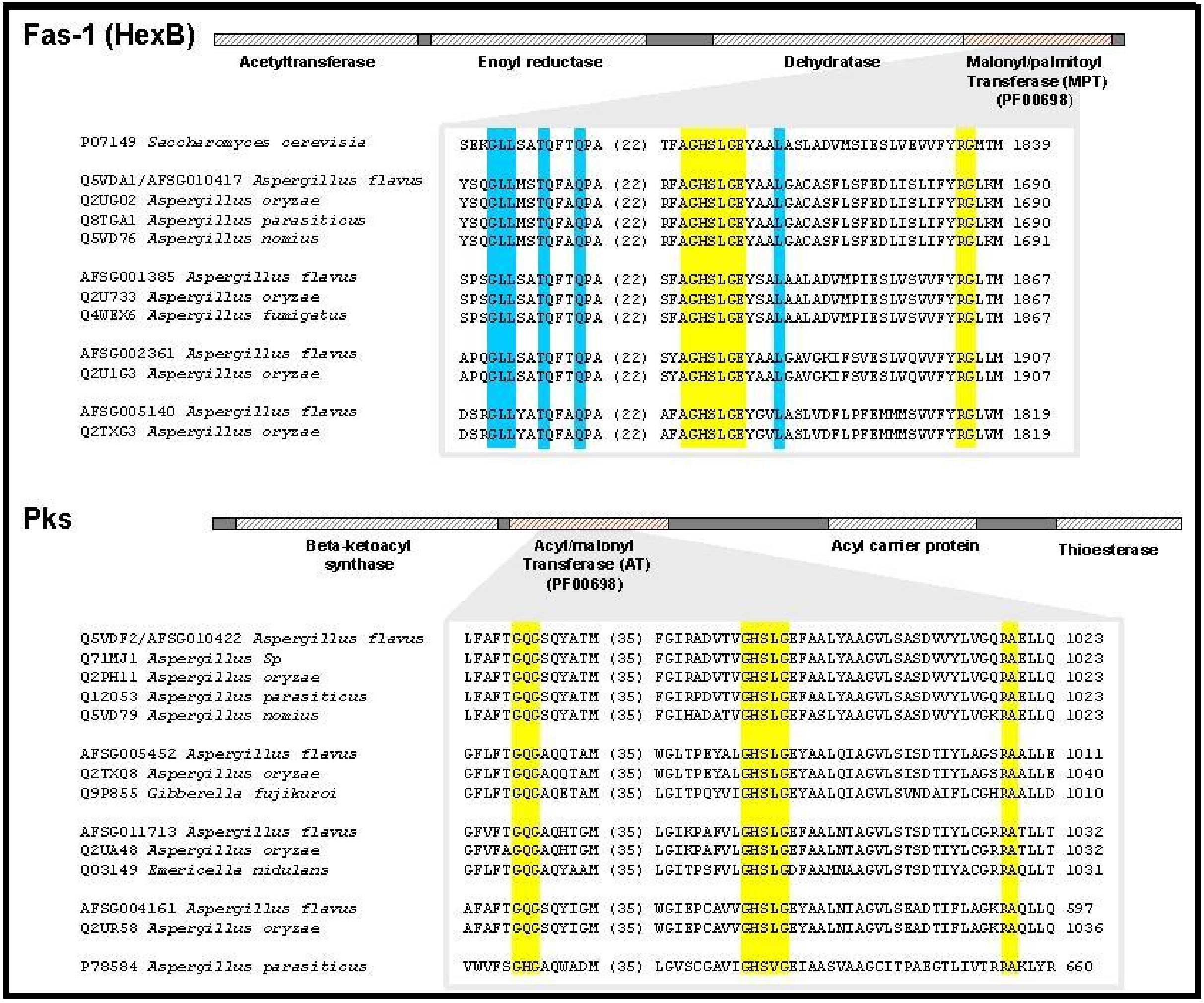

Fas-1 (HexB)
Acetyltransferase
Enoyl reductase
Dehydratase
Malonyl/palmitoyl Transferase (MPT) (PF00698)
P07149 Saccharomyces cerevisia SEKGLLSATQFTQPA (22) TFAGHSLGEYAALASLADVMSIESLVEVVFYRGMTM 1839
Q5VDA1/AFSG010417 Aspergillus flavus YSQGLLMSTQFAQPA (22) RFAGHSLGEYAALGACASFLSFEDLISLIFYRGLKM 1690
Q2UG02 Aspergillus oryzae YSQGLLMSTQFAQPA (22) RFAGHSLGEYAALGACASFLSFEDLISLIFYRGLKM 1690
Q8TGA1 Aspergillus parasiticus YSQGLLMSTQFAQPA (22) RFAGHSLGEYAALGACASFLSFEDLISLIFYRGLKM 1690
Q5VD76 Aspergillus nomius YSQGLLMSTQFAQPA (22) RFAGHSLGEYAALGACASFLSFEDLISLIFYRGLKM 1691
AFSG001385 Aspergillus flavus SPSGLLSATQFTQPA (22) SFAGHSLGEYSALAALADVMPIESLVSVVFYRGLTM 1867
Q2U733 Aspergillus oryzae SPSGLLSATQFTQPA (22) SFAGHSLGEYSALAALADVMPIESLVSVVFYRGLTM 1867
Q4WEX6 Aspergillus fumigatus SPSGLLSATQFTQPA (22) SFAGHSLGEYSALAALADVMPIESLVSVVFYRGLTM 1867
AFSG002361 Aspergillus flavus APQGLLSATQFTQPA (22) SYAGHSLGEYAALGAVGKIFSVESLVQVVFYRGLLM 1907
Q2U1G3 Aspergillus oryzae APQGLLSATQFTQPA (22) SYAGHSLGEYAALGAVGKIFSVESLVQVVFYRGLLM 1907
AFSG005140 Aspergillus flavus DSRGLLYATQFAQPA (22) AFAGHSLGEYGVLASLVDFLPFEMMMSVVFYRGLVM 1819
Q2TXG3 Aspergillus oryzae DSRGLLYATQFAQPA (22) AFAGHSLGEYGVLASLVDFLPFEMMMSVVFYRGLVM 1819
Pks
Beta-ketoacyl synthase
Acyl/malonyl Transferase (AT) (PF00698)
Acyl carrier protein
Thioesterase
Q5VDF2/AFSG010422 Aspergillus flavus LFAFTGQGSQYATM (35) FGIRADVTVGHSLGEFAALYAAGVLSASDVVYLVGQRAELLQ 1023
Q71MJ1 Aspergillus Sp LFAFTGQGSQYATM (35) FGIRADVTVGHSLGEFAALYAAGVLSASDVVYLVGQRAELLQ 1023
Q2PH11 Aspergillus oryzae LFAFTGQGSQYATM (35) FGIRADVTVGHSLGEFAALYAAGVLSASDVVYLVGQRAELLQ 1023
Q12053 Aspergillus parasiticus LFAFTGQGSQYATM (35) FGIRPDVTVGHSLGEFAALYAAGVLSASDVVYLVGQRAELLQ 1023
Q5VD79 Aspergillus nomius LFAFTGQGSQYATM (35) FGIHADATVGHSLGEFASLYAAGVLSASDVVYLVGKRAELLQ 1023
AFSG005452 Aspergillus flavus GFLFTGQGAQQTAM (35) WGLTPEYALGHSLGEYAALQIAGVLSISDTIYLAGSRAALLE 1011
Q2TXQ8 Aspergillus oryzae GFLFTGQGAQQTAM (35) WGLTPEYALGHSLGEYAALQIAGVLSISDTIYLAGSRAALLE 1040
Q9P855 Gibberella fujikuroi GFLFTGQGAQETAM (35) LGITPQYVIGHSLGEYAALQIAGVLSVNDAIFLCGHRAALLD 1010
AFSG011713 Aspergillus flavus GFVFTGQGAQHTGM (35) LGIKPAFVLGHSLGEFAALNTAGVLSTSDTIYLCGRRATLLT 1032
Q2UA48 Aspergillus oryzae GFVFAGQGAQHTGM (35) LGIKPAFVLGHSLGEFAALNTAGVLSTSDTIYLCGRRATLLT 1032
Q03149 Emericella nidulans GFLFTGQGAQYAAM (35) LGITPSFVLGHSLGDFAAMNAAGVLSTSDTIYACGRRAQLLT 1031
AFSG004161 Aspergillus flavus AFAFTGQGSQYIGM (35) WGIEPCAVVGHSLGEYAALNIAGVLSEADTIFLAGKRAQLLQ 597
Q2UR58 Aspergillus oryzae AFAFTGQGSQYIGM (35) WGIEPCAVVGHSLGEYAALNIAGVLSEADTIFLAGKRAQLLQ 1036
P78584 Aspergillus parasiticus VWVFSGHGAQWADM (35) LGVSCGAVIGHSVGEIAASVAAGCITPAEGTLIVTRRAKLYR 660

Figure 4: Extract of the HexB and Pks multialigments from selected sequences corresponding to the predicted MPT and AT domains (Pfam references are denoted). Diagrams of sequence domain organization are shown. Highlighted in blue are the residues that are strictly conserved in this fragment of the fungal multialigment. Highlighted in yellow are the residues that are also conserved in bacterial sequences. Protein sequences are named by their UniProt code and species name. Numbers on the right side represent the residue position on the N-terminus. Numbers between brackets denote the length of the omitted fragments.

In this study, we found that for the HexA and HexB proteins, five homologues of the protein that belongs to the aflatoxin gene cluster were found in *A. oryzae*; four in both *E. nidulans* and *A. flavus;* and only one in *A. fumigatus*. Of the four homologues in *A. flavus*, the one resulting from fatty acid synthesis was found in the three other organisms as it is an essential part of fungal primary metabolism. Another homologue belongs to the aflatoxin gene cluster. The similarity between *A. flavus* and *A. oryzae* can be seen by comparing the proteins belonging to this cluster. Despite being unable to produce aflatoxins, the proteins from the *A. oryzae* cluster are closer to *A. flavus* than any other aflatoxin-producing organism. While *E. nidulans* has a sterigmatocystin homologue within the selected sequences, *A. fumigatus* has no other homologue to the fatty acid synthase proteins, which shows that it lacks any secondary metabolism related to aflatoxin production. The function of the remaining proteins remains unknown but since *A. fumigatus* has none of these homologues we believe they may be related to some kind of undiscovered secondary metabolism. As far as the Pks proteins are concerned, *A. oryzae* and *A. nidulans* both have seven homologues while *A. fumigatus* and *A. flavus* have only four.

Acknowledgment:
We thank John Bates of the Language Service of the Rovira i Virgili University for his help with writing the manuscript. This study was funded by the "Ministerio de Educación y Ciencia", project reference AGL2007-65678.

References:

[01] J. Yu *et al., Environ Microbiol.,* 70: 1253 (2004) [PMID: 15006741]
[02] G. A. Payne and M. P. Brown, *Annu Rev Phytopathol.,* 36: 329 (1998) [PMID: 15012504]
[03] N. P. Keller *et al., Nat Rev Microbiol.*, 3: 937 (2005) [PMID: 16322742]

[04] K. Yabe and H. Nakajima, *Appl Microbiol Biotechnol*., 64: 745 (2004) [PMID: 15022028]

[05] I. Carbone *et al., BMC Evol Biol.,* 9: 111 (2007) [PMID: 17620135]

[06] C. M. Watanabe and C. A. Townsend, *Chem Biol.,* 9: 981 (2002) [PMID: 12323372]

[07] S. W. White *et al., Annu Rev Biochem*., 74: 791 (2005) [PMID: 15952903]

[08] S. Jenni *et al., Science,* 311: 1263 (2006) [PMID: 16513976]

[09] T. Maier *et al., Science,* 311: 1258 (2006) [PMID: 16513975]

[10] J. M. Crawford *et al., Proc Natl Acad Sci USA.*, 103: 16728 (2006) [PMID: 17071746]

[11] S. F. Altschul *et al., Nucleic Acids Res*., 25: 3389 (1997) [PMID: 9254694]

[12] J. D. Thompson *et al., Nucleic Acids Res*., 22: 4673 (1994) [PMID: 7984417]

[13] S. Kumar *et al., Brief Bioinform*., 5: 150 (2004) [PMID: 15260895]

[14] D. M. Geiser *et al., Fungal Genet Biol*., 31: 169 (2000) [PMID: 11273679]

[15] S. Kroken *et al., Proc Natl Acad Sci USA*., 100: 15670 (2003) [PMID: 14676319]

Edited by P. Kangueane
Citation: Marcet-Houben *et al.*, Bioinformation 3(1): 33-40 (2008)