Fungal Genetics and Biology 48 (2011) 353–358

Contents lists available at ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

Technological Advancement

FungiFun: A web-based application for functional categorization of fungal genes and proteins

Steffen Priebe [a,*], Jörg Linde [a,1], Daniela Albrecht [a,1], Reinhard Guthke [a,1], Axel A. Brakhage [b,c,2]

[a] Research Group Systems Biology/Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute (HKI), Jena, Germany
[b] Department of Molecular and Applied Microbiology, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute (HKI), Jena, Germany
[c] Institute of Microbiology, Friedrich Schiller University, Jena, Germany

ARTICLE INFO

Article history:
Received 22 September 2010
Accepted 1 November 2010
Available online 10 November 2010

Keywords:
Functional annotation
Fungi
Gene Ontology
FunCat
KEGG
Classification
Enrichment analysis
Candida
Aspergillus
Yeast

ABSTRACT

FungiFun assigns functional annotations to fungal genes or proteins and performs gene set enrichment analysis. Based on three different classification methods (FunCat, GO and KEGG), FungiFun categorizes genes and proteins for several fungal species on different levels of annotation detail. It is web-based and accessible to users without any programming skills. FungiFun is the first tool offering gene set enrichment analysis including the FunCat categorization. Two biological datasets for *Aspergillus fumigatus* and *Candida albicans* were analyzed using FungiFun, providing an overview of the usage and functions of the tool. FungiFun is freely accessible at https://www.omnifung.hki-jena.de/FungiFun/.

1. Introduction

High-throughput methods are frequently applied in fungal research. Using microarrays and protein arrays a multitude of biological data can be measured simultaneously. In addition, the rapid development of novel high-throughput methods like next-generation sequencing facilitates even more detailed measurements (Shendure, 2008). Global investigation of transcriptomic, proteomic or metabolomic data is essential to understand comprehensively gene regulatory mechanisms. A result of many global studies is a long list of gene or protein names. The interpretation of such gene lists is often difficult, but it is of great importance to elucidate their biological meaning. This is prerequisite to understand how a biological system reacts under a given environmental stimulus. One way of deciphering such gene or protein lists is gene set enrichment analysis (Hedegaard et al., 2009).

Several categorization approaches, which annotate specific functions of genes and their proteins, have been developed. At the same time, plenty of tools and software packages for functional categorization are on-hand. Generally, these enrichment tools can be separated into three classes: singular enrichment analysis (SEA), gene set enrichment analysis (GSEA) and modular enrichment analysis (MEA) (Huang et al., 2009a,b). FungiFun belongs to the first class, i.e. SEA, that presents the most traditional strategy for enrichment analysis. Despite the plethora of tools for well investigated species, there exists no web-based solution, which is usable without programming skills for many fungal species. The widely applied annotation tool DAVID (Huang et al., 2009a,b) does not support the common fungal gene IDs (e.g., CaO19.6385 for *Candida albicans* or AFUA_3G14490 for *Aspergillus fumigatus*). The same problem occurs with tools like GO Slimmer (Carbon et al., 2009) or FuncAssociate 2.0 (Berriz et al., 2009). GeneTrail (Backes et al., 2007) only supports *A. fumigatus* but no other fungi. In GeneCoDis2 (Carmona-Saez et al., 2007) only three yeast species are included. Other tools seem to support fungal gene IDs, but do not provide significant categories based on *p*-values (e.g., CGD GO Slim Mapper (Skrzypek et al., 2010)).

FungiFun attempts to fill this gap. The tool assigns functional annotations to a list of IDs representing fungal genes or proteins. Based on different classification methods like FunCat (Functional

* Corresponding author. Fax: +49 (0) 3641 532 2161.
E-mail addresses: steffen.priebe@hki-jena.de (S. Priebe), joerg.linde@hki-jena.de (J. Linde), daniela.albrecht@hki-jena.de (D. Albrecht), reinhard.guthke@hki-jena.de (R. Guthke), axel.brakhage@hki-jena.de (A.A. Brakhage).
[1] Fax: +49 (0) 3641 532 0803.
[2] Fax: +49 (0) 3641 532 0802.

1087-1845/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2010.11.001

Catalog; Ruepp et al., 2004), GO (Gene Ontology; Ashburner et al., 2000) and KEGG (Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000), FungiFun categorizes genes and proteins for currently 29 fungal species, including different human-pathogenic fungi, like *C. albicans*, *A. fumigatus* or *Candida glabrata*. These species are of major importance in the field of infection biology (Nucci and Marr, 2005; Dagenais and Keller, 2009; Brakhage et al., 2010) since they can cause superficial or allergic diseases but also invasive infections with mortality rates of up to 90% (Brakhage and Langfelder, 2002; Tekaia and Latge, 2005; Askew, 2008). Especially immunocompromised patients are at high risk to gain fungal infections, whereas candidiasis is the most frequent systemic fungal infection (Wilson et al., 2002; Brakhage and Zipfel, 2011). Beside human-pathogenic fungi, several plant-pathogenic fungi like *Botryotinia fuckeliana*, *Phaeosphaeria nodorum* or *Ustilago maydis* are also included in FungiFun. Model organisms (*Aspergillus nidulans*, *Neurospora crassa*) and industrial important species (*Aspergillus niger*, *Aspergillus terreus*, *Kluyveromyces lactis*, *Saccharomyces cerevisiae*) conclude the list of species supported by FungiFun.

Several hierarchical levels for categorization are eligible by the user to get insight into gene functions at different levels of detail. For each encountered category, an enrichment analysis is performed using Fisher's exact test. A pie chart provides an overview of the categorization. All results are available for download.

2. Material and methods

FungiFun uses a list of gene or protein names as input and performs functional classification. For each classification category, a gene set enrichment analysis is performed. The user can choose one of three different classification methods.

2.1. FunCat

FunCat is a hierarchically structured, organism-independent, flexible and scalable controlled classification system enabling the functional description of proteins from any organism. Over the last decade, FunCat has been established as a robust and stable annotation scheme that offers both, meaningful and manageable functional classification as well as ease of perception. FunCat consists of 28 main functional categories that cover general fields like cellular transport, metabolism and cellular communication/signal transduction. It exhibits a hierarchical structure with up to six levels of increasing specificity and, in total, version 2.1 includes 1362 functional categories.

Until now, there exists no enrichment analysis for the FunCat categorization. Thus, this is a novel feature within FungiFun.

2.2. GO

The goal of the GO Consortium is to produce a dynamic, controlled vocabulary that can be applied to all eukaryotes even as knowledge about the role of genes and proteins in cells is accumulating and changing. It is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The GO Consortium has developed three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. GO is structured in directed acyclic graphs, where nodes represent GO terms, and edges the relationships between the terms. A GO term can have several more specialized child terms and several more general parent terms at the same time. This circumstance complicates a generalization of GO categories at a specific hierarchical level. Therefore, FungiFun does not support this option. FungiFun uses the common top terms for generalization and subdivides all specialized categories into the namespaces biological process, cellular component and molecular function.

2.3. KEGG

The Kyoto Encyclopedia of Genes and Genomes is a knowledge base for systematic analysis of gene functions, linking genomic information with higher order functional information. It offers species dependent pathway maps representing current knowledge on the molecular interaction and reaction networks. The pathway maps are subdivided into Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Human Diseases and Drug Development. Each of these top categories has several 2nd level categories containing specific molecular functions which are again subdivided on a third level. All in all, the KEGG pathway contains 413 functional categories.

For all methods, the categorization performed by FungiFun is based on flat files, which are downloaded from the respective servers (www.uniprot.org, http://pedant.gsf.de, www.geneontology.org, www.genome.jp/kegg) and adjusted to be easily accessible by Perl scripts. These files are updated at regular intervals, because annotations are constantly improved. The file names as well as their date of last change are stated. Consequently, the user can realize how up to date the categorization is. Due to flat file based categorization, FungiFun performs all the work in only a few seconds.

Because functional categorization of biological units is an ongoing process, not every ID submitted by the user might be annotated by a functional category. On the other hand, some IDs have no one-to-one functional mapping, since proteins accomplish multifunctional tasks. The fraction of IDs with annotations as well as the number of non-unique mappings are part of the output of FungiFun.

Results are presented by tables containing all assigned categories at the chosen and the most detailed level as well as the number of identified IDs for each category (hits). In addition, a pie chart is shown which gives an overview about the encountered categories (Fig. 3). Categories are sorted by ascending p-values calculated by Fisher's exact test. The p-value indicates the significance of the number of hits for each category in the input dataset using the number of hits for the whole genome/proteome of the organism as background. More specifically, the p-values describe the probability that the encountered number of hits would have been found by chance. Thus, the smaller the p-values the more significant the categories are. It is probable that some categories are only covered by few hits. Therefore, Fisher's exact test is used instead of a χ^2-approximation. The calculation is performed by the Text::NSP::Measures::2D::Fisher Perl package (Banerjee and Pedersen, 2003). All results are available for download as tab-separated text file.

Besides using the whole genome as background list, the user can upload a specific background list, when the analysis is based on a subset of the whole genome. In this case, not every gene for which annotation is available will be used for classification. This feature can be used for subset gene lists descending from custom microarrays. In this case, the user should use a custom-made background list, only including the subset of measured genes. The use of a specialized background list will affect the results significantly.

3. Results

3.1. FungiFun utilization

The use of FungiFun via the web-interface is self-explanatory and easy (Fig. 1). After selecting one of the currently supported species (Table 1), the user chooses a text file, which contains a list of gene or protein IDs. The user can also enter the IDs directly into a textbox on the website. For each species, some example IDs can be

displayed to enable testing the tool before using it on own data. These examples also show the structure of the input that is needed. For most of the genes/proteins more than one possible name exists. To establish uniformity and comparability, only the most common IDs, which are consistent with the three methods are supported by FungiFun (see examples in Table 1). When the identification of the entered IDs fails, an adaption with a converter tool is possible (FungiFun ID-converter, included in FungiFun). In case of employing a user specific background ID list, it has to be uploaded only once, afterwards it will be available for selection.

After input of the IDs, the next step is selecting one of the categorization methods. Currently, FungiFun supports FunCat, GO and KEGG. In most cases all methods are selectable, except for some species for which a respective annotation is not available (see Table 1). Additionally, a p-value cutoff can be defined to constrain the resulting categories. Other optional parameters are selectable which adapt the output format. FunCat and KEGG are hierarchically structured and categorization is generated from the most general top-level down to the more detailed levels (four for FunCat and two for KEGG). The top-level is the most general one; each further level shows more details. For FungiFun categorization with GO, the user can select one of the three ontologies: biological process, cellular component or molecular function in Section 3. All detected categories, sorted in ascending order of p-values, are selectable to get an overview about assigned genes. Genes that were part of the input and are included in the found category are shown in green, the remaining quantity of genes in red. For KEGG categorization, links to the corresponding pathway maps are provided and related genes or proteins of the input set are highlighted. For downstream analysis, additional gene or protein names and further information about the found categories, the tab-separated result file should be used. A detailed instruction for the usage of FungiFun is available at the website.

To test the different categorization methods, two sample applications for the human-pathogenic fungi C. albicans and A. fumigatus will be demonstrated in the following sections.

3.2. Example 1: experimental oral infection by C. albicans

Transcript profiling data of *ex vivo* oral infection on reconstituted oral epithelium (RHE) were analyzed (Zakikhany et al., 2007). Preprocessing revealed 1320 differentially expressed genes during oral infection. In order to identify key processes during infection, this gene set was analyzed with the help of FungiFun. Gene Ontology (GO), FunCat and KEGG annotations were used to scan for categories significantly enriched within the set of differentially expressed genes. Significant enrichment was defined by the *p*-value cutoff 0.05. A number of interesting categories were identified in this way (Fig. 2). The GO process "nitrogen compound metabolic process" consists of gene products necessary to utilize various organic and inorganic compounds. This process is significantly enriched with differentially expressed genes, reflecting the nitrogen shift the fungus faces during infection. A number of proteins annotated with this process belong to the family of secreted aspartic proteinases. These proteins have been shown to play an important role during the infection process, giving *C. albicans* the possibility to acquire nutrients from host proteins (Hube, 2004). The importance of this protein family is further strengthened by the enriched molecular

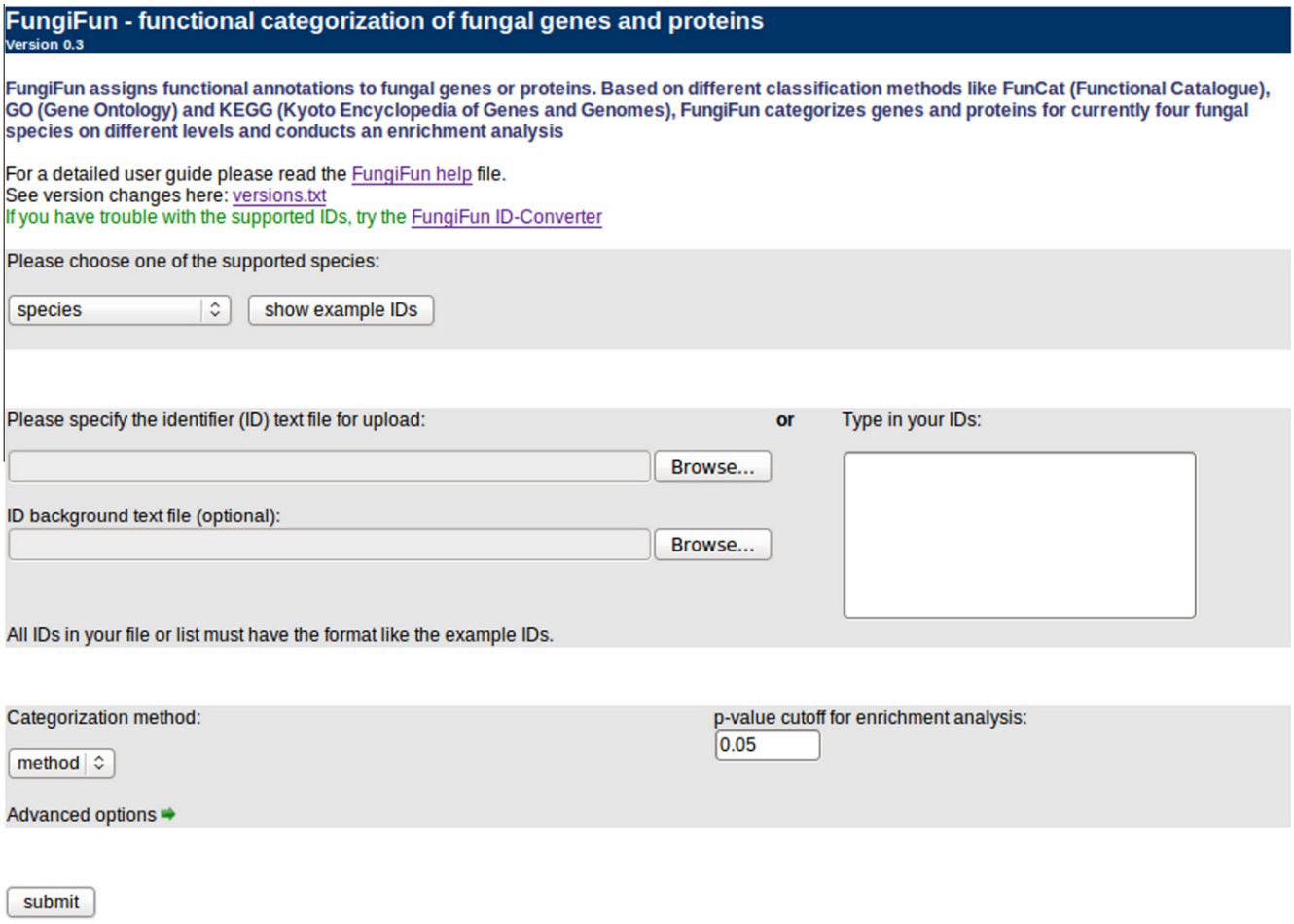

Fig. 1. Screenshot of the FungiFun webpage.

Table 1
Supported species and available methods of FungiFun. Example IDs are chosen randomly.

Species	FunCat	GO	KEGG	Example IDs
Ashbya gossypii	X	X	X	AAL162C, ABL102C, ACR106C
Aspergillus clavatus		X	X	ACLA_010450, ACLA_000680, ACLA_069440
Aspergillus flavus		X	X	AFLA_048690, AFLA_057030, AFLA_014190
Aspergillus fumigatus Af293	X	X	X	AFUA_1G06810, AFUA_1G17180, AFUA_3G14490
Aspergillus fumigatus A1163		X		AFUB_034200, AFUB_068250, AFUB_078360
Aspergillus nidulans	X	X	X	AN0034.2, AN0044.2, AN0046.2
Aspergillus niger		X	X	An06g00170, An02g11150, An06g00290
Aspergillus oryzae	X	X	X	AO090672000003, AO090102000550, AO090023000577
Aspergillus terreus	X	X		ATEG_00364, ATEG_09210, ATEG_03138
Botryotinia fuckeliana	X	X	X	BC1G_00061, BC1G_12399, BC1G_13689
Candida albicans	X	X	X	CaO19.6385, CaO19.3962, CaO19.8363
Candida dubliniensis		X	X	CD36_03650, CD36_19890, CD36_85860
Candida glabrata	X	X	X	CAGL0B04213g, CAGL0G09449g, CAGL0G09757g
Candida tropicalis	X		X	CTRG_05794, CTRG_01181, CTRG_01187
Cryptococcus neoformans JEC21		X	X	CNB00190, CNA04190, CNL04800
Cryptococcus neoformans B-3501A		X	X	CNBB3880, CNBI2040, CNBE3120
Debaryomyces hansenii		X	X	DEHA2D12936g, DEHA2F18744g, DEHA2E05676g
Encephalitozoon cuniculi		X	X	ECU05_1590, ECU01_0650, ECU01_0240
Kluyveromyces lactis	X	X	X	KLLA0A00286g, KLLA0A00374g, KLLA0A00418g
Lodderomyces elongisporus	X	X	X	LELG_00015, LELG_00121, LELG_00630
Neosartorya fischeri		X	X	NFIA_097440, NFIA_075290, NFIA_054850
Neurospora crassa	X	X	X	NCU00106, NCU05235, NCU06063
Phaeosphaeria nodorum		X	X	SNOG_06553, SNOG_03344, SNOG_13776
Pichia guilliermondii	X	X	X	PGUG_00035, PGUG_00611, PGUG_01239
Pichia stipitis		X	X	PICST_76205, PICST_68558, PICST_27980
Saccharomyces cerevisiae	X	X	X	YAL062W, YBL098W, YBR065C
Schizosaccharomyces pombe	X	X	X	SPAC977.16c, SPBC685.09, SPBC1105.05
Ustilago maydis	X	X	X	UM06453, UM05408, UM05776
Yarrowia lipolytica	X	X	X	YALI0A00264g, YALI0B00572g, YALI0F20790g

function "aspartic-type endopeptidase activity". A number of environmental parameters are changed during infection causing different kinds of stress the fungus has to deal with. The overrepresented category "stress response" (FunCat) provides evidence for this fact. More specifically, the categories "response to oxidative stress" of GO and "oxidative stress response" of FunCat contain more differentially expressed genes, than randomly expected. This shows that both methods are able to identify similar biological processes. As an aerobically growing yeast, *C. albicans* produces reactive oxygen species (ROS) as by-product of the oxidative metabolism. Furthermore, the fungus is often faced with the oxidative burst generated by host immune effector cells. To cope with this harsh environment, the fungus applies specific signaling and adaptation processes (Chauhan et al., 2006). During infection, the available amount of iron is limited. For this reason, *C. albicans* possesses ways to acquire iron (Almeida et al., 2009) and to cope with a small intracellular iron level. A significant number of genes belonging to the GO category "heme biosynthetic process", the FunCat category "metabolism of porphyrins" and the KEGG category "porphyrin and chlorophyll metabolism" are downregulated (Fig. 3). There is an overlap between genes belonging to these categories showing that all three methods allow to find similar results. The downregulation of iron binding proteins might reflect the need of having a higher intracellular iron level. Furthermore, the GO and the FunCat categories "siderophore-iron transport" are overrepresented. *C. albicans* uses siderophores of other organisms as iron source (Lan et al., 2004). Since there are no other microorganisms in the RHE model, the regulation of siderophore genes can be explained in two ways: either the fungus uses proteins annotated to siderophore transport for high affinity iron uptake, or the fungus senses contact to oral epithelia and predicts there are siderophores as a potential iron source. This kind of adaptive prediction has already been shown for other fungi (Mitchell et al., 2009). The importance of iron uptake is further strengthened by the GO molecular function "ferric-chelate reductase activity" which is significantly overrepresented.

3.3. Example 2: *A. fumigatus* thermotolerance

Differential in gel electrophoresis (DIGE) data of a temperature shift from 30 °C to 48 °C obtained from *A. fumigatus* were analyzed

Results:

biological_process molecular_function cellular_component all namespaces

IDs with annotation: 1014 (76.82%)
IDs without annotation: 306 (23.18%)

GO number	category name	category namespace	hits	p-value
GO:0030476	ascospore wall assembly	biological_process	15	0.000668429948185271
GO:0015892	siderophore-iron transport	biological_process	8	0.00182321383027179
GO:0006783	heme biosynthetic process	biological_process	6	0.00413659156415911
GO:0006979	response to oxidative stress	biological_process	29	0.00431007728630947
GO:0006612	protein targeting to membrane	biological_process	6	0.008498429463807
GO:0044270	cellular nitrogen compound catabolic process	biological_process	3	0.00948221610673934
GO:0006807	nitrogen compound metabolic process	biological_process	9	0.00969793287942132

Fig. 2. Screenshot of the FungiFun GO result section for the oral infection example dataset. Only categories with a *p*-value below 0.01 are shown. The GO namespaces are selectable on top. The fraction of genes with and without annotations are stated below, followed by the found categories ordered according ascending *p*-values.

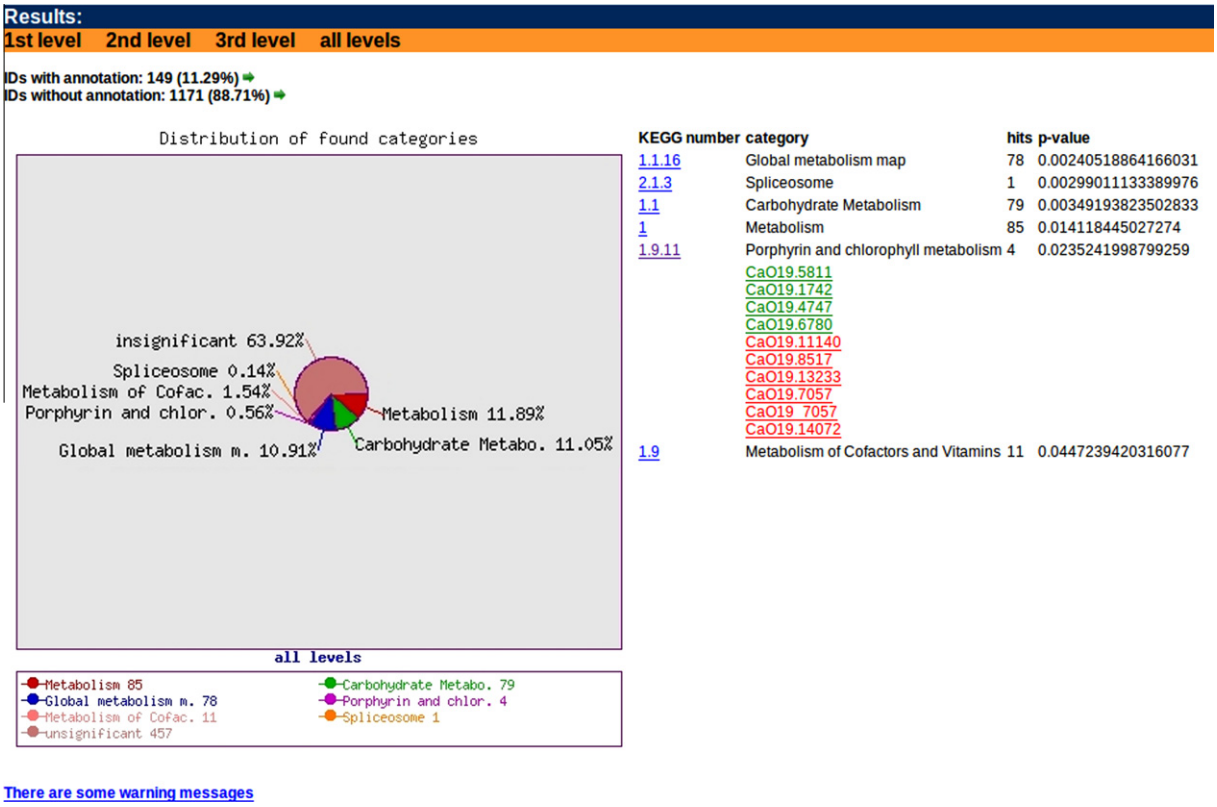

Fig. 3. Screenshot of the FungiFun KEGG result section for example dataset obtained with the oral infection model. The found categories are clickable, leading to the genes which belong to them. The genes also provide links to the appropriate KEGG pathway map, in which included genes are colored. The pie chart is an optional feature, showing the distribution of genes within the found categories. Based on calculated *p*-values beyond the cutoff, the assignment of a majority of these categories is insignificant.

(Albrecht et al., 2010). Preprocessing of data resulted in a list of 64 differentially regulated proteins. The dataset was analyzed using GO, FunCat and KEGG to identify biological processes that are important for the thermotolerance of the fungus. Enrichment analysis was conducted, considering categories with a *p*-value below 0.05 as significant. GO and FunCat contain annotations to nearly all proteins (61 and 63, respectively), whereas in KEGG only about half of the proteins are annotated (31 proteins). Protein un- and re-folding is the most important process in the heat shock response (Burnie et al., 2006). This can be deduced easily from the fact that 10 out of the 64 differentially regulated proteins are chaperones. This finding is reflected by the overrepresented categories "protein folding and stabilization" and "unfolded protein response" in FunCat and categories "protein folding" and "response to unfolded protein" in GO. Moreover, the categories "heat shock response", "stress response" and "temperature perception and response" in FunCat as well as "response to stress" in GO indicate that the fungus is under stress and, in addition, has mechanisms to cope with that. It is known from *S. cerevisiae*, that heat shock enhances oxygen respiration which results in an increase of the formation of ROS and an activation of the oxidative stress response (Sugiyama et al., 2000). This is reflected by the significantly overrepresented FunCat categories "oxygen and radical detoxification" and "oxidative stress response" but also by the GO categories "oxygen transport" and "oxygen binding". A heat shock-dependent regulation of glycolytic enzymes was demonstrated for different fungi (e.g. Boucherie et al., 1996). An increase in glycolysis enzymes could provide energy needed for the ATP-dependent protein re-folding by chaperones (Gasch et al., 2000). This increase can be found in the categories "glycolysis and gluconeogenesis" of FunCat, "glycolysis" of GO as well as "glycolysis/gluconeogenesis" and "energy metabolism" in KEGG. Heat shock in *A. fumigatus* also leads to an increased protein turnover and hence to a higher amino acid biosynthesis rate. This is reflected by many overrepresented categories such as "aminoacyl-tRNA synthetases" in FunCat, "aminoacyl-tRNA ligase activity" in GO as well as "nitrogen metabolism", "cysteine and methionine metabolism" and "metabolism of other amino acids" in KEGG. By contrast, in *S. cerevisiae* amino acid biosynthesis is repressed after heat shock (Ye et al., 2009).

4. Discussion

Elucidating significant biological functions and processes within large data sets is a challenging and important goal. Novel high-throughput technologies provide gene expression data in an unprecedented magnitude, even for species which have not been investigated extensively so far. Hence, gene set enrichment analysis tools are crucial and needed for those species as well. The currently 29 supported fungal species in FungiFun are a result of the available overlapping annotation information from FunCat, GO and KEGG. For some species not all methods are available but they will be included as soon as annotation files are on-hand. In the future, additional species can be incorporated too. There are a number of genes of the species under consideration, which do not have any annotated function. This missing information is a well known problem but annotation is improved continuously, leading to an improved gene coverage. However, FungiFun is already able to produce meaningful biological results. Due to the power of the applied statistical test, important overrepresented categories can be found. Furthermore, the tool helps to hypothesize new gene functions. For example, if a specific function was significantly overrepresented within a list of genes, it would be likely that the non-annotated genes of this list have a similar function. Besides the common application of GO by most of the existing classification tools, FungiFun additionally uses FunCat and KEGG. FungiFun is the first

tool that offers a statistical test for gene set enrichment with FunCat categories. Especially for some fungal species like *Ashbya gossypii*, *Aspergillus flavus* or *K. lactis* the usage of FunCat annotation might be helpful, since it contains more genes with associated description than GO or KEGG. However, on average most of the fungal genes are annotated in GO whereas KEGG contains generally only a subset of all genes. For two different example data sets, all three categorization methods found meaningful categories. These categories can help understanding key properties of complex biological phenomena and provide new hypotheses for gene functions. In some cases an overlap between the three methods was observed, confirming the results. For example, ''biosynthesis of porphyrins'' was found by all three approaches during the oral infection of *C. albicans*. Nevertheless, each of the three categorization methods also provides unique annotations. Hence, the usage of all three can be helpful to expand the knowledge on gene/protein functions. Within the plethora of tools, FungiFun is a specialized application, due to its restriction to fungal species. Therefore it is customized for fungal gene names/IDs and thus, can easily be used by scientists working in the field of fungal biology.

Acknowledgments

This work is part of the research program of the Jena Centre for Systems Biology of Ageing – JenAge funded by the German Ministry for Education and Research (Bundesministerium für Bildung und Forschung – BMBF; Support Code: 0315581) (SP). Part of this work was supported by the DFG Priority Program 1160 *'Colonisation and infection by human-pathogenic fungi*' (DA) to A.A.B. and R.G., and the excellence graduate school *Jena School for Microbial Communication* (JSMC) (JL).

References

Albrecht, D., Guthke, R., Brakhage, A.A., Kniemeyer, O., 2010. Integrative analysis of the heat shock response in *Aspergillus fumigatus*. BMC Genom. 11 (1), 32.

Almeida, R.S., Wilson, D., Hube, B., 2009. *Candida albicans* iron acquisition within the host. FEMS Yeast Res. 9 (7), 1000–1012.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25 (1), 25–29.

Askew, D.S., 2008. *Aspergillus fumigatus*: virulence genes in a street-smart mold. Curr. Opin. Microbiol. 11, 331–337.

Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y.A., Müller, R., Meese, E., Lenhof, H.P., 2007. GeneTrail – advanced gene set enrichment analysis. Nucl. Acid Res. 35 (Web Server Issue: W186-92).

Banerjee, S., Pedersen, T., 2003. The design, implementation and use of the Ngram statistics package. In: Proc. of the Fourth Int. Conf. on Int. Text Proc. and Comp. Ling, pp. 370–381.

Berriz, G.F., Beaver, J.E., Cenik, C., Tasan, M., Roth, F.P., 2009. Next generation software for functional trend analysis. Bioinformatics 25 (22), 3043–3044.

Boucherie, H., Sagliocco, F., Joubert, R., Maillet, I., Labarre, J., Perrot, M., 1996. Two-dimensional gel protein database of *Saccharomyces cerevisiae*. Electrophoresis 17, 1683–1699.

Brakhage, A.A., Langfelder, K., 2002. Menacing mold: the molecular biology of *Aspergillus fumigatus*. Annu. Rev. Microbiol. 56, 433–455.

Brakhage, A.A., Zipfel, P.F., 2011. Overview about fungal pathogens. In: Kaufmann, S.H.E., Rouse, B.T., Sacks, D.L. (Eds.), The immune response to infection. ASM Press.

Brakhage, A.A., Bruns, S., Thywissen, A., Zipfel, P.F., Behnsen, J., 2010. Interaction of phagocytes with filamentous fungi. Curr. Opin. Microbiol. 13 (4), 409–415.

Burnie, J.P., Carter, T.L., Hodgetts, S.J., Matthews, R.C., 2006. Fungal heat-shock proteins in human disease. FEMS Microbiol. Rev. 30, 53–88.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25 (2), 288–289.

Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J.M., Pascual-Montano, A., 2007. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Gen. Biol. 8 (1), R3.

Chauhan, N., Latge, J.P., Calderone, R., 2006. Signalling and oxidant adaptation in *Candida albicans* and *Aspergillus fumigatus*. Nat. Rev. Microbiol. 4 (6), 435–444.

Dagenais, T.R., Keller, N.P., 2009. Pathogenesis of *Aspergillus fumigatus* in invasive aspergillosis. Clin. Microbiol. Rev. 22, 447–465.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., Brown, P.O., 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11, 4241–4257.

Hedegaard, J., Arce, C., Bicciato, S., Bonnet, A., Buitenhuis, B., Collado-Romero, M., Conley, L.N., Sancristobal, M., Ferrari, F., Garrido, J.J., Groenen, M.A., Hornshøj, H., Hulsegge, I., Jiang, L., Jiménez-Marín, A., Kommadath, A., Lagarrigue, S., Leunissen, J.A., Liaubet, L., Neerincx, P.B., Nie, H., van der Poel, J., Prickett, D., Ramirez-Boo, M., Rebel, J.M., Robert-Granié, C., Skarman, A., Smits, M.A., Sørensen, P., Tosser-Klopp, G., Watson, M., 2009. Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. BMC Proc. 3 (Suppl 4) (S5).

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37 (1), 1–13.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (1), 44–57.

Hube, B., 2004. From commensal to pathogen: stage- and tissue-specific gene expression of *Candida albicans*. Curr. Opin. Microbiol. 7 (4), 336–341.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28 (1), 27–30.

Lan, C.Y., Rodarte, G., Murillo, L.A., Jones, T., Davis, R.W., Dungan, J., Newport, G., Agabian, N., 2004. Regulatory networks affected by iron availability in *Candida albicans*. Mol. Microbiol. 53 (5), 1451–1469.

Mitchell, A., Romano, G.H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., Pilpel, Y., 2009. Adaptive prediction of environmental changes by microorganisms. Nature 460 (7252), 220–224.

Nucci, M., Marr, K.A., 2005. Emerging fungal diseases. Clin. Infect. Dis. 41 (4), 521–526.

Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Gueldener, U., Mannhaupt, G., Muensterkoetter, M., Mewes, H.W., 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucl. Acids Res. 32, 5539–5545.

Shendure, J., 2008. The beginning of the end for microarrays? Nat. Meth. 5 (7), 585–587.

Skrzypek, M.S., Arnaud, M.B., Costanzo, M.C., Inglis, D.O., Shah, P., Binkley, G., Miyasato, S.R., Sherlock, G., 2010. New tools at the Candida genome database: biochemical pathways and full-text literature search. Nucl. Acids Res. 38, 428–432.

Sugiyama, K., Izawa, S., Inoue, Y., 2000. The Yap1p-dependent induction of glutathione synthesis in heat shock response of *Saccharomyces cerevisiae*. J. Biol. Chem. 275, 15535–15540.

Tekaia, F., Latge, J.P., 2005. *Aspergillus fumigatus*: saprophyte or pathogen? Curr. Opin. Microbiol. 8, 385–392.

Wilson, L.S., Reyes, C.M., Stolpman, M., Speckman, J., Allen, K., Beney, J., 2002. The direct cost and incidence of systemic fungal infections. Value Health 5 (1), 26–34.

Ye, Y., Zhu, Y., Pan, L., Li, L., Wang, X., Lin, Y., 2009. Gaining insight into the response logic of *Saccharomyces cerevisiae* to heat shock by combining expression profiles with metabolic pathways. Biochem. Biophys. Res. Commun. 385, 357–362.

Zakikhany, K., Naglik, J.R., Schmidt-Westhausen, A., Holland, G., Schaller, M., Hube, B., 2007. In vivo transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination. Cell. Microbiol. 9, 2938–2954.