FUNGIpath: a new tool for analysing the evolution of fungal metabolic pathways

Sandrine Grossetête, Bernard Labedan, Olivier Lespinet

Institut de Génétique et de Microbiologie, UMR 8621,
Université Paris Sud, Bâtiment 400, 91405 Orsay Cedex

{sandrine.grossetete,bernard.labedan,olivier.lespinet}@igmors.u-psud.fr

Abstract: *FUNGIpath is a new tool dedicated to perform in-depth analysis of fungal metabolic pathways. It is freely accessible at http://www.fungipath.u-psud.fr. FUNGIpath consists in a collection of orthologous groups of proteins that have been predicted using complementary methods of detection and further mapped on KEGG and MetaCyc pathways. It allows an easy comparison of the primary and secondary metabolisms afforded by the different fungal species present in the database with the possibility to assess the level of specificity of various pathways at different taxonomic distances. As more and more fungal genomes are expected to be decrypted in the next years, this tool is expected to help to progressively reconstruct what were the primary and secondary metabolisms of the ancestors of the main branches of the fungi tree and to understand how these ancestral fungal metabolisms evolved to various specific derived metabolisms.*

Keywords: Metabolism, evolution, fungi.

1 Introduction

Fungi constitute one of the eukaryotic taxonomic group that present today (April 2009) the highest number of species for which the complete sequence of the nuclear genome has been published and is available to the scientific community (26 genomes according to [1]). This relative abundance is mainly due to their moderate genome size, and to the fact that several species have been model organisms for fundamental, medical, or agronomical and industrial studies (e.g. *Saccharomyces cerevisiae, Candida albicans, Yarrowia lipolytica*).

Therefore, fungal genomes appear today to be a suitable material for large-scale comparative studies. Indeed, several teams have already performed extensive comparison of a few fungal genomes to predict clusters of orthologous groups of proteins, which can be accessed by tools such as OrthoDB [2] or e-Fungi [3]. Such approaches open the way to study the evolution of fungal genomes [4].

However, information about the metabolism of fungi is presently rather scarce and heterogeneous in major public databases. Although we found a moderate or low amount of data in dedicated databases such as KEGG [6] or MetaCyc [7], there are almost no data on fungi metabolism in Swiss-Prot [5] with the noticeable exception of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (data not shown). In addition, beside a preliminary attempt to identify enzymes in pathogenic fungi for a limited number of metabolic pathways [8], there is presently no tool allowing performing large-scale analysis of fungal metabolism.

Here, we describe FUNGIpath that is, to our knowledge, the first tool allowing to mining genomic data in order to perform in-depth analysis of fungi metabolism. This new tool presents two efficient features: it is based on several complementary approaches combining to define reliable groups of orthologous genes and it allows mapping these groups on every pathway that are available in the KEGG [6] and MetaCyc [7]

databases.

2 Organizing Metabolic Data by Comparing Fungal Genomes

Primary (sequences and pathways) and secondary data (orthologous group) were assembled in a database that is made freely available to the community through FUNGIpath, a user-friendly website implemented in PHP, HTML and Javascript.

2.1 Primary Data

Sequences data are summarized in two tables, one describing genome informations and the other one listing the amino acid sequences encoded by the genomes of 20 fungal species: *Aspergillus nidulans*, *Aspergillus oryzae*, *Batrachochytrium dendrobatidis*, *Chaetomium globosum*, *Coprinus cinereus*, *Fusarium oxysporum*, *Laccaria bicolor*, *Magnaporthe bicolor*, *Mycospherella graminicolla*, *Neurospora crassa*, *Phycomyces blakesleeanus*, *Podospora anserina*, *Puccinia graminis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Sclerotinia sclerotiorum*, *Stagonospora nodorum*, *Trichoderma reesei*, *Ustilago maydis*, *Yarrowia lipolytica.*

The complete list of genome sources and url files are available on supplementary data table 1. For each genome, we removed sequences that are 100% identical. Supplementary data Table 1 provides sources for the respective genomic data used in FUNGIpath.

Metabolism data downloaded from either KEGG (6) or MetaCyc (7) and enriched with predicted annotations are organized in several tables to speed up data access time.

2.2 Predicting Orthologs

Different methods have been published to predict orthologs but none of them appears completely reliable since they poorly overlap (supplementary data, table 2). Thus, we found necessary to use independent methods to collect as many potential orthologs as possible. Moreover, exploring several methods raised the probability to have a consistent group, corresponding to their overlapping. Accordingly, we are using three different and complementary approaches based on similarity searches and another one based on the analysis of phylogenetic trees.

First, we adapted two methods already published with their respective default parameters: OrthoMCL [9] allows defining consistent groups of orthologs that are strongly related. Inparanoid [10] permits to differentiate orthologs and inparalogs (genes recently duplicated after the last speciation event) in pairwise comparison of all genomes. Moreover, the classical Best Reciprocal Hits (BRH) approach has been entirely automated by a Perl script. To improve the definition of orthologs we filtered the BLAST [11] results by specifying two parameters, the alignment percent and the score ratio. Dividing the alignment length of each aligned sequence by its total length permits to avoid local conservation. The score ratio is computed by dividing the crude BLAST score obtained when aligning sequence 1 against sequence 2 by maximum BLAST score, i.e. BLAST score obtained when sequence 1 is aligned against itself. We keep only results with score ratio superior to 0.2 and alignment percent superior to 60%.

These different methods based on sequence similarity allow to get more or less stringent clusters of orthologous genes depending if we used single (e.g. Inparanoid) or multiple (BRH) linkage.

Beside these methods based on similarity approaches, we also used a phylogeny approach to get orthologous groups using the automatic tree analysis previously developed by Lemoine et al. [12]. We first build families of homologous protein detected by BLASTP [11] with the following requirements: an E-value less than 0.001 and an alignment percent larger than 70 % of the length of the shorter sequence of the aligned pair. For each family, a multiple alignment was built with Muscle [13], and the deduced phylogenetic tree was reconstructed with PhyML [14]. The program Retree from Phylip package [15] was further used to root the tree in order to distinguish orthologs and paralogs with the automatic tree analysis [12].

Once we got the orthologous groups, we compared groups obtained by the different methods and merged groups that overlap (supplementary data Table 2). To help the user to evaluate the reliability of the predictions we computed a confidence score based on the number of methods that found independently the same group. We suppose that a group found by several methods is more reliable than a group found only by one of the available methods (see supplementary data for score computation). We obtained orthologous group with group size ranging from 2 to 2694 sequences. As the homogeneity of the largest groups is most propably doubtful, we kept only orthologous groups with a score superior to 1.5. Such a threshold value, helped to limit the group size to a maximum of 400 sequences.

2.3 Reconstructing Pathways

Annotating the putative enzymatic activities

Once the orthologous groups have been defined, we attemped to predict functionnal annotation by using an HMM approach. For each orthologous group, we built the corresponding HMM profile with hmmbuild [16] and used hmmsearch [16] to search its similarity with sequences that display an enzymatic annotation in Swiss-Prot [5]. The annotation was transferred to the orthologous group analyzed, if the E-value of the best hit obtained is lower or equal to 10^{-80}.

This annotation assigned 843 different EC numbers to 1261 groups of orthologous proteins. Among these groups, our HMM approach contribute to annotate 360 groups (29%) which the belonging sequences did not get any annotation in the Swiss-Prot database [5]. Nearly one half (396) of the EC numbers is present in all genomes, and 90% (764) of the assigned EC numbers are found in at least 50% of the genomes.

Assembling the putative EC numbers in pathways

Once the different putative orthologs have been annotated as described above, we used them to exhaustively reconstruct the different metabolic pathways in fungi. To do that, we used two reliable public databases, KEGG [6] and MetaCyc [7] that differ in their way to define pathways.

Useful information was extracted from the reaction file generated by KEGG [6] and the corresponding GIF maps were downloaded. BIOPAX files defined in MetaCyc [7] were downloaded and we generated automatically the corresponding map pictures by directed graph building. Accordingly, we collected 151 pathways in KEGG and 1143 pathways in MetaCyc that define mainly anabolic and catabolic ways.

3 Querying FUNGIpath and exploring pathways

FUNGIpath (http://www.fungipath.u-psud.fr) has been designed to allow studying fungi metabolism by performing various queries on our database.

FUNGIpath allows checking and visualizing the conservation of pathways between different fungi. We can make such a search using several ways: one can start from a defined EC number (Fig. 1), from a known pathway (Fig. 2), or by using a user-defined pathway delineated in a simplified BIOPAX format (data not shown).

Searching a specific EC number allows assessing the level of conservation of this EC number in each taxonomic group and to directly access to all the pathways in which this EC number is involved (Fig. 1).

For instance, Fig. 1 shows that EC 3.5.1.4 corresponds to an amidase (Acymlamide amidohydrolase that cleaves carbon–nitrogen bonds in amides) that is very well conserved in all fungi and is involved in at least 6 different pathways in both KEGG and MetaCyc databases. Moreover, we found that this EC number has been assigned to several distinct orthologous groups (data not shown).

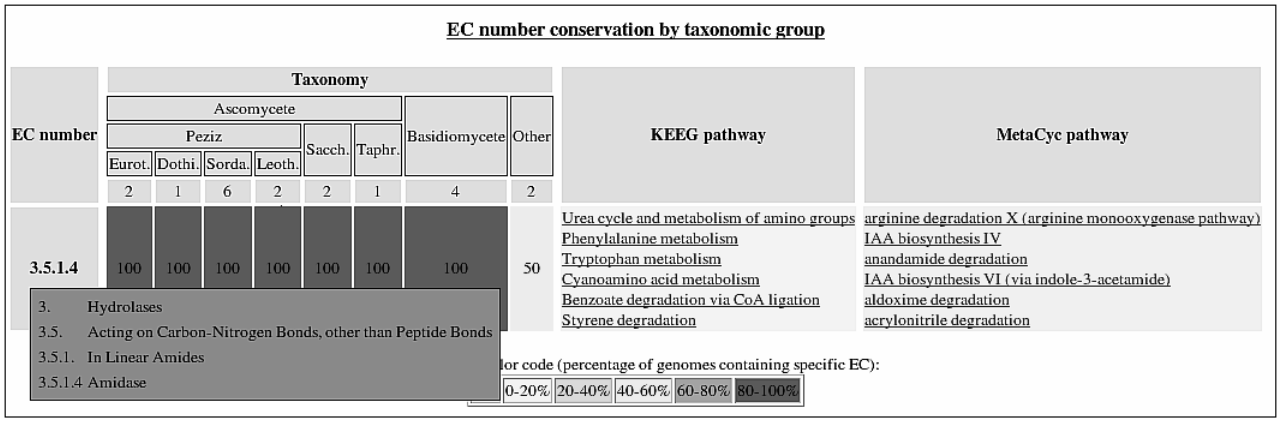

Figure 1. Exploring pathways using a specific EC number (3.5.1.4). The level of occurrence in the different species belonging to the different taxonomic groups of fungi is indicated with a color code (scale from white to red) in the taxonomy column. The respective lists of the pathways that contain the requested EC number are indicated in the KEGG and MetaCyc columns, respectively. Note that the pathway names are different in both databases.

Fig. 2 illustrate the case of the 'Biotin metabolism' (KEGG database), the results displayed when a pathway is queried.

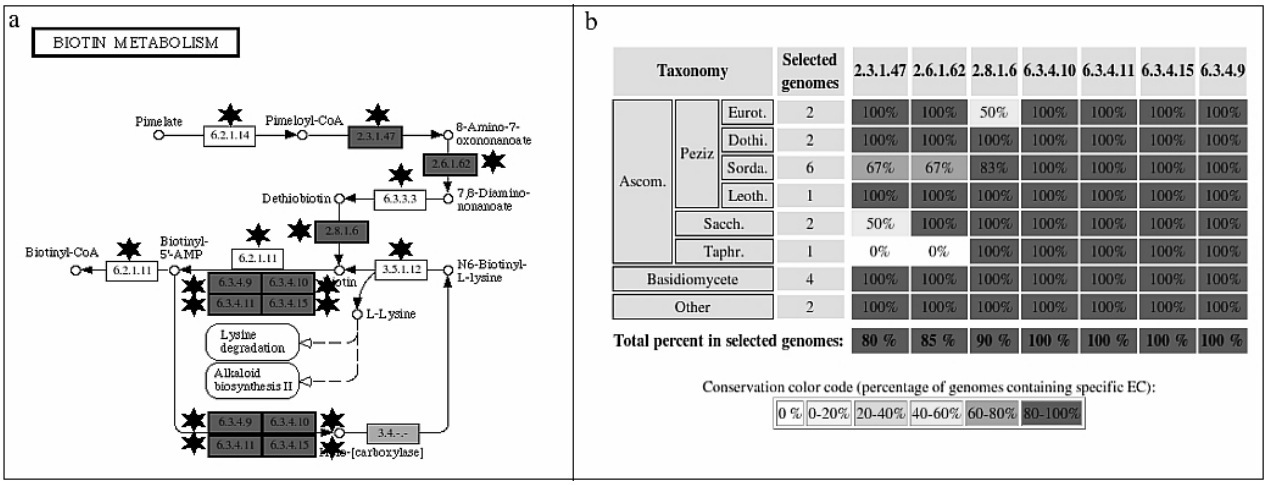

Figure 2. Exploring pathways using a specific pathway name. **a:** The 'Biotin metabolism' pathway is displayed using the KEGG map. For each EC a color code indicate the level of conservation. The black star indicates the EC numbers that are specific to the 'Biotin metabolism' **b:** The table lists the percentage of conservation of this pathway with the same color code (white to red) in each species belonging to the different taxonomic groups of fungi. Moreover, the relative presence in all 20 fungi is given in the last line of this table, this value defining the color used in the KEGG map, in. 2a.

To facilitate the analysis of these results, a color code has been associated to the conservation level of EC numbers. Results are presented both as a KEGG gif map (Fig. 2a) and summarized in a table (Fig. 2b). For each EC number the corresponding orthologous groups of proteins can easily be accessed by using the genomes feature table (not shown) and can be downloaded for further studies.

4 Discussion

4.1 Improving functional annotation of fungi genomes

To challenge the validity of our predictions of functional annotation, we compared as a control all our EC numbers predictions for the yeast *S.cerevisiae* with 4 different curated public databases (KEGG [6], MetaCyc [7], Swiss-Prot [8], and SGD [17]). The results are displayed in Table 1. We call ID-EC the unique

pair formed between an ID and its EC (or one of its EC). Thus, an EC can belong to several pairs, if several IDs have the same annotation. Likewise, an ID can be present in several pairs, if the ID gets several EC numbers in case of multifunctional proteins.

Database	Total number of ID-EC in FUNGIpath	Number (percent) of identical predicted ID-EC	Number of different ID-EC	Digit position which is different			
				1st d.	2nd d.	3rd d.	4th d.
KEGG	1062	843 (79,4%)	47	2	2	7	36
METACYC	523	409 (78,2%)	68	16	3	7	42
SGD	512	408 (79,7%)	41	9	2	3	27
SP	1114	1030 (92,5%)	40	2	1	5	32
Prediction	1299						

Table 1. Comparison of enzymatic data for *S. cerevisiae* between 4 databases and our prediction

According to our predictions made from orthologous groups we get 1299 ID-EC for the yeast. The observed high overlapping with Swiss-Prot figures is not surprising since we used mainly this database for our prediction. On the other hand, our prediction corresponds to nearly 80% of the EC numbers found in the three other databases. In addition, most of the IDs pairing with different EC numbers diverge only at the level of the last digit. Thus, the reliability of the predictions of our group of orthologous proteins and of their enzyme function appears to be comparable with that of the independently curated public databases.

Moreover, we compared the functional annotations for 12 species between KEGG [6] and FUNGIpath (Supplementary data Table 3). More than 75% of KEGG data are found in FUNGIpath. However, contrarily to KEGG, we work only with complete EC numbers. Thus, the comparison of the incomplete KEGG EC numbers with our data allows to recover some of them. For instance, in the case of *Schizosaccharomyces pombe*, the total number of annotated enzymes are very close (1267 EC numbers in KEGG and 1231 ones in FUNGIpath), but among the 258 EC numbers that are incomplete in KEGG, 62 have been completed in FUNGIpath. Thus, FUNGIpath is a efficient tool for functional annotation of fungi and thus for studying their metabolism.

4.2 Studying annotation and evolution of metabolism in fungi

One of the main problems encountered when trying to reconstruct entire pathways from orthology data is the occurrence of missing data. The absence of an EC number (*orphan metabolic activities* [18]), may be due to a too low percent identity of the corresponding amino acid sequence or to its replacement by another protein. Alternatively, the simultaneous absence of several EC numbers that belong to a specific pathway most likely suggests that this entire pathway is not present in the studied species. However, one cannot dismiss the hypotheses that this absence is simply due to a major annotation problem or to the replacement of this pathway by an alternate one.

For instance, Fig. 2 shows that most of the EC numbers involved in the 'Biotin metabolism' defined by KEGG [6], are found in our database and appear specific to this pathway. However, this Biotin metabolism pathway appears to be incomplete in many fungi since several of its specific enzymatic activities are not found such as 3.5.1.12, 6.3.3.3 and 6.2.1.11 (white EC with black stars in Fig. 2a). We can suppose that either these EC numbers exist in fungi but they are presently not detectable, or fungi use other EC numbers to catalyse these reactions. Of the 11 EC numbers involved in the 'Biotin metabolism', four (6.3.4.9, 6.3.4.10, 6.3.4.11, 6.3.4.15) are found in all the selected species and one (2.8.1.6) in all species excepted one genome (*A.orizae*). Two EC numbers (2.3.1.47, 2.6.1.62) are absent in several genomes (*M. grisea, N. crassa, S. pombe* plus *S. cerevisiae* for 2.3.1.47). These genomes may use an alternative way to realize the first steps of the 'Biotin metabolism'.

5. Conclusion

FUNGIpath appears a reliable tool that helps to analyse the metabolism of fungi. It will be especially

useful to annotate newly-sequenced genomes.

Moreover, FUNGIpath allows an easy comparison of the respective metabolisms afforded by the different taxons. For instance, 163 EC numbers are found uniquely in Ascomycetes (data not shown) and may help to delineate the metabolic specificities of the common ancestor to this group.

As more and more genomes are expected to be decrypted in the next years, FUNGIpath we will be especially usefull to progressively reconstruct what were the primary and secondary metabolisms of the ancestors of the main branches of the fungi tree and to understand how these ancestral fungal metabolisms evolved to various specific derived metabolisms.

Acknowledgements

We are grateful to Philippe Silar for his help during the designing process of the web site and his helpful comments about this work. The computations were performed on the MIGALE platform (INRA, Jouy-en-Josas, France). SG is a PhD student supported by a doctorant CNRS fellowship.

References

[1] K. Liolios, K. Mavrommatis, N. Tavernarakis, N.C. Kyrpides, The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res., 36, D475-479, 2007.

[2] E.V. Kriventseva, N. Rahman, O. Espinosa, E.M. Zdobnov, OrthoDB: the hierarchical catalog of eukaryotic orthologs. Nucleic Acids Res., 36, D271-275, 2008.

[3] C. Hedeler, H.M. Wong, M.J. Cornell, I. Alam, D.M. Soanes, M. Rattray, S.J. Hubbard, N.J. Talbot, S.G. Oliver, N.W. Paton, e-Fungi: a data resource for comparative analysis of fungal genomes. BMC Genomics, 8:426, 2007.

[4] D.M. Soanes, I. Alam, M. Cornell, H.M.Wong, C. Hedeler, N.W. Paton, M. Rattray, S.J. Hubbard, S.G. Oliver, N.J. Talbot, Comparative Genome Analysis of Filamentous Fungi Reveals Gene Family Expansions Associated with Fungal Pathogenesis. PLoS ONE, 3, 6:e2300, 2008.

[5] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivany, R.D. Appel, A. Bairoch, Nucleic Acids Res., 31:3784-3788, 2003.

[6] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res., 27:29-34, 1999.

[7] R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S.Y. Rhee, A.G. Shearer, C. Tissier, T.C. Walk, P. Zhang, P.D. Karp, The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res., 36:623-31, 2008.

[8] P.F. Giles, D.M. Soanes, N.J. Talbot, A relational database for the discovery of genes encoding amino acid biosynthetic enzymes in pathogenic fungi. Comp Funct Genomics. 4(1):4-15, 2003.

[9] L. Li, CJ. Jr. Stoeckert, D.S. Roos, OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res., 13, 2178-89, 2003.

[10] M. Remm, C.E. Storm, E.L. Sonnhammer, Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol., 314, 1041-52, 2001.

[11] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. J Mol Biol., 215:403-10, 1990.

[12] F. Lemoine, O. Lespinet, B. Labedan, Assessing the evolutionary rate of positional orthologous genes in prokaryotes using synteny data. BMC Evol Biol., 7, 237, 2007.

[13] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res., 32:1792-7, 2004.

[14] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol., 52:696-704, 2003.

[15] J. Felsenstein, PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, 2005.

[16] SR. Eddy, Hidden Markov models. Curr Opin Struct Biol. 6(3):361-5, 1996.

[17] JM. Cherry, C. Ball, S. Weng, G. Juvik, R. Schmidt, C. Adler, B. Dunn, S. Dwight, L. Riles, RK. Mortimer, D. Botstein, Genetic and physical maps of Saccharomyces cerevisiae. Nature 387(6632 Suppl):67-73,1997.

[18] L. Chen, D. Vitkup, Predicting genes for orphan metabolic activities using phylogenetic profiles, Genome Biol., 7:R17, 2006.

Supplementary data

Genome Name	Source	url
Aspergillus nidulans	Broad Institute	http://www.broad.mit.edu/annotation/genome/aspergillus_nid…
Aspergillus oryzae	NITE	ftp://ftp.bio.nite.go.jp/pub/ao/
Batrachochytrium dendrobatidis	Broad Institute	http://www.broad.mit.edu/annotation/genome/batrachochytriu…
Chaetomium globosum	Broad Institute	http://www.broad.mit.edu/annotation/genome/chaetomium_gl…
Coprinus cinereus	Broad Institute	http://www.broad.mit.edu/annotation/genome/coprinus_cinere…
Fusarium graminearum	Broad Institute	http://www.broad.mit.edu/annotation/genome/fusarium_grami…
Laccaria bicolor	JGI	ftp://ftp.jgi-psf.org/pub/JGI_data/Laccaria_bicolor/
Magnaporthe grisea	Broad Institute	http://www.broad.mit.edu/annotation/genome/magnaporthe_g…
Mycosphaerella graminicolla	JGI	ftp://ftp.jgi-psf.org/pub/JGI_data/Mycosphaerella_graminicola…
Neurospora crassa	Broad Institute	http://www.broad.mit.edu/annotation/genome/neurospora/Do…
Phycomyces blakesleeanus	JGI	ftp://ftp.jgi-psf.org/pub/JGI_data/Phycomyces_blakesleeanus/…
Podospora anserina	IGM	http://podospora.igmors.u-psud.fr/download_fr.html
Puccinia graminis	Broad Institute	http://www.broad.mit.edu/annotation/genome/puccinia_gramin…
Saccharomyces cerevisiae	Stanford	ftp://genome-ftp.stanford.edu/pub/yeast/data_download/sequ…
Schizosaccharomyces pombe	SANGER	ftp://ftp.sanger.ac.uk/pub/yeast/pombe/
Sclerotinia sclerotiorum	Broad Institute	http://www.broad.mit.edu/annotation/genome/sclerotinia_scle…
Stagonospora nodorum	Broad Institute	http://www.broad.mit.edu/annotation/genome/stagonospora_n…
Trichoderma reesei	JGI	http://genome.jgi-psf.org/Trire2/Trire2.download.html
Ustilago maydis	Broad Institute	http://www.broad.mit.edu/annotation/genome/ustilago_maydi…

Table 1. Source and url of the genomic data used in FUNGIpath.

	BRH	Inparanoid	OrthoMCL	Phylogeny
BRH		4,8%	3,7%	5,8%
Inparanoid	8,0%		22,4%	8,5%
OrthoMCL	12,4%	16,3%		8,3%
Phylogeny	32,4%	25,9%	32,6%	

Table 2. Comparison of groups predicted by the different methods. The upper triangular matrix displays the percent of identical group between the 2 methods, while the lower triangular matrix shows the percent of specific groups.

Confidence score used to evaluate the prediction reliability:

The score has been computed according the formula (1).

$$Score_{Final\ \ group} = \left(\sum_{m=methods} \frac{\#\,\mathrm{ID}_{\ \text{Final group}} \cap ID_{Method\ m}}{\#\,\mathrm{ID}_{\text{Final group}} . \#\,\mathrm{groups}_{\text{Method m}} \subset group_{Final}} \right) . \frac{10}{\#\,\text{methods}} \qquad (1)$$

For each method m, we compute the ratio of ID (Number of ID in the intersection between the final group and the orthologous groups predicted by the method m divided by the numbers of ID in the final group) and we divide this ratio by the number of groups predicted by the method m which belongs to the final group. Thus, if the method m has predicted the final group, the score is 1. If the method m did not predict exactly the final group, the score is between 0 and 1 (if the method built 2 orthologous groups with the same ID, the score is 0.5). Then, we computed the average score (we summed the score for each method and divided it by the total number of methods) and multiplied by 10 to scale the score from 0 to 10 (all methods correctly predict the same group). Higher is the score, better is the agreement between the 4 methods. Thus, we have an orthologous database built from different methods and whose a score permits to evaluate the reliability of the prediction.

.

Species	Number (Percent) of identical ID-EC		Number (Percent) of ID-EC with incomplete EC in KEGG but completed in FUNGIpath	
Aspergillus nidulans	700	77,0%	30	19,9%
Aspergillus oryzae	869	78,2%	33	18,6%
Fusarium graminearum	552	77,2%	4	5,1%
Laccaria bicolor	489	76,2%	3	4,0%
Magnaporthe grisea	765	73,5%	42	21,9%
Neurospora crassa	663	78,1%	16	14,5%
Podospora anserina	485	73,7%	3	4,4%
Saccharomyces cerevisiae	843	79,4%	117	30,2%
Schizosaccharomyces pombe	760	78,1%	55	21,3%
Sclerotinia sclerotiorum	511	78,6%	2	3,0%
Ustilago maydis	564	74,7%	19	16,0%
Yarrowia lipolytica	718	80,9%	22	15,8%

Table 3. Comparison of enzymatic data between KEGG and FUNGIpath (based on 12 shared species).

Species	Estimated number of CDS	FUNGIpath		KEGG		Swiss-Prot	
		Number of ID with EC	Percent of ID with EC	Number of ID with EC	Percent of ID with EC	Number of ID with EC	Percent of ID with EC
Aspergillus nidulans	9727	1750	18,0%	924	9,5%	149	1,5 %
Aspergillus oryzae	12074	1940	16,1%	1094	9,1%	82	0,7 %
Batrachochytrium dendrobatidis	8818	1087	12,3%				
Chaetomium globosum	11124	1550	13,9%			22	0,2 %
Coprinus cinereus	13544	1794	13,2%			5	0,0 %
Fusarium graminearum	11640	1711	14,7%	693	6,0%	41	0,4 %
Laccaria bicolor	20000	1500	7,5%	650	3,3%	3	0,0 %
Magnaporthe grisea	13146	1800	13,7%	1013	7,7%	30	0,2 %
Mycosphaerella graminicola	11395	1582	13,9%			4	0,0 %
Neurospora crassa	10082	1396	13,8%	808	8,0%	213	2,1 %
Phycomyces blakesleeanus	14792	1628	11,0%			8	0,1 %
Podospora anserina	10545	1587	15,0%	633	6,0%	22	0,2 %
Puccinia graminis	20567	1366	6,6%				%
Saccharomyces cerevisiae	5860	1159	19,8%	1027	17,5%	1163	19,8 %
Schizosaccharomyces pombe	5004	1013	20,2%	922	18,4%	876	17,5 %
Sclerotinia sclerotiorum	14522	1565	10,8%	615	4,2%	4	0,0 %
Stagonospora nodorum	16597	1772	10,7%			23	0,1 %
Trichoderma reesei	9129	1485	16,3%			21	0,2 %
Ustilago maydis	6631	1191	18,0%	725	10,9%	73	1,1 %
Yarrowia lipolytica	6520	1243	19,1%	860	13,2%	131	2,0 %

Table 4. Comparison of data for the FUNGIpath genomes.