Computational Molecular Biology 2014, Vol. 4, No. 7, 1-17
http://cmb.biopublisher.ca

Research Article **Open Access**

FunSecKB2: a fungal protein subcellular location knowledgebase

John Meinken[1,4], David K. Asch[2,4], Kofi A. Neizer-Ashun[3], Guang-Hwa Chang[3], Chester R.Cooper JR[2,4], Xiang Jia Min[2,4]

1. Department of Computer Science and Information Systems, Youngstown State University, OH 44555, USA
2. Department of Biological Sciences, Youngstown State University, Youngstown, OH 44555, USA
3. Department of Mathematics & Statistics, Youngstown State University, Youngstown, OH 44555, USA
4. Center for Applied Chemical Biology, Youngstown State University, Youngstown, OH 44555, USA

Corresponding author email: xmin@ysu.edu Phone: (330) 941-1945 Fax: (330) 941-1483

Computational Molecular Biology, 2014, Vol.4, No.7 doi: 10.5376/cmb.2014.04.0007
Received: 05 Aug., 2014
Accepted: 21 Sep., 2014
Published: 22 Oct., 2014

Preferred citation for this article:
Meinken et al., 2014, FunSecKB2: a fungal protein subcellular location knowledgebase, Computational Molecular Biology, Vol.4, No.7, 1-17 (doi: 10.5376/cmb.2014.04.0007)

Abstract FunSecKB2 is an improved and updated version of the fungal secretome and subcellular proteome, i. e. protein subcellular location, knowledgebase. The fungal protein sequence data were retrieved from UniProtKB, consisting of nearly 2 million entries with 167 species having a complete proteome. The assignments of protein subcellular locations were based on curated information and prediction using seven computational tools. The tools used for subcellular location prediction include SignalP, WoLF PSORT, Phobius, TargetP, TMHMM, FragAnchor, and PS-Scan. Secreted proteins, i.e. secretomes, along with 15 other subcellular proteomes were predicted. The database can be searched by users using several different types of identifiers, gene name or keyword(s). A subcellular proteome from a species can be searched or downloaded. BLAST searching whole fungal protein data or secretomes is available. Community annotation of subcelluar locations based on experimental evidence is also supported. A primary analysis revealed that the secretome size of a fungal species is one of the determining factors to its lifestyle. The Gene Ontology and protein domain analysis of fungal secretomes revealed that fungal secretomes contain a large number of hydrolases, peptidases, oxidoreductases, and lysases, which may have potential applications in bio-processing of chemical wastes or biofuel production. The database provides an important and rich resource for the fungal community looking for protein subcellular location information and performing comparative subcellular proteome analysis.

Database URL: http://proteomics.ysu.edu/secretomes/fungi2/index.php

Keywords Computational prediction; Fungi; Secreted protein; Secretome; Signal peptide; Subcellular location; Subcellular proteome

Introduction

Fungi play important roles in nature and in our daily life. In nature, fungal species serve as decomposers of biomass, which is critical for carbon and nutrient cycling. In our daily life, edible mushrooms are well-known examples of fungi. *Saccharomyces cerevisiae*, known as a baker's yeast, is widely used in winemaking, baking and brewing. Some fungi are also known as producers for drugs, such as antibiotics. Fungal species are also important pathogens in insects, animals, human and plants.

Fungi belong to one of the four kingdoms of eukaryotic organisms. Fungal cells contain multiple subcellular compartments for performing different subcellular activities. For example, a mitochondrion, which is a membrane-enclosed structure, is mainly used to provide cellular energy; and a nucleus is a place for storing genetic materials and a site for controlling gene transcription. In this work, we define a secretome as all proteins secreted outside the plasma membrane in a species. These proteins include cell wall proteins, extracellular matrix proteins, and secreted soluble proteins that may serve as a hormone or signal molecule or an enzyme. However, the proteins in the secretory pathway machinery were not included, which is slightly different form the original definition of a secretome (Tjalsma et al., 2000; Lum and Min, 2011a). Secreted proteins in biotrophic fungi are identified as the main effectors responsible for

pathogenic or symbiotic interactions between plants and fungi (Girard et al., 2013). Saprophytic fungi secrete a large number of families of hydrolytic enzymes such as glycoside hydrolases for breaking down complex biomaterials like lignin and cellulose (Martinez et al., 2004; Martinez et al., 2009; Murphy et al., 2011). Recently, along with complete genome sequencing of many fungi, identification and analysis of secretomes in fungi has been an important subject of research, using both computational and experimental approaches (Bouws et al., 2008). For example, the secretomes have been reported in following fungi including *Aspergillus niger* (Tsang et al., 2009; Braaksma et al., 2010), *Aspergillus fumigatus* (Powers-Fletcher et al., 2011), *Candida albicans* (Lee et al., 2003; Ene et al., 2012), *Doratomyces stemonitis* C8 (Peterson et al., 2011), *Fusarium graminearum* (Paper et al., 2007; Brown et al., 2012), *Irpex lacteus* (Salvachúa et al., 2013), *Magnaporthe oryzae* (Jung et al., 2012), *Mycosphaerella graminicola* (Morais et al., 2012), Paracoccidioides (a complex of several phylogenetic species) (Weber et al., 2012), *Penicillium echinulatum* (Ribeiro et al., 2012), *Phanerochaete chrysosporium* (Wymelenberg et al., 2005), *Sclerotinia sclerotiorum* (Yajima and Kav, 2006), *Trichoderma harzianum* (Do Vale et al., 2012), and *Ustilago maydis* (Mueller et al., 2008).

Two fungal specific secretome databases, the Fungal Secretome Database (FSD, http://fsd.snu.ac.kr/) and the Fungal Secretome Knowledgebase (FunSecKB, http://proteomics.ysu.edu/secretomes/fungi.php) have been constructed for the community to search fungal secretome related information (Choi et al., 2010; Lum and Min, 2011). FSD was constructed using a three-layer hierarchical identification rule based on 9 different programs (Choi et al., 2010). We developed the FunSecKB using 6 different tools for predicting secreted proteins from RefSeq data set of fungi (Lum and Min, 2011). However, since the release of FunSecKB, the available fungal protein data have been increased tremendously. In this work, we describe FunSecKB2, a fungal protein subcellular location knowledgebase, also known as the Fungal Secretome and Subcellular Proteome Knowledgebase (Version 2), that is, an expanded, updated, and improved version of FunSecKB. FunSecKB2 is constructed with a refined protocol for including curated subcellular information and predicted information on secretomes and other subcellular proteomes of 15 subcellular locations. This improved fungal protein knowledgebase is expected to serve as a central portal for providing information on fungal protein subcellular locations to users in the fungal research and industrial community who are interested in exploiting fungi for a global development of the bioeconomy (Lange et al., 2012).

1 Data Collection and Database Implementation

1.1 Data collection

The protein sequences for all fungi were retrieved from the UniProtKB/Swiss-Prot dataset and the UniProtKB/TrEMBL dataset (release 2013_08) (http://www.uniprot.org/downloads). The UniProtKB/Swiss-Prot dataset contains manually annotated non-redundant protein sequences with information extracted from literature of experimental results and curator-evaluated computational analysis (The UniProt Consortium, 2014). The UniProtKB/TrEMBL contains protein sequences associated with computationally generated annotation and large-scale functional characterization. The dataset consisted of a total of 1,976,832 fungal proteins with 30,859 and 1,945,973 entries retrieved from the UniProtKB/Swiss-Prot dataset and the TrEMBL dataset, respectively.

1.2 Methods for protein subcellular location assignment

The fungal protein sequences were processed using the following programs: SignalP (version 3.0 and 4.0, http://www.cbs.dtu.dk/services/SignalP/), (Bendtsen et al., 2004b; Petersen et al., 2011), Phobius (http://phobius.binf.ku.dk/) (Käll et al., 2007), WoLF PSORT (http://wolfpsort.org/) (Horton et al., 2007), and TargetP (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al., 2007) for signal peptide and subcellular location prediction. These predictors were previously evaluated favorably and are widely used by the fungal secretome research community (Min, 2010). TMHMM (http://www.cbs.dtu.dk/services/TMHMM) was used to identify proteins having transmembrane domains (Krogh et al., 2001) and Scan-Prosite (called PS-Scan in standalone version) (http://www.expasy.

org/tools/scanprosite/) was used to scan endoplasmic reticulum (ER) targeting sequence (Prosite: PS00014) (de Castro et al., 2006; Sigrist et al., 2010). For predicting membrane proteins using TMHMM, the entries having membrane domains not located within the N-terminus (the first 70 amino acids) were treated as real membrane proteins. Protein sequences predicted to have a signal peptide by SignalP (version 3) were further processed using the FragAnchor webserver to identify the glycosylphosphatidyinositol (GPI) anchors (http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html) (Poisson et al., 2007). With the exception of FragAnchor, we used the standalone tools installed on a local Linux system for data processing. The commands for how to run these tools often can be found in the "readme" page in each downloaded package and were summarized by Lum and Min (2013).

The categories of fungal protein subcellular locations include: secreted proteins, mitochondrial (membrane or non-membrane), ER (membrane or lumen), cytosol (cytoplasm), cytoskeleton, Golgi apparatus (membrane or lumen), nuclear (membrane or non-membrane), vacuolar (membrane or non-membrane), lysosome, peroxisome, plasma membrane, and other membrane proteins. For assigning a protein subcellular location, the UniProtKB annotation and our curated subcellular information was considered prior to using prediction information. For proteins not having annotated subcellular information, their subcellular location assignments are based on prediction. Our recent accuracy evaluation of the computational tools revealed that the highest prediction accuracy (92.1% in sensitivity and 98.9% in specificity) for fungal secretomes was achieved by combining SignalP, WoLF PSORT, and Phobius for signal peptide prediction, with TMHMM for eliminating membrane proteins and PS-Scan for removing ER targeting proteins (Min, 2010). Thus, the secretome was limited to include manually curated secreted proteins and proteins predicted having a signal peptide at their N-terminus by all the three programs but not having a transmembrane domain or an ER targeting signal. In this work, SignalP4 is used to replace SignalP3 as SignalP4 improves the prediction accuracy (Petersen et al., 2011; Melhem et al., 2013). However, the information generated by SignalP3 was also included as it predicts signal peptide cleavage sites more accurately than SignalP4 (Petersen et al., 2011). The detailed methods for assigning a protein subcellular location are described below.

Secreted protein

Secreted proteins are further divided as curated secreted proteins, highly likely secreted, likely secreted, and weakly likely secreted proteins. Curated secreted proteins include proteins that are annotated to be "secreted" or "extracellular" or "cell wall" in subcellular location from the UniProtKB/Swiss-Prot data set which are "reviewed". It also includes manually collected secreted proteins from recent literature by our curators. Three predictors consisting of SignalP4, Phobius, and WoLF PSORT are used for protein secretory signal peptide or subcellular location prediction. The highly likely secreted, likely secreted, and weakly likely secreted proteins are proteins that are predicted to be secreted or contain a secretory signal peptide by three, two, or one of the three predictors, respectively. These proteins do not have a transmembrane domain or an ER retention signal.

ER proteins

ER proteins were predicted by WoLF PSORT and PS-Scan. Proteins predicted to contain a signal peptide by SignalP 4.0 and an ER target signal (Prosite: PS00014) by PS-Scan were treated as luminal ER proteins. Further, if they contain one or more transmembrane domains, they are classified as ER membrane proteins.

GPI-anchored proteins

Signal peptide containing proteins that were predicted to have a GPI anchor by FragAnchor were further classified as GPI-anchored proteins. Protein sequences predicted to have a signal peptide and a GPI anchor may attach to the outer leaflet of the plasma membrane or be secreted becoming components of the cell wall.

Proteins in other subcellular locations

Other subcellular locations including mitochondria, cytosol (cytoplasm), cytoskeleton, Golgi apparatus,

lysosome, nucleus, peroxisome, plasma membrane and vacuole proteins were predicted by WoLF PSORT. For proteins predicted as located in mitochondria, Golgi apparatus, nucleus, and vacuole, if a protein contains one or more transmembrane domain, it is further classified as a membrane protein in that specific subcellular location.

1.3 Database implementation

The data were stored in a relational database using MySQL hosted in a Linux server. The user interface and modules to access the data were implemented using PHP. BLAST utility and community annotation submission can be accessed from links on the main user interface at http://proteomics.ysu.edu/secretomes/fungi2/index.php. The Supplementary Tables and other data described in the work can be downloaded at http://proteomics.ysu.edu/publication/data/FunSecKB2/.

2 Results

2.1 Evaluation of prediction accuracies of protein subcellular locations

The prediction methods we employed as described above were based on our previous evaluation of computational tools (Min, 2010; Meinken and Min, 2012; Melhem et al., 2013). To further estimate the prediction accuracies of our methods for each subcellular location in this dataset we retrieved 14884 proteins having an annotated, unique subcellular location from UniProtKB/Swiss-Prot set. Proteins having multiple subcellular locations or labeled as "fragment" were excluded. The prediction accuracies were measured as the sensitivity, the specificity, and Matthews correlation coefficient (MCC) based on formulas used previously (Min, 2010). The accuracy results are shown in Table 1. The prediction accuracies from plasma membrane and lysosome were not included as the numbers of positive proteins were too few (<20). In comparing with methods using a single tool, our method - i.e. using a combination of multiple tools including SignalP 4.0, WoLF PSORT and Phobius for secretory signal peptide prediction and PS-Scan for removing ER proteins and TMHMM for removing membrane proteins - significantly improved the prediction accuracy for secretomes (Min, 2010; Meinken and Min, 2012). For prediction of secretome size in a given species, the predicted set of highly likely secreted proteins would provide a relatively accurate estimation as this method has the highest specificity (>0.99), and interestingly, the number of false negatives is close to the number of false positives in the dataset used for evaluation. Including the predicted likely secreted protein set into a secretome only slightly decreased the MCC value as only a small number of entries belong to this category. However, the predicted set of weakly likely secreted proteins needs to be treated with caution as the number of false positives was far more than the number decrease of the false negatives (Table 1).

Table 1 Evaluation of prediction accuracies of fungal protein subcellular locations

Subcellular location	True positive	False positive	True negative	False Negative	Sn	Sp	MCC
HLS	1364	130	13269	121	0.919	0.990	0.906
HLS+LS	1401	188	13211	84	0.943	0.986	0.902
HLS+LS+WLS	1412	337	13062	73	0.951	0.975	0.862
Mitochondria	1595	887	12015	387	0.805	0.931	0.671
ER	19	11	13873	981	0.019	0.999	0.102
Golgi apparatus	5	2	14527	350	0.014	1.000	0.098
Nucleus	4483	2771	6823	807	0.847	0.711	0.535
Vacuole	0	0	14389	495	0.000	1.000	
Peroxisome	9	15	14722	138	0.061	0.999	0.148
Cytoplasm	1293	762	10611	2218	0.368	0.933	0.371
Cytoskeleton	87	234	14055	508	0.146	0.984	0.175

Note: HLS: highly likely secreted; LS: likely secreted; WLS: weakly likely secreted; ER: Endoplasmic reticulum; Sn: sensitivity; Sp:specificity; MCC: Matthews correlation coefficient.

We also compared the accuracy of mitochondrial proteins predicted by WoLF PSORT and TargetP. We found that the MCC values were 0.67 for WoLF PSORT and 0.56 for TargetP, and we also found using both tools increased the mitochondrial protein prediction specificity, from 0.93 using WoLF PSORT only to >0.98 when both were used. However, using both tools did not improve the MCC value due to the decrease in prediction sensitivity. Thus, we selected WoLF PSORT for assigning mitochondrial proteins. However, a user should be aware that if both WoLF PSORT and TargetP predicted the protein is a mitochondrial protein, the prediction is more reliable than prediction just from one of them.

The prediction accuracies for other subcellular locations vary significantly. Prediction of nuclear proteins had 0.85 in sensitivity, 0.71 in specificity, and 0.53 in MCC. The accuracies for other subcellular locations including the ER, Golgi apparatus, vacuole, peroxisome, cytoplasm, and cytoskeleton were very low in MCC (<0.4) (Table 1). However, it should be noted that the low accuracies were caused by very low sensitivities, and in fact, the specificities were relatively high (>0.98). Thus, there are a good number of proteins located in these subcellular locations that cannot be predicted. However, if a protein is predicted to be located in such a location, the prediction is most likely correct. Nonetheless, the accuracies for predicting these subcellular locations of fungal proteins need to be improved.

2.2 Overview of subcellular proteome distribution in different species

The database contains predicted subcellular location information of proteins generated from 16554 fungal species or varieties (strains) with 189 of them each having at least 1000 protein entries. The species names, some of which may have more than one strain or variety, can be found on the user interface, which facilitate species specific searching or downloading. Species having <1000 protein entries can also searched with a species name provided by the user. The distributions of subcellular proteomes in different fungal species are summarized in Table 2 and Supplementary Table 1. Table 2 includes the following subcellular locations: secreted proteins (4 subcategories), mitochondrial membrane and mitochondrial non-membrane, cytoplasm (cytosol), cytoskeleton, nuclear membrane and nuclear non-membrane, plasma membrane, and glycosylphosphatidylinositol (GPI) anchored proteins. The category of secreted proteins includes the following subcategories: curated secreted, highly likely secreted, likely secreted, and weakly secreted proteins. Information on other subcellular protein locations including endoplasmic reticulum (membrane or lumen), Golgi apparatus (membrane or lumen), lysosome, peroxisome, vacuole (membrane or non-membrane), other membrane, and other curated locations can be found in Supplementary Table 1.

The variability of genome sizes and thus the proteome sizes is pretty large in different fungal species. However, it should be noted that in the database, as showed in Table 2, the total proteins of a given species is not necessarily the proteome size, but rather a collection of all proteins available from the species. For example, for *Saccharomyces cerevisiae*, its reference proteome size as compiled UniProtKB consists only of 6,621 proteins, there are a total of 79,093 proteins in our database under the name of *Saccharomyces cerevisiae*, thus obviously consisting of proteins obtained from multiple strains. The subcellular distributions of fungal proteins were estimated based on the pooled data for each phylum for Ascomycota, Basidiomycota and Microsporidia. Interestingly, we found that the nucleus represents the largest compartment for protein destination: 39.2% in Ascomycota, 39.2% in Basidiomycota, and 57.4% in Microsporidia, respectively, were predicted to be located in the nucleus. Mitochondria represent another large compartment for protein targeting: 19.5% in Ascomycota, 21.1% in Basidiomycota, and 16.7% in Microsporidia, respectively, were located in mitochondria. Approximately 18 – 21% of proteins are located in cytosol or cytoplasm. The proportions of secretomes vary from 0.3% to 10.5% with an average of 4.6% in Ascomycota, from 1.9% to 7.4% with an average of 4.4% in Basidiomycota, and from 0.5% to 1.7% with an average of 0.9% in Microsporidia, respectively. However, here the secretome is limited to including curated secreted proteins and highly likely secreted proteins, thus the

number represents a lower bound of a species secretome. Including other proteins predicted as likely secreted and weakly likely secreted proteins, the size of secretome certainly will be significantly increased, but there would be an increase in the number of false positives, i.e., non-secreted proteins in the set.

2.3 Relationship of lifestyle and secretome size in different fungi

Similar to our previous analysis in FunSecKB work (Lum and Min, 2011), the secretome size (Y) was highly correlated with its proteome size (X) in a species (r = 0.87) with a regression as Y = 0.081X - 271. (Figure 1). However, species having different lifestyles showed differences in secretome size and proportion of secreted proteins. Lowe and Howlett (2012) examined the relationship between lifestyle and secretome size and found that fungi with biphasic lifestyle have a large proportion of secreted proteins and animal pathogens have fewer genes than saprophytes or plant interacting fungi do, and a lower proportion of predicted secreted. In the work of Lowe and Howlett (2012), the secretome prediction was only used SignalP, and thus, its size may be over estimated. Using the data we collected in this work, we examined the relationship between fungal lifestyles and their secretome sizes (Figure 1, Supplementary Table 2). As the data for each species in the database contain redundant or duplicated protein entries, we only used the proteins in datasets of reference or complete proteomes compiled by UniProt (http://www.uniprot.org/taxonomy/complete-proteomes). We collected species having a complete proteome and a lifestyle in the category of animal or/and human pathogen, plant pathogen, and saprophyte. Some of them may be classified into more than one category and these entries are annotated (see Supplementary Table 2). In general agreement with Lowe and Howlett (2012) reported, human and animal pathogens, including entomopathogens and some nematode killing fungal parasites have a relatively smaller proteome size – the majority of them have <12000 protein sequences, some of them are known as Microsporidian parasites having a genome encoding a total of 2000 - 4000 proteins, with less than 1% of them being secreted (Figure 1). The proportion of secreted proteins varied from 0.3 to 7.9% with an average of 2.8% in human/animal pathogens. On other hand, plant pathogens and saprophytes have much more variable proteome sizes from ~ 4000 to 18000 and a relatively higher proportion of secreted proteins, though variable, from 1.3 to 7.1% with an average of 4.2% in saprophytes and from 1.7 to 10.5% with an average of 6.3% in plant pathogens. Clearly, these results show that secretome size is one of the important determining factors in controlling fungal lifestyles. However, as species having a similar size of secretome may have different lifestyles, the composition within each secretome may play a more critical role in determining its lifestyle in each species.

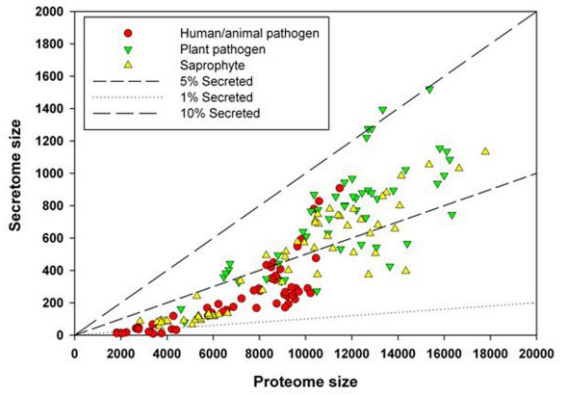

Figure 1 Relationship between proteome size and secretome size in fungal species having different lifestyles

2.4 Functional analysis of fungal secreted proteins

To provide an overview of the functionalities of all fungal secreted proteins, we carried out Gene Ontology (GO) analysis. The secreted protein set including curated and predicted highly likely secreted proteins only was used to search the UniProtKB/Swiss-Prot dataset with BLASTP with a cutoff E-value of 1e-10. GO information was retrieved from UniProt ID mapping data (http://www.uniprot.org/downloads) and analyzed using GO SlimViewer with generic GO terms (McCarthy et al., 2006). GO biological process and molecular function classification of the secretomes are summarized in Table 3. Molecular function classification revealed that fungal secreted proteins consist of a large number of hydrolases (~33.7%), proteins having ion binding

activity (21.1%), peptidase (15.7%), oxidoreducatases (14%), and some other enzymatic activities. Fungal secreted proteins are involved in more than 60 different biological processes. The main biological processes include catabolic process (24.6%), carbohydrate (22.0%) or lipid (4.0%) metabolic process, cell wall organization or biogenesis (6.4%), response to stress, small molecule and nitrogen metabolic process, etc. It should be noted that GO classification was only an estimate of the distributions of each category as ~54% of the predicted secreted proteins do not have GO annotation information.

Table 3 Gene Ontology (GO) classification of fungal secreted proteins

GO ID	Count	%	GO description
Molecular function			
GO:0016798	16132	30.9	hydrolase activity, acting on glycosyl bonds
GO:0043167	11011	21.1	ion binding
GO:0008233	8182	15.7	peptidase activity
GO:0016491	7305	14.0	oxidoreductase activity
GO:0016829	1710	3.3	lyase activity
GO:0016791	1439	2.8	phosphatase activity
GO:0016810	1242	2.4	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
GO:0016853	1010	1.9	isomerase activity
Others	4136	7.9	including 32 other GO categories
Total	52167		
Biological process			
GO:0009056	21356	24.6	catabolic process
GO:0005975	19039	22.0	carbohydrate metabolic process
GO:0071554	5584	6.4	cell wall organization or biogenesis
GO:0009058	3612	4.2	biosynthetic process
GO:0006629	3463	4.0	lipid metabolic process
GO:0006950	3405	3.9	response to stress
GO:0044281	3356	3.9	small molecule metabolic process
GO:0034641	3076	3.5	cellular nitrogen compound metabolic process
Others	23845	27.5	including 60 other GO categories
Total	86736		

We further categorized the functions of predicted secreted fungal proteins using the rpsBLAST tool to search the Pfam database with a cutoff E-value of 1e-10. Among a total of 93430 predicted secreted proteins, 43953 protein sequences have a Pfam match and a total of 880 protein families were detected. The summary of the Pfam analysis with 33 highly encoded secreted protein families in fungi is shown in Table 4. A complete list can be downloaded (http://proteomics.ysu.edu/publicaiton/data/). The top 10 highly encoded secreted protein families in fungi were eukaryotic aspartyl protease, carboxylesterase family, FAD binding domain containing family, subtilase family, glycosyl hydrolase family 61, glycosyl hydrolases family 28, glycosyl hydrolases family 18, GMC oxidoreductase, serine carboxypeptidase, and glycosyl hydrolase family 3. These proteases identified here such as aspartyl protease, subtilase, and other peptidase families are likely to be required for synergistic degradation of the proteins present in the various growth medium or substrate materials in the environments (Druzhinina et al. 2012; Girard et al. 2013). GO analysis and functional domain analysis are consistent in showing these proteins are mainly involved in biodegrading complex bio-molecules including carbohydrates, proteins, lipids, and other molecules.

Table 4 Highly encoded secreted protein families in fungi

Pfam ID	Members	%[a]	Pfam	Function
pfam00026	1473	3.4	Asp	Eukaryotic aspartyl protease
pfam00135	1419	3.2	COesterase	Carboxylesterase family
pfam01565	1395	3.2	FAD_binding_4	FAD binding domain
pfam00082	1279	2.9	Peptidase_S8	Subtilase family
pfam03443	1150	2.6	Glyco_hydro_61	Glycosyl hydrolase family 61
pfam00295	924	2.1	Glyco_hydro_28	Glycosyl hydrolases family 28
pfam00704	924	2.1	Glyco_hydro_18	Glycosyl hydrolases family 18
pfam05199	873	2.0	GMC_oxred_C	GMC oxidoreductase
pfam00450	845	1.9	Peptidase_S10	Serine carboxypeptidase
pfam00933	809	1.8	Glyco_hydro_3	Glycosyl hydrolase family 3 N terminal
pfam04389	695	1.6	Peptidase_M28	Peptidase family M28
pfam07732	651	1.5	Cu-oxidase_3	Multicopper oxidase
pfam00264	631	1.4	Tyrosinase	Common central domain of tyrosinase
pfam04616	591	1.3	Glyco_hydro_43	Glycosyl hydrolases family 43
pfam01083	569	1.3	Cutinase	Cutinase
pfam09286	519	1.2	Pro-kuma_activ	Pro-kumamolisin
pfam01522	486	1.1	Polysacc_deac_1	Polysaccharide deacetylase
pfam00150	454	1.0	Cellulase	Cellulase (glycosyl hydrolase family 5)
pfam09362	450	1.0	DUF1996	Domain of unknown function (DUF1996)
pfam00328	417	0.9	His_Phos_2	Histidine phosphatase superfamily (branch
pfam00840	410	0.9	Glyco_hydro_7	Glycosyl hydrolase family 7
pfam00188	400	0.9	CAP	Cysteine-rich secretory protein family
pfam01764	397	0.9	Lipase_3	Lipase (class 3)
pfam00544	393	0.9	Pec_lyase_C	Pectate lyase
pfam00331	381	0.9	Glyco_hydro_10	Glycosyl hydrolase family 10
pfam00457	377	0.9	Glyco_hydro_11	Glycosyl hydrolases family 11
pfam01055	366	0.8	Glyco_hydro_31	Glycosyl hydrolases family 31
pfam00246	348	0.8	Peptidase_M14	Zinc carboxypeptidase
pfam12708	337	0.8	Pectate_lyase_3	Pectate lyase superfamily protein
pfam07519	331	0.8	Tannase	Tannase and feruloyl esterase
pfam00722	325	0.7	Glyco_hydro_16	Glycosyl hydrolases family 16
pfam00394	303	0.7	Cu-oxidase	Multicopper oxidase
pfam13668	301	0.7	Ferritin_2	Ferritin-like domain

Note: [a] The percentage (%) was calculated based on a total of 43853 proteins having a Pfam match. The complete list can be downloaded (see text for details)

3 Discussion

We constructed the fungal protein subcellular location database and named it Fungal Secretome and Subcellular Proteome Knowledgebase (FunSecKB2). Comparing with FunSecKB (Lum and Min 2011), the number of total protein entries increased from 478,073 in FunSecKB to 1,976,832 in FunSecKB2, and the number of fungal species including different varieties and strains having a complete proteome increased from 52 in FunSecKB to 210 in FunSecKB2. The subcellular locations in FunSecKB2 were also expanded to include not only secretomes but also other subcellular locations including mitochondria, cytosol (cytoplasm), cytoskeleton, Golgi apparatus, lysosome, nucleus, peroxisome, plasma membrane and vacuole. In addition, for the secretomes, we further classified them as curated, predicted to be highly likely secreted, likely secreted, and weakly likely secreted protein subsets. This refinement of classifications of secreted proteins and other

subcellular locations would greatly enhance comparative analysis of subcellular proteomes in different species. However, as the protein sequence data were obtained from the UniProtKB and some duplicated entries are present, thus for proteome-wide analysis for a given species the non-redundant reference or complete proteome dataset needs to be used and that can be downloaded at UniProt (http://www.uniprot.org/taxonomy/complete-proteomes). It also should be noted that for a given species in the list if no specific strain or sub-genotype is specified, the entries for that specific species included all available proteins from the species.

We also provided the BLAST tool to allow users to search all fungal protein data or the predicted fungal secreted protein data with their own protein sequences. This utility facilitates identifying protein homologs with their potential subcellular locations. Otherwise, for any anonymous protein sequence users can predict protein subcelluar locations using the tools we have used in this work. Other available tools for prediction of secretomes and other protein subcellular locations were summarized by Meinken and Min (2012) and Caccia et al. (2013). Recently Cortázar et al. (2014) implemented a webserver, named SECRETOOL, which integrated several tools for predicting fungal secretomes. As some of the tools implemented in the server are the same tools as we used, we expect the server generates fairly reliable results for fungal secretome prediction, thus, it is particularly useful for newly generated proteomes (Cortázar et al., 2013; Lum and Min, 2011). In addition, another available database, named the fungal secretome database (FSD), which was constructed using a slightly different suite of tools, may provide extra subcellular location information for these fungal proteins (Choi et al., 2010).

Fungal species have a secretome adapted to their environment and the selection pressure exerted by environmental constrains led to the species with varying complexity in their secretome compositions (Girad et al., 2013; Alfaro et al., 2014). Depending on the lifestyle, fungal species which belong to saprotrophs mainly have degrading hydrolases in their secretomes, biotrophic species have both degrading hydrolases and compatibility effectors, mycorrhiza species have degrading hydrolases, compatibility effectors, and exchange effectors, and necrotrophic species have degrading hydrolases and killing effectors (Girad et al., 2013, Alfaro et al., 2014). The basal secretome contains generally two pools of proteins: a large proportion represented by the polysaccharide degrading enzymes, i.e. hydrolases acting on glycosyl bonds, and a minor part including the proteases, lipases, and oxidoreductases, etc. (see Table 3). In this work, the secretome identification was limited to classical secreted proteins, i.e., signal peptide containing proteins, and curated proteins which may include both classical and leadless secreted proteins (LSP). SecretomeP was a tool implemented for predicting these LSPs in bacteria and mammals (http://www.cbs.dtu.dk/services/SecretomeP/) (Bendtsen et al., 2004a). Because the tool has not been trained with fungal data and the prediction accuracy could not be evaluated, we did not include this tool in our data processing. We would like to request the fungal research community to submit fungal protein subcellular locations, particularly LSPs, with experimental evidence traceable from literature to the database. Genome-wide computational prediction of a secretome for a species provides the first step for experimental validation and characterization of secreted proteins under various changing environments or culture conditions (Alfaro et al., 2014). Along with our published plant secretome and subcellular proteome knowledgebase (PlantSecKB) (Lum et al., 2014), we expect that FunSecKB2 will serve the community a useful resource for genome-wide comparative analysis and for further exploring the potential applications of fungal secreted proteins in biofuel production, environmental remediation, and prevention and treatment of plant and human fungal pathogens.

Authors' contributions

JM implemented the database, DA collected the lifestyle data, KA and GZ participated in method development, XJM and CC conceived of the study, designed the procedure of data processing. XJM, JM, DA and CC analyzed the data and prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Gengkon Lum and Dr. Feng Yu for their assistance in maintaining the server and Jessica Orr and Stephanie Frazier for manually curating secreted proteins.

Funding

The work is funded by Youngstown State University (YSU) Research Council. The work is also supported by a YSU Research Professorship award and the College of Science, Technology, Engineering, and Mathematics Dean's reassigned time for research to XJM. JM was supported with a graduate research assistantship by the Center for Applied Chemical Biology at YSU.

References

Alfaro M., Oguiza J.A., Ramírez L. et al. 2014, Comparative analysis of secretomes in basidiomycete fungi, J Proteomics, 102C: 28-43
http://dx.doi.org/10.1016/j.jprot.2014.03.001

Bendtsen J.D., Jensen L.J., Blom N. et al. 2004a, Feature based prediction of non-classical and leaderless protein secretion, Protein Eng Des Sel, 17: 349-356
http://dx.doi.org/10.1093/protein/gzh037

Bendtsen J.D., Nielsen H., von Heijne, G. et al. 2004b, Improved prediction of signal peptides: SignalP 3.0, J Mol Biol, 340: 783-795
http://dx.doi.org/10.1016/j.jmb.2004.05.028

Bouws H., Wattenberg A. and Zorn H, 2008, Fungal secretomes-nature's toolbox for white biotechnology. Appl. Microbiol. Biotechnol. 80: 381-388
http://dx.doi.org/10.1007/s00253-008-1572-5

Braaksma M., Martens-Uzunova E.S., et al. 2010, An inventory of the *Aspergillus niger* secretome by combining in silico predictions with shotgun proteomics data, *BMC Genomics*, 19: 584
http://dx.doi.org/10.1186/1471-2164-11-584

Brown N.A., Antoniw J., and Hammond-Kosack K.E., 2012, The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis, PLoS One, 7: e33731
http://dx.doi.org/10.1371/journal.pone.0033731

Caccia D., Dugo M., Callari M., et al. (2013) Bioinformatics tools for secretome analysis, Biochim. Biophys. Acta., 1834: 2442-2453
http://dx.doi.org/10.1016/j.bbapap.2013.01.039

Choi J., Park J., Kim D., et al. 2010, Fungal secretome database: integrated platform for annotation of fungal secretomes, BMC Genomics, 11: 105
http://dx.doi.org/10.1186/1471-2164-11-105

Cortázar A.R., Aransay A.M., Alfaro M., et al. 2014, SECRETOOL: integrated secretome analysis tool for fungi, Amino Acids, 46: 471-473
http://dx.doi.org/10.1007/s00726-013-1649-z

de Castro E., Sigrist C.J., Gattiker A., et al. 2001 ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins, Nucleic Acids Res., 34: W362-365

Do Vale L.H., Gómez-Mendoza D.P., Kim M.S., et al. 2012, Secretome analysis of the fungus *Trichoderma harzianum* grown on cellulose, Proteomics, 12: 2716-2728
http://dx.doi.org/10.1002/pmic.201200063

Druzhinina I.S., Shelest E., and Kubicek C.P., 2012, Novel traits of Trichoderma predicted through the analysis of its secretome, FEMS Microbiol Lett., 337: 1-9
http://dx.doi.org/10.1111/j.1574-6968.2012.02665.x

Emanuelsson O., Brunak S., von Heijne G., et al. 2007, Locating proteins in the cell using TargetP, SignalP and related tools, Nat. Protoc., 2: 953-971
http://dx.doi.org/10.1038/nprot.2007.131

Ene I.V., Heilmann C.J., Sorgo A.G., et al. (2012) Carbon source-induced reprogramming of the cell wall proteome and secretome modulates the adherence and drug resistance of the fungal pathogen *Candida albicans*, Proteomics, 12: 3164-3179
http://dx.doi.org/10.1002/pmic.201200228

Girard V., Dieryckx C., Job C. et al. 2013, Secretomes: the fungal strike force, Proteomics, 13: 597-608
http://dx.doi.org/10.1002/pmic.201200282

Horton P., Park K.-J., Obayashi T., et al. 2007, WoLF PSORT: protein localization predictor, Nucleic Acids Res., 35: W585-587
http://dx.doi.org/10.1093/nar/gkm259

Jung Y.H., Jeong S.H., Kim S.H., et al. 2012, Secretome analysis of *Magnaporthe oryzae* using in vitro systems, Proteomics, 12: 878-900
http://dx.doi.org/10.1002/pmic.201100142

Käll L., Krogh A., and Sonnhammer E.L.L., 2007, Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server, Nucleic Acids Res., 35: W429-432
http://dx.doi.org/10.1093/nar/gkm256

Krogh A., Larsson B., von Heijne G., et al. 2001, Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes, J. Mol. Biol., 305: 567-580
http://dx.doi.org/10.1006/jmbi.2000.4315

Lange L., Bech L., Busk P.K., et al. 2012, The importance of fungi and of mycology for a global development of the bioeconomy, IMA Fungus, 3: 87-92
http://dx.doi.org/10.5598/imafungus.2012.03.01.09

Lee S.A., Wormsley S., Kamoun S., et al. 2003, An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms, Yeast, 20: 595-610
http://dx.doi.org/10.1002/yea.988

Lowe R.G., and Howlett B.J., 2012, Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi, PLoS pathogens, 8: e1002515
http://dx.doi.org/10.1371/journal.ppat.1002515

Lum G., and Min X.J., 2011, FunSecKB: the fungal secretome knowledgebase, *Database (Oxford)*, 2011, doi: 10.1093/database/bar001
http://dx.doi.org/10.1093/database/bar001

Lum G., and MinX.J., 2013, Bioinformatic protocols and the knowledge-base for secretomes in fungi, In: Gupta V.K., Tuohy M.G., Ayyachamy M., Turner K.M. and O'Donovan A. (eds), Laboratory Protocols in Fungal Biology: Current Methods in Fungal Biology, Springer, pp 545-557
http://dx.doi.org/10.1007/978-1-4614-2356-0_54

Lum G., Meinken J., Orr J., et al. 2014, PlantSecKB: the plant secretome and subcellular proteome knowledgebase. Comput. Mole. Biol., 4: 1-17

Martinez D., Challacombe J., Morgenstern I., et al. 2009, Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion, Proc Natl Acad Sci U S A,106: 1954-1959
http://dx.doi.org/10.1073/pnas.0809575106

Martinez D., Larrondo L.F., Putnam N., et al. 2004, Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78, Nat Biotechnol. 22: 695-700
http://dx.doi.org/10.1038/nbt967
http://dx.doi.org/10.1038/nbt0704-899b
http://dx.doi.org/10.1038/nbt0704-899a

McCarthy F.M., Wang N., Magee G.B., et al. 2006, AgBase: a functional genomics resource for agriculture, BMC Genomics, 7: 229
http://dx.doi.org/10.1186/1471-2164-7-229

Meinken J., and Min X.J., 2012, Computational prediction of protein subcellular locations in eukaryotes: an experience report, Comput. Mole. Biol., 2: 1-7

Melhem H., Min X.J., and Butler G., 2013, The impact of SignalP 4.0 on the prediction of secreted proteins. IEEE Symposium Series on Computational Intelligence (IIEEE SSCI 2013): The 10th annual IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, Singapore, pp.16-22 (doi: 10.1109/CIBCB.2013.6595383)

Min X.J., 2010, Evaluation of computational methods for secreted protein prediction in different eukaryotes, J. Proteomics Bioinform., 3: 143-147

Morais do Amaral A., Antoniw J., Rudd J.J., et al. 2012, Defining the predicted protein secretome of the fungal wheat leaf pathogen *Mycosphaerella graminicola*, PLoS One. 7: e49904
http://dx.doi.org/10.1371/journal.pone.0049904

Mueller O., Kahmann R., Aguilar G., et al. 2008, The secretome of the maize pathogen *Ustilago maydis*, Fungal Genet. Biol., 1: S63-S70
http://dx.doi.org/10.1016/j.fgb.2008.03.012

Murphy C., Powlowski J., Wu M., et al. 2011, Curation of characterized glycoside hydrolases of fungal origin, Database (Oxford). 2011, doi: 10.1093/database/bar020
http://dx.doi.org/10.1093/database/bar020

Paper J.M., Scott-Craig J.S., Adhikari N.D., et al. 2007, Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus *Fusarium graminearum*, Proteomics, 7: 3171-3183
http://dx.doi.org/10.1002/pmic.200700184

PetersenT.N., Brunak S., von Heijne G., et al. 2011, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nature Methods, 8: 785-786
http://dx.doi.org/10.1038/nmeth.1701

Peterson R., Grinyer J., and Nevalainen H., 2011, Secretome of the coprophilous fungus *Doratomyces stemonitis* C8, isolated from koala feces, Appl. Environ. Microbiol., 77: 3793-3801
http://dx.doi.org/10.1128/AEM.00252-11

Poisson G., Chauve C., Chen X., et al. 2007, FragAnchor a large scale all Eukaryota predictor of Glycosylphosphatidylinositol-anchor in protein sequences by qualitative scoring, Genomics Proteomics Bioinform., 5: 121-130
http://dx.doi.org/10.1016/S1672-0229(07)60022-9

Powers-Fletcher M.V., Jambunathan K., Brewer J.L., et al. 2011, Impact of the lectin chaperone calnexin on the stress response, virulence and proteolytic secretome of the fungal pathogen *Aspergillus fumigatus*, PLoS One, 6: e28865
http://dx.doi.org/10.1371/journal.pone.0028865

Ribeiro D.A., Cota J., Alvarez T.M., et al. 2012, The *Penicillium echinulatum* secretome on sugar cane bagasse, PloS One, 7: e50571
http://dx.doi.org/10.1371/journal.pone.0050571

Salvachúa D., Martínez A.T., Tien M, et al. 2013, Differential proteomic analysis of the secretome of Irpex lacteus and other white-rot fungi during wheat straw pretreatment, Biotechnol. Biofuels. 6: 115
http://dx.doi.org/10.1186/1754-6834-6-115

Sigrist C.J.A., Cerutti L., de Casro E., et al. 2010, PROSITE, a protein domain database for functional characterization and annotation, Nucleic Acids Res., 38: 161-166
http://dx.doi.org/10.1093/nar/gkp885

The UniProt Consortium, 2014, Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res., 42:D191-198
http://dx.doi.org/10.1093/nar/gkt1140

Tjalsma H., Bolhuis A., Jongbloed J.D., et al. 2000, Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome, Microbiol. Mol. Biol. Rev., 64: 515-547
http://dx.doi.org/10.1128/MMBR.64.3.515-547.2000

Tsang A., Butler G., Powlowski J., et al. 2009, Analytical and computational approaches to define the *Aspergillus niger* secretome, Fungal Genetics Biol., 46:S153-160
http://dx.doi.org/10.1016/j.fgb.2008.07.014

Weber S.S., Parente A.F.A., Borges C.L., et al. 2012, Analysis of the secretomes of Paracoccidioides mycelia and yeast cells, PloS One, 7: e52470
http://dx.doi.org/10.1371/journal.pone.0052470

Wymelenberg A.V., Sabat G., Martinez D., et al. 2005, The *Phanerochaete chrysosporium* secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium, J. Biotechnol., 118: 17-34
http://dx.doi.org/10.1016/j.jbiotec.2005.03.010

Yajima W., and Kav N.N., 2006, The proteome of the phytopathogenic fungus *Sclerotinia sclerotiorum*, Proteomics, 6: 5995-6007
http://dx.doi.org/10.1002/pmic.200600424

Table 2 Summary of some major subcellular locations of proteins in different fungal different species. Data of other subcellular locations of fungal proteins are in Supplementary Table 1.

	Total	CS	HLS	LS	WLS	Mt mem	Mt non-m	Cyp	Cyk	Nuc mem	Nuc non-m	Plas mem	GPI	Sec (%)
Ascomycota														
Ajellomyces capsulata	37457	6	895	550	1760	565	8242	7068	1048	304	15640	3613	148	2.4
Ajellomyces dermatitidis	29246	7	853	477	1396	444	6712	5293	794	249	11594	3020	156	2.9
Arthrobotrys oligospora	11491	2	908	364	460	140	1754	2441	281	114	4476	1267	121	7.9
Arthroderma benhamiae	8067	56	248	143	377	142	1211	1643	234	146	3240	1050	39	3.8
Arthroderma gypseum	8918	27	380	151	396	117	1631	1763	264	97	3496	1098	49	4.6
Arthroderma otae	8813	53	312	135	392	125	1625	1724	308	78	3455	1102	47	4.1
Ashbya gossypii	9553	1	163	124	338	255	1934	1967	192	103	4023	976	67	1.7
Ashbya gossypii FDAG1	4762	0	83	68	184	97	1020	908	89	49	2009	532	34	1.7
Aspergillus clavatus	9182	53	471	170	438	169	1688	1921	273	74	3128	1221	70	5.7
Aspergillus flavus	14041	88	825	267	721	192	2565	3098	596	93	4366	1956	77	6.5
Aspergillus kawachii	11506	11	732	189	522	162	1936	2492	389	82	3865	1696	84	6.5
Aspergillus niger	25597	295	1261	467	1351	376	4988	5450	880	192	8150	3581	136	6.1
Aspergillus oryzae	23947	100	1464	487	1294	358	4194	5372	881	173	7200	3669	126	6.5
Aspergillus terreus	10550	75	649	200	503	162	1823	2287	336	71	3367	1566	60	6.9
Baudoinia compniacensis UAMH 10762	10508	0	374	163	573	160	2672	2288	299	69	3373	1087	50	3.6
Beauveria bassiana	10798	2	806	293	554	167	2026	2376	318	80	3272	1411	116	7.5
Bipolaris maydis ATCC 48331	12705	0	896	275	659	158	2452	2593	309	113	4345	1615	92	7.1
Bipolaris maydis C5	12857	0	880	266	655	178	2491	2648	320	117	4432	1584	93	6.8
Bipolaris sorokiniana ND90Pr	12174	0	847	274	580	195	2413	2426	282	113	4132	1559	92	7.0
Blumeria graminis f. sp. hordei DH14	6459	0	359	314	296	101	1210	942	136	80	3010	680	19	5.6
Botryotinia fuckeliana	27965	6	1523	523	1193	330	4823	5467	870	231	11560	3034	169	5.5
Botryotinia fuckeliana BcDW1	11018	0	719	215	405	140	1513	2347	305	106	4450	1367	85	6.5
Candida albicans	16194	58	454	304	663	581	2085	2366	79	479	8164	1597	184	3.2
Candida dubliniensis	5896	0	169	108	169	140	699	855	24	160	3221	575	77	2.9
Candida glabrata	5492	7	101	75	150	160	607	1048	93	117	2895	480	76	2.0
Candida maltosa Xu316	5976	0	194	82	143	117	574	1118	29	135	3210	575	56	3.2
Candida orthopsilosis	5758	0	143	93	177	118	714	932	36	146	3028	641	49	2.5
Candida parapsilosis	5920	2	175	106	167	132	654	1055	50	134	3069	677	75	3.0
Candida tenuis	6052	0	116	74	205	98	759	1229	52	115	3025	717	42	1.9
Candida tropicalis	6413	1	194	129	193	157	716	1022	30	167	3356	697	79	3.0
Chaetomium globosum	11080	2	779	237	579	177	2618	2379	253	59	3334	1168	67	7.0

Continued Table 2

	Total	CS	HLS	LS	WLS	Mt mem	Mt non-m	Cyp	Cyk	Nuc mem	Nuc non-m	Plas mem	GPI	Sec (%)
Chaetomium thermophilum	7237	0	346	147	363	132	1490	1491	138	61	2583	867	61	4.8
Claviceps purpurea 20.1	8807	0	494	158	360	140	1689	1818	250	64	3430	865	59	5.6
Clavispora lusitaniae	6006	0	138	101	310	104	1014	1100	87	98	2634	747	43	2.3
Coccidioides immitis	9773	2	278	153	418	143	2311	1716	279	91	3904	1022	61	2.9
Coccidioides posadasii	17595	33	483	270	721	252	3877	3299	478	175	7008	1993	113	2.9
Colletotrichum gloeosporioides	15636	5	1531	330	734	163	2342	3925	485	101	4362	2358	99	9.8
Colletotrichum graminicola	12268	6	1025	295	607	179	2202	2775	313	69	3579	1770	131	8.4
Colletotrichum higginsianum	16264	0	1217	391	909	188	3261	3709	493	71	4862	2022	74	7.5
Colletotrichum orbiculare	13358	0	1399	302	629	211	2456	2887	329	90	3679	1974	126	10.5
Coniosporium apollinis CBS 100218	9306	0	386	137	396	142	1854	2204	256	55	3140	1149	68	4.1
Cordyceps militaris	9744	2	548	269	578	156	2143	1997	221	66	2830	1384	93	5.6
Debaryomyces hansenii	6331	1	129	73	141	163	747	1134	92	143	3344	618	46	2.1
Dekkera bruxellensis AWRI1499	4853	0	52	48	146	64	583	1224	67	67	2287	356	2	1.1
Dothistroma septosporum NZE10	12414	0	560	234	568	154	2614	2675	368	78	4457	1313	82	4.5
Emericella nidulans	13302	100	687	224	582	211	2398	2835	348	124	4552	1844	83	5.9
Eremothecium cymbalariae	4444	0	67	54	136	90	670	732	54	73	2361	446	23	1.5
Eutypa lata UCREL1	11682	0	945	268	596	138	1771	3350	437	56	3226	1435	84	8.1
Exophiala dermatitidis	9426	0	295	147	468	144	1835	1954	203	76	3465	1274	78	3.1
Fusarium oxysporum	64825	5	4398	1338	2851	640	9745	15590	2293	407	22169	8930	433	6.8
Fusarium oxysporum f. sp. cubense race 1	15345	0	1053	319	662	154	2321	3711	527	107	5158	2216	126	6.9
Fusarium oxysporum f. sp. cubense race 4	14147	0	984	294	608	141	2173	3381	491	86	4693	2049	104	7.0
Fusarium oxysporum f. sp. lycopersici	16735	4	1066	360	795	170	2621	4086	674	85	5695	2192	101	6.4
Fusarium pseudograminearum	12530	1	879	276	561	161	1826	2841	332	106	4440	1753	126	7.0
Gaeumannomyces graminis var. tritici	14634	1	1035	410	980	273	3995	2766	291	86	4065	1527	145	7.1
Geomyces destructans	9178	1	265	143	415	140	1992	2044	250	62	3419	955	60	2.9
Gibberella zeae	13576	4	888	291	668	188	2059	3122	415	102	4859	1773	113	6.6
Glarea lozoyensis	7907	0	333	108	347	79	1266	2383	358	22	2830	691	34	4.2
Grosmannia clavigera	8394	2	346	159	553	136	1695	2015	213	47	2440	1115	74	4.1
Hypocrea atroviridis	11922	2	700	254	638	145	2171	2794	385	84	3926	1482	71	5.9
Hypocrea jecorina	9359	14	521	169	474	142	1737	2183	243	74	3006	1220	69	5.7
Hypocrea virens	12537	2	748	233	617	139	2174	3068	447	112	4012	1608	84	6.0
Kazachstania africana	5359	0	113	76	143	114	636	780	40	130	3030	514	49	2.1

Continued Table 2

	Total	CS	HLS	LS	WLS	Mt mem	Mt non-m	Cyp	Cyk	Nuc mem	Nuc non-m	Plas mem	GPI	Sec (%)
Kazachstania naganishii	5304	0	100	72	178	79	803	1026	61	76	2604	575	41	1.9
Kluyveromyces lactis	5243	6	92	52	137	150	636	999	73	105	2729	494	40	1.9
Komagataella pastoris	10303	0	194	138	289	189	1273	1742	112	225	5614	1068	97	1.9
Lachancea thermotolerans	5105	0	110	56	161	101	832	954	71	64	2457	576	38	2.2
Leptosphaeria maculans	12742	5	732	295	781	197	3215	2092	334	96	4400	1389	63	5.8
Lodderomyces elongisporus	5794	0	120	77	197	168	716	1008	35	176	2978	607	46	2.1
Macrophomina phaseolina	13813	0	898	266	741	215	2855	2986	335	83	4276	1795	73	6.5
Magnaporthe oryzae	39211	10	3935	1217	2145	566	8761	7298	881	254	11673	4382	385	10.1
Magnaporthe oryzae P131	12711	1	1274	404	683	172	2820	2378	283	83	3767	1434	125	10.0
Magnaporthe oryzae Y34	12858	1	1274	405	690	178	2844	2429	291	81	3841	1442	125	9.9
Magnaporthe poae	11326	1	857	311	700	199	2783	2178	266	48	3232	1279	90	7.6
Marssonina brunnea f. sp. multigermtubi	10034	1	611	267	451	203	2048	1961	235	92	3482	1233	76	6.1
Metarhizium acridum	9870	1	610	243	482	154	1876	2171	272	72	3195	1320	90	6.2
Metarhizium anisopliae	10860	3	892	323	544	158	1962	2413	283	75	3292	1453	103	8.2
Meyerozyma guilliermondii	5945	0	142	109	240	121	838	1117	70	108	2733	800	45	2.4
Mycosphaerella graminicola	11258	1	630	239	510	160	2156	2654	408	79	3761	1264	63	5.6
Mycosphaerella populorum SO2202	10152	0	568	208	436	134	2042	2127	261	92	3679	1122	72	5.6
Naumovozyma castellii	5650	0	106	75	152	102	700	880	34	123	3120	581	46	1.9
Naumovozyma dairenensis	5536	0	91	74	126	111	632	806	31	169	3213	516	41	1.6
Nectria haematococca	15790	3	937	360	790	178	2482	3970	520	106	4855	2440	137	6.0
Neofusicoccum parvum UCRNP2	10360	0	869	184	537	130	1809	2711	242	42	2779	1487	74	8.4
Neosartorya fischeri	10452	77	624	165	466	173	1829	2292	287	81	3505	1434	76	6.7
Neosartorya fumigata	20377	171	1041	318	1077	365	3908	4085	577	162	6745	2864	154	5.9
Neurospora crassa	13456	13	691	231	631	300	2617	2778	292	104	5120	1353	93	5.2
Neurospora tetrasperma	21639	0	1073	360	1070	303	4778	4171	505	160	8149	2161	149	5.0
Paracoccidioides brasiliensis	26076	8	562	373	1233	435	5494	4937	753	236	10921	2709	113	2.2
Penicillium chrysogenum	13109	11	649	205	623	206	2450	2695	469	102	4641	1707	96	5.0
Penicillium digitatum	18148	1	687	262	794	275	3343	3882	586	152	6827	2297	122	3.8
Penicillium marneffei	10652	3	484	159	443	153	1590	2403	318	105	4016	1447	74	4.6
Phaeosphaeria nodorum	16124	8	1018	349	769	198	3220	3448	515	115	5646	1734	82	6.4
Pichia angusta	4418	1	90	45	128	71	527	1030	35	87	2164	465	27	2.1
Pichia sorbitophila	8851	0	174	123	239	145	1194	1411	42	199	4790	951	64	2.0

Continued Table 2

	Total	CS	HLS	LS	WLS	Mt mem	Mt non-m	Cyp	Cyk	Nuc mem	Nuc non-m	Plas mem	GPI	Sec (%)
Pneumocystis jiroveci	3662	0	17	51	138	77	547	456	69	88	2229	271	4	0.5
Pneumocystis murina B123	3761	0	11	48	75	77	515	370	49	106	2407	305	4	0.3
Podospora anserina	10959	3	762	245	588	201	2316	2334	248	81	3377	1348	97	7.0
Pseudocercospora fijiensis CIRAD86	13062	0	543	254	718	191	2752	2757	385	100	4572	1525	64	4.2
Pyrenophora teres f. teres	11765	3	801	219	457	152	2005	2570	313	100	4388	1372	79	6.8
Pyrenophora tritici-repentis	12106	5	857	255	523	170	2140	2627	315	93	4400	1394	81	7.1
Saccharomyces arboricola	3655	0	56	44	106	63	475	615	23	77	1997	387	34	1.5
Saccharomyces cerevisiae	79093	120	1531	947	2266	1853	10305	14841	978	1667	40230	7642	500	2.1
Saccharomyces cerevisiae CEN.PK113-7D	5438	0	101	62	142	84	679	861	41	126	3011	583	44	1.9
Saccharomyces cerevisiae x S. kudriavzevii VIN7	9076	0	190	120	248	169	1118	1661	93	199	4779	971	53	2.1
Saccharomyces kudriavzevii	3820	0	78	51	120	76	529	650	48	86	2050	363	29	2.0
Scheffersomyces stipitis	5835	0	131	60	196	112	620	1225	44	116	2975	665	44	2.2
Schizophyllum commune	13269	7	686	213	720	174	2948	3071	342	72	4361	1367	76	5.2
Schizosaccharomyces japonicus	4807	0	90	65	169	81	745	900	81	61	2416	531	25	1.9
Schizosaccharomyces pombe	5165	45	20	10	35	220	541	1966	194	96	2407	34	24	1.3
Sclerotinia sclerotiorum	14845	5	574	220	636	164	2719	2941	606	114	6101	1338	65	3.9
Setosphaeria turcica Et28A	11687	0	798	270	575	143	2262	2296	264	85	4100	1471	83	6.8
Sordaria macrospora	10047	2	642	158	447	176	1815	2080	211	84	3797	1100	80	6.4
Spathaspora passalidarum	5979	0	174	97	164	92	735	1106	37	127	3122	618	90	2.9
Talaromyces stipitatus	13036	1	505	190	605	148	2037	2736	394	123	5251	1755	55	3.9
Taphrina deformans PYCC 5710	4618	0	162	69	187	63	775	1074	125	34	1885	545	42	3.5
Tetrapisispora blattae	5385	0	72	86	123	123	578	695	27	182	3171	531	29	1.3
Tetrapisispora phaffii	5245	0	77	67	114	133	632	768	22	141	3038	463	41	1.5
Thielavia heterothallica	9095	4	481	186	451	145	2042	2012	178	64	2929	1024	75	5.3
Thielavia terrestris	9761	1	567	197	594	159	2302	2187	170	60	2833	1129	76	5.8
Togninia minima UCRPA7	8833	0	443	147	420	100	1312	2770	305	42	2433	1258	73	5.0
Torulaspora delbrueckii	4996	2	99	54	119	86	686	876	55	73	2634	557	37	2.0
Trichophyton equinum	8703	17	343	143	379	121	1680	1641	264	89	3557	1001	53	4.1
Trichophyton rubrum	8814	20	369	144	387	146	1707	1625	245	88	3548	1055	49	4.4
Trichophyton tonsurans	8556	18	351	126	393	135	1584	1621	270	87	3468	1009	49	4.3
Trichophyton verrucosum	8050	49	231	153	393	131	1232	1636	237	140	3256	1020	31	3.5

Continued Table 2

	Total	CS	HLS	LS	WLS	Mt mem	Mt non-m	Cyp	Cyk	Nuc mem	Nuc non-m	Plas mem	GPI	Sec (%)
Trichosporon asahii var. asahii	16824	0	853	268	880	231	3865	3943	349	118	5344	1815	190	5.1
Tuber melanosporum	7530	0	222	136	360	141	1812	1628	226	48	2461	841	50	2.9
Uncinocarpus reesii	7770	9	268	131	316	98	1452	1623	212	82	3120	885	44	3.6
Vanderwaltozyma polyspora	5370	1	98	67	127	148	586	870	50	140	3069	450	33	1.8
Verticillium albo-atrum	10277	4	762	230	544	156	2157	2169	253	59	3222	1294	88	7.5
Verticillium dahliae	10780	1	784	256	575	167	2210	2309	251	67	3308	1371	108	7.3
Wickerhamomyces ciferrii	6725	0	246	89	154	127	648	1032	24	205	3736	717	59	3.7
Yarrowia lipolytica	6594	4	273	123	258	169	985	1615	131	76	2538	758	106	4.2
Zygosaccharomyces rouxii	5446	0	107	58	134	112	766	909	71	88	2888	582	30	2.0
Basidiomycota														
Agaricus bisporus var. bisporus	10409	0	512	153	422	110	1784	2111	329	79	4282	1221	63	4.9
Agaricus bisporus var. burnettii	11211	0	508	155	492	115	2065	2169	381	88	4656	1218	58	4.5
Auricularia delicata	5290	0	241	118	411	49	1525	1204	146	23	1509	380	29	4.6
Ceriporiopsis subvermispora B	12078	0	510	233	659	136	2787	2496	406	55	3909	1495	67	4.2
Coniophora puteana	1026	0	65	19	48	5	197	242	42	3	335	131	3	6.3
Coprinopsis cinerea	13534	8	887	225	523	173	2446	2639	348	124	5326	1545	108	6.6
Cryptococcus gattii serotype B	6560	0	150	98	313	107	1462	1346	180	56	2412	804	46	2.3
Cryptococcus neoformans var. grubii serotype A	6977	1	174	107	291	136	1566	1482	181	65	2506	884	55	2.5
Cryptococcus neoformans var. neoformans serotype D	13006	2	298	201	617	245	2857	2817	297	138	4592	1686	101	2.3
Dacryopinax sp.	10232	0	445	197	599	142	2355	2358	420	44	3006	1233	67	4.3
Dichomitus squalens	7187	0	413	135	412	74	1607	1628	224	30	2196	833	54	5.7
Fibroporia radiculosa	9251	0	401	183	484	125	1889	1959	288	59	3072	1215	57	4.3
Laccaria bicolor	17929	0	555	359	1180	211	4111	3325	586	124	6855	1856	79	3.1
Malassezia globosa	4282	0	119	61	177	85	999	811	159	42	1654	445	14	2.8
Malassezia sympodialis ATCC 42132	3400	0	66	40	183	58	971	752	118	18	1050	373	9	1.9
Melampsora larici-populina	16255	0	1086	604	649	155	2397	2151	298	222	8829	1120	68	6.7
Mixia osmundae	6727	0	442	177	432	127	1426	1071	118	91	2421	833	55	6.6
Moniliophthora perniciosa	13703	0	452	165	857	162	2502	3840	734	39	4905	1080	11	3.3
Phanerochaete carnosa	13868	0	658	297	868	168	3079	3159	467	51	4258	1705	67	4.7
Piriformospora indica	11824	0	594	200	514	168	2211	2178	401	88	4573	1386	80	5.0
Postia placenta	9164	0	332	251	611	101	2091	2016	318	62	2863	1156	19	3.6
Pseudozyma antarctica T-34	6640	0	343	171	513	109	1534	1031	69	70	2334	816	56	5.2
Pseudozyma hubeiensis SY62	7472	0	252	156	511	102	1799	1204	105	76	2949	796	29	3.4
Puccinia graminis f. sp. tritici	15837	0	1171	683	712	144	2693	1920	282	183	8347	1063	73	7.4
Puccinia triticina	11560	0	534	230	579	113	2137	1932	198	112	5703	735	49	4.6

Continued Table 2

	Total	CS	HLS	LS	WLS	Mt mem	Mt non-m	Cyp	Cyk	Nuc mem	Nuc non-m	Plas mem	GPI	Sec (%)
Punctularia strigosozonata	2096	0	138	37	139	20	432	480	68	3	685	213	17	6.6
Rhizoctonia solani AG-1 IA	10499	0	273	198	739	160	3167	1858	375	59	3205	1180	26	2.6
Rhizoctonia solani AG-1 IB	12197	0	773	161	608	101	2139	3193	507	45	4363	1106	46	6.3
Rhodosporidium toruloides NP11	8135	0	277	147	528	152	2156	1630	108	63	2610	960	80	3.4
Rhodotorula glutinis	2872	0	85	66	187	65	792	501	43	32	968	315	23	3.0
Serpula lacrymans var. lacrymans	27064	0	768	435	1580	277	5886	5666	862	136	10783	2728	90	2.8
Sporisorium reilianum	6717	0	405	165	462	114	1422	1193	73	69	2407	791	65	6.0
Stereum hirsutum	1617	0	50	29	103	12	311	363	71	9	597	179	10	3.1
Trametes versicolor	1095	3	73	13	70	14	272	257	36	4	328	75	2	6.9
Tremella mesenterica	1470	0	28	23	66	17	288	251	52	10	702	141	4	1.9
Ustilago hordei	7189	0	329	158	442	100	1461	1351	113	90	2869	745	47	4.6
Ustilago maydis	6929	2	436	174	440	147	1361	1120	72	79	2729	788	55	6.3
Wallemia ichthyophaga EXF-994	4834	0	152	67	166	54	669	866	56	57	2505	514	28	3.1
Wallemia sebi	5268	0	153	77	149	77	670	1055	55	89	2701	520	27	2.9
Chytridiomycota														
Batrachochytrium dendrobatidis	8623	0	450	347	418	97	1197	1680	257	90	3934	928	62	5.2
Microsporidia														
Edhazardia aedis	4210	0	37	245	242	255	572	414	12	200	2416	241	7	0.9
Encephalitozoon cuniculi	3857	2	24	100	107	35	481	897	239	47	1971	450	5	0.7
Encephalitozoon hellem	1878	0	16	48	39	28	196	416	103	24	1051	179	1	0.9
Encephalitozoon intestinalis	1853	0	15	46	33	30	181	390	85	28	1079	182	2	0.8
Encephalitozoon romaleae	1867	0	12	32	53	33	217	407	78	32	1057	185	0	0.6
Enterocytozoon bieneusi	3317	0	17	43	76	155	573	736	78	117	1689	133	0	0.5
Nematocida parisii	5383	0	84	197	305	177	712	804	101	275	2732	512	3	1.6
Nematocida sp. 1	2769	0	35	136	139	69	389	484	91	82	1341	311	3	1.3
Nosema bombycis CQ1	4398	0	33	265	179	174	572	822	27	155	2437	234	3	0.8
Nosema ceranae	2065	0	11	88	74	99	248	382	6	68	1212	97	2	0.5
Trachipleistophora hominis	3220	0	21	121	128	42	533	590	35	61	1837	212	3	0.7
Vavraia culicis	2774	0	46	131	137	74	427	414	25	71	1514	247	1	1.7
Vittaforma corneae	2237	0	17	74	63	103	257	397	22	53	1312	202	7	0.8
Zygomycota														
Rhizopus delemar	16998	0	414	187	467	171	2653	2813	308	202	9648	1268	60	2.4
Total for all Species	1976832	1922	91482	37240	91610	33519	358594	420973	53796	20008	756436	218998	12601	4.7

Note: Abbreviation: CS: curated secreted protein; HLS: highly likely secreted; LS: likely secreted; WLS: weakly likely secreted; Mt mem: mitochondrial membrane; Mt non-m: mitochondrial non-membrane; Cyp: cytoplasm (or cytosol); Cyk: cytoskeleton; Nuc mem: nuclear membrane; Nuc non-m: nuclear non-membrane; Pla mem: plasma membrane; GPI: glycosylphosphatidyinositol anchored; Sec: secretome.