Fungal Secretome Database: Integrated platform for annotation of fungal secretomes
Jaeyoung Choi1,2,3, Jongsun Park1,2,3,4, Donghan Kim1,2,3, Kyongyong Jung1,2,3, Seogchan Kang6, Yong-Hwan Lee1,2,3,4,5*

Abstract
Background: Fungi secrete various proteins that have diverse functions. Prediction of secretory proteins using only one program is unsatisfactory. To enhance prediction accuracy, we constructed Fungal Secretome Database (FSD).

Description: A three-layer hierarchical identification rule based on nine prediction programs was used to identify putative secretory proteins in 158 fungal/oomycete genomes (208,883 proteins, 15.21% of the total proteome). The presence of putative effectors containing known host targeting signals such as RXLX [EDQ] and RXLR was investigated, presenting the degree of bias along with the species. The FSD’s user-friendly interface provides summaries of prediction results and diverse web-based analysis functions through Favorite, a personalized repository.

Conclusions: The FSD can serve as an integrated platform supporting researches on secretory proteins in the fungal kingdom. All data and functions described in this study can be accessed on the FSD web site at http://fsd.snu.ac.kr/.

Background
The “secretome” refers to the collection of proteins that contain a signal peptide and are processed via the endoplasmic reticulum and Golgi apparatus before secretion [1]. In organisms from bacteria to humans, secretory proteins are common and perform diverse functions. These functions include immune system [2], roles as neurotransmitters in the nervous system [3], roles as hormones/pheromones [4], acquisition of nutrients [5-7], building and remodeling of cell walls [8], signaling and environmental sensing [9], and competition with other organisms [10-13]. Some secretory proteins in pathogens function as effectors that manipulate and/or destroy host cells with special signatures. In Plasmodium and Phytophthora species, effectors carry the RXLX [EDQ] or RXLR motifs as host targeting signals [11-13]. With the aid of advanced genome sequencing technologies [14], the rapid increase of sequenced fungal genomes offers many opportunities to study the function and evolution of secretory proteins at the genome level [15,16]. The Comparative Fungal Genomics Platform (CFGP; http://cfgp.snu.ac.kr/) [16] now archives 235 genomes from 120 fungal/oomycete species. The accurate prediction of secretory proteins in sequenced genomes is the key to realizing such opportunities.

The widely used SignalP 3.0 program [17] detected 89.81% of the 2,512 experimentally verified sequences in SPdb [18], a database containing proteins with signal peptides. To improve the accuracy of prediction, we built a hierarchical identification pipeline based on nine prediction programs (Table 1). Through this pipeline, putative secretory proteins, including pathogen effectors, encoded by 158 fungal and oomycete genomes were identified. The Fungal Secretome Database (FSD; http://fsd.snu.ac.kr/) was established to support not only the archiving of fungal secretory proteins but also the management and use of the resulting data. The FSD also has a user-friendly web interface and offers several data analysis functions via Favorite, a personalized data repository implemented in the CFGP (http://cfgp.snu.ac.kr/)[16].

Construction and content
Evaluation of the pipeline for predicting secretory proteins
To evaluate the capabilities of four programs SignalP 3.0 [17], SigCleave [19], SigPred [20], and RPSP [21] for

* Correspondence: yonglee@snu.ac.kr
1 Fungal Bioinformatics Laboratory, Seoul National University, Seoul 151-921, Korea
© 2010 Choi et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
predicting signal peptides, we analyzed the secretory proteins collected in SPdb [18]. SignalP 3.0 identified 89.81% of 2,512 proteins; while adding the other three programs, in combination, 87.50% of the proteins, which were not predicted by SignalP 3.0, were identified. The remaining proteins (1.31% of 2,512 proteins) were investigated by using two programs that predicted subcellular localization: PSort II [22] and TargetP 1.1b [23]. We found that 34.38% of the proteins were predicted to be extracellular proteins, increasing the coverage to 99.16%.

For the 1,093 characterized fungal/oomycete secretory proteins (Table 2), the combinatory pipeline raised the prediction coverage from 75.30% to 84.17% in comparison to SignalP 3.0. In addition, 98.14% of 24,921 experimentally unverified sequences in the SPdb were predicted as secretory proteins by the pipeline, while SignalP 3.0 caught 80.22% of them as positive. To assess robustness of the pipeline with non-secretory proteins, we prepared yeast proteins localized in cytosol, endoplasmic reticulum, nucleus, or mitochondrion [24]. When the 1,955 proteins were subjected to the FSD pipeline and SignalP 3.0, the numbers of false positives were almost same (84 and 82, respectively). Together, these results suggest that this ensemble approach could

| Prediction Program | Description | Ref |
|--------------------|-------------|-----|
| SignalP 3.0        | A program to predict whether a protein has the signal peptidase site I or not | [17] |
| SigCleave          | A program to predict whether a protein has signal peptides or not | [19] |
| SigPred            | A program to predict whether a protein has signal peptides or not | [20] |
| RPSP               | A program to predict whether a protein has signal peptides or not | [21] |
| TMHMM 2.0c         | A program to predict whether a protein has trans-membrane helix(es) or not | [26] |
| TargetP 1.1b       | A program to predict a site where a protein probably resides | [23] |
| PSort II           | A program to predict a site where a protein probably resides | [22] |
| SecretomeP 1.0f    | A program to predict whether a protein is secreted by non-classical pathways or not | [25] |
| predictNLS         | A program to predict whether a protein has nuclear localization signal or not | [28] |

Table 1 List of prediction programs used in FSD

| Title                                                  | Total Identified Proteins | Class SP | Class SP\(^3\) | Class SL | Putative Secretome |
|--------------------------------------------------------|----------------------------|----------|----------------|----------|-------------------|
| Crucial Role of Antioxidant Proteins and Hydrolytic Enzymes in Pathogenicity of *Penicillium expansum*: Analysis Based on Proteomics Approach (Secretory) | 21                         |           |               |          | 6                 |
| Crucial Role of Antioxidant Proteins and Hydrolytic Enzymes in Pathogenicity of *Penicillium expansum*: Analysis Based on Proteomics Approach (Non-secretory) | 21                         | 1        | 2              | 0        | 3                 |
| The *Phanerochaete chrysosporium* secretome: Database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium | 49                         | 25       | 5              | 0        | 30                |
| An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms (Secretory) | 46                         | 28       | 19             | 2        | 49                |
| An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms (Non-secretory) | 45                         | 0        | 5              | 1        | 6                 |
| The secretome of the maize pathogen *Ustilago maydis* (Without known functions) | 386                        | 352      | 18             | 10       | 380               |
| The secretome of the maize pathogen *Ustilago maydis* (With known functions) | 168                        | 147      | 15             | 5        | 167               |
| A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes | 25                         | 22       | 1              | 0        | 23                 |
| Fungal degradation of wood: initial proteomic analysis of extra cellular proteins of *Phanerochaete chrysosporium* grown on oak substrate | 11                         | 8        | 0              | 0        | 8                 |
| Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus *Fusarium graminearum* | 120                        | 63       | 8              | 0        | 71                |
| Expression analysis of extracellular proteins from *Phanerochaete chrysosporium* grown on different liquid and solid substrates | 27                         | 16       | 4              | 0        | 20                |
| Dandruff-associated Malassezia genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens | 34                         | 28       | 0              | 0        | 28                |
| Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes | 79                         | 79       | 0              | 0        | 79                |
| Genome, transcriptome, and secretome analysis of wood decay fungus Postia placenta supports unique mechanisms of lignocellulose conversion. | 47                         | 29       | 3              | 1        | 33                |
| Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response | 14                         | 12       | 0              | 1        | 13                |
| **Total**                                              | **1,093**                   | **815**  | **81**         | **20**   | **916**           |

Table 2 List of references and annotation results of characterized fungal secretory proteins
compensate for some of the weaknesses of individual programs, resulting in more robust predictions. Additionally, SecretomeP 1.0f [25], which can predict non-classical secretory proteins, was integrated into the FSD.

The FSD contains an identification pipeline that sequentially analyzes proteomes of interest using i) SignalP 3.0; ii) a combination of SigCleave, SigPred, and RPSP to screen those proteins not considered positive by SignalP 3.0; and iii) PSort II and TargetP 1.1b to analyze the negatives from the previous step. Additionally, SecretomeP 1.0f was integrated to provide information related to non-classical secretory proteins. To eliminate potential false positives, we filtered proteins that i) contain more than one transmembrane helix predicted by TMHMM 2.0c [26] and/or ii) the endoplasmic reticulum retention signal ([(KRHQA]-[DENIQ]-E-L; classified as false-positive; Figure 1A) [27]. In addition, iii) nuclear proteins predicted by both predictNLS [28] and PSort II [22] and iv) mitochondrial proteins predicted by PSort II [22] as well as TargetP 1.1b [23] were eliminated because two subcellular localizations are not related to secretory proteins.

Following analysis via the pipeline, the resulting putative secretory proteins after removing potential false positives are divided into four classes: i) SP contains all proteins predicted by SignalP 3.0; ii) SP$^3$ contains the proteins predicted by SigPred, SigCleave, or RPSP but not by SignalP 3.0; iii) SL contains the proteins predicted by PSort II and/or TargetP 1.1b but not by the first two steps; and iv) NS contains the proteins predicted by SecretomeP 1.0f but not by SignalP 3.0 (Figure 1A; Table 3).

Table 3: Class definitions used in FSD

| Class  | Description* |
|--------|--------------|
| Class SP | Proteins which are predicted by SignalP 3.0 |
| Class SP$^3$ | Proteins which are predicted by SigPred, SigCleave, or RPSP |
| Class SL | Proteins which are predicted by PSort II or TargetP 1.1b, but are not predicted by SignalP 3.0, SigPred, SigCleave, RPSP, or SecretomeP 1.0f |
| Class NS | Proteins which are predicted by SecretomeP 1.0f, but are not predicted by SignalP 3.0, SigPred, SigCleave, or RPSP |

* Proteins as follows were removed from all four classes described in this table: proteins which i) contain more than one trans-membrane helices, ii) have ER retention signals, iii) predicted as mitochondrial proteins by PSort II and TargetP 1.1b, and iv) predicted as nuclear proteins by TargetP 1.1b and predictNLS.
### Table 4 List and distribution of secretion-associated proteins of the fungal genomes belonging to the subphylum Pezizomycotina archived in FSD

| Species                     | Size (Mb) | # of ORFs | Class SP | Class SP^2 | Class SL | Putative Secretome | Ref          |
|-----------------------------|-----------|-----------|----------|-------------|----------|--------------------|--------------|
| **Ascomycota (Phylum)**     |           |           |          |             |          |                    |              |
| Aspergillus clavatus        | 27.9      | 9,121     | 754      | 732         | 81       | 1,567              | [53,54]      |
| Aspergillus flavus          | 36.8      | 12,604    | 1,200    | 900         | 142      | 2,332              | [55]         |
| Aspergillus fumigatus A1163 | 29.2      | 9,929     | 807      | 878         | 67       | 1,752              | [54]         |
| Aspergillus fumigatus AF293 | 29.4      | 9,887     | 781      | 909         | 84       | 1,774              | [56]         |
| Aspergillus nidulans        | 30.1      | 10,568    | 922      | 877         | 96       | 1,895              | [57]         |
| Aspergillus niger ATCC1015  | 37.2      | 12,200    | 860      | 883         | 88       | 1,831              | N            |
| Aspergillus niger CBS1338   | 34.0      | 14,063    | 1,142    | 1,230       | 154      | 2,616              | [58]         |
| Aspergillus terreus         | 29.3      | 10,406    | 934      | 916         | 81       | 1,931              | [53]         |
| Botrytis cinerea            | 42.7      | 16,448    | 1,163    | 1,287       | 182      | 2,632              | N            |
| Chaetomium globosum         | 34.9      | 11,124    | 1,121    | 923         | 99       | 2,143              | N            |
| Coccidioides immitis H538.4 | 27.7      | 10,663    | 957      | 957         | 80       | 1,585              | N            |
| Coccidioides immitis RMSCC 2394 | 28.8   | 10,408    | 752      | 920         | 66       | 1,561              | N            |
| Coccidioides immitis RMSCC 3703 | 27.6   | 10,465    | 539      | 892         | 65       | 1,496              | N            |
| Coccidioides immitis RS     | 28.9      | 10,457    | 476      | 855         | 102      | 1,433              | [60]         |
| Coccidioides posadasii RMSCC 3488 | 28.1   | 9,964     | 546      | 838         | 95       | 1,479              | N            |
| Coccidioides posadasii Silveira | 27.5   | 10,125    | 558      | 869         | 91       | 1,518              | N            |
| Cochliobolus heterostrophus C5 | 34.9   | 9,433     | 932      | 725         | 83       | 1,740              | N            |
| Cryptococcus albidus         | 43.9      | 11,184    | 1,040    | 1,064       | 145      | 2,269              | [59]         |
| Fusarium graminearum GZ3639c | 15.1     | 6,604     | 373      | 386         | 47       | 806                | [61]         |
| Fusarium graminearum MIPS   | 36.1      | 13,920    | 1,370    | 1,072       | 118      | 2,560              | N            |
| Fusarium graminearum PH-1   | 36.6      | 13,339    | 1,282    | 1,004       | 118      | 2,404              | [61]         |
| Fusarium oxysporum          | 61.4      | 17,608    | 1,613    | 1,297       | 147      | 3,057              | N            |
| Fusarium solani             | 51.3      | 15,707    | 1,381    | 1,242       | 155      | 2,778              | [62]         |
| Fusarium verticillioides    | 41.9      | 14,199    | 1,347    | 1,071       | 116      | 2,534              | N            |
| Histoplasma capsulatum G186AR | 29.9   | 7,454     | 357      | 578         | 96       | 1,031              | N            |
| Histoplasma capsulatum G217B | 41.3     | 8,038     | 393      | 583         | 103      | 1,079              | N            |
| Histoplasma capsulatum H143 | 39.0      | 9,547     | 468      | 842         | 87       | 1,397              | N            |
| Histoplasma capsulatum H88  | 37.9      | 9,445     | 492      | 832         | 99       | 1,423              | N            |
| Histoplasma capsulatum Nam1 | 33.0      | 9,349     | 398      | 736         | 79       | 1,213              | [60]         |
| Magnaporthia oryzae         | 41.7      | 11,069    | 1,573    | 833         | 64       | 2,470              | [63]         |
| Microsporum canis           | 23.3      | 8,777     | 564      | 702         | 88       | 1,354              | N            |
| Microsporum gypseum         | 23.3      | 8,876     | 629      | 669         | 52       | 1,350              | N            |
| Mycosphaerella fijiensis    | 73.4      | 10,327    | 770      | 778         | 81       | 1,629              | N            |
| Mycosphaerella gramineicola | 41.9      | 11,395    | 979      | 913         | 81       | 1,973              | N            |
| Neosartorya fischeri        | 32.6      | 10,603    | 959      | 818         | 84       | 1,861              | [54]         |
| Neurospora crassa           | 39.2      | 9,842     | 817      | 788         | 61       | 1,666              | [64]         |
| Neurospora crassa MIPS      | 34.2      | 9,572     | 788      | 749         | 78       | 1,615              | N            |
| Neurospora discrepate discrete | 37.3  | 9,948     | 823      | 800         | 88       | 1,711              | N            |
| Neurospora tetrospora        | 37.8      | 10,640    | 849      | 895         | 73       | 1,817              | N            |
| Paracoccidioides brasiliensis Pb01 | 33.0  | 9,136     | 402      | 808         | 71       | 1,281              | N            |
| Paracoccidioides brasiliensis Pb03 | 29.1  | 9,264     | 470      | 823         | 92       | 1,385              | N            |
| Paracoccidioides brasiliensis Pb18 | 30.0  | 8,741     | 425      | 743         | 55       | 1,233              | N            |
| Penicillium chrysogenum      | 32.2      | 12,791    | 947      | 1,008       | 127      | 2,082              | [65]         |
| Penicillium mamei            | 28.6      | 10,638    | 713      | 792         | 109      | 1,614              | N            |
| Podospora anserina           | 35.7      | 10,596    | 1,127    | 893         | 124      | 2,144              | [66]         |
| Pyrenophora tritici-repentis | 38.0      | 12,169    | 1,228    | 912         | 123      | 2,263              | N            |

*Choi et al. BMC Genomics 2010, 11:105*  
http://www.biomedcentral.com/1471-2164/11/105
System structure of the FSD

To improve the expandability and flexibility of the FSD, we adopted a three-layer structure (i.e., data warehouse, analysis pipeline, and user interface) in its design. The data warehouse was established using the standardized genome warehouse managed by the CFGP (http://cfgp.snu.ac.kr)[16] that has been used in various bioinformatics systems [15,29-35]. The pipeline layer was built with a series of Perl programs.

In addition to the prediction programs described above, ChloroP 1.1 as well as hydropathy plots [36] were included in the FSD to provide additional information on secretory proteins. Whenever new fungal genomes become available, the automated pipeline classifies them based on the predictions of nine programs, thus keeping the FSD current (Figure 1B).

MySQL 5.0.67 and PHP 5.2.9 were used to maintain database and to develop web-based user interfaces that present complex information intuitively. Web pages were serviced through Apache 2.2.11. Favorite, a personal data repository used in the CFGP (http://cfgp.snu.ac.kr)[16], was integrated to provide thirteen functions for further analyses.

Utility and Discussion

**Discussion**

**Secretory proteins in 158 fungal/oomycete genomes**

To survey the genome-wide distribution of secretory proteins in fungi and oomycetes, we used the pipeline to analyze all predicted proteins encoded by 158 fungal/oomycete genomes. Of the 1,373,444 open reading frames (ORFs) analyzed, 92,926 (6.77%), 103,224 (7.52%), and 12,733 (0.93%) proteins belonged to classes SP, SP3, and SL, respectively (Table 4, 5, and 6). In total, 208,883 ORFs (15.21%) were denoted putative secretory proteins. The proteins belonging to class NS were not included in the putative secretome because they represented more than 40% of whole proteome.

To determine the phylum-level distribution of classes SP, SP3, and SL within fungi, we investigated the proportions of the three classes among subphyla (Figure 2). Class SP3 was the largest, class SP was a little smaller, and the class SL was much smaller; this was consistent over every subphylum. Only in *Plasmodium* species, oomycetes, and the kingdom Metazoa class SP was dominant. Class SL did not exceed 2.10% of the whole genome, except in *Plasmodium* species (4.52%). *Plasmodium* species also showed the lowest variance among the three classes, which may reflect signal peptide-independent types of secretory proteins such as vacuolar transport signals (VTSs) [12]. These results may be partially affected by the composition of the training data for each prediction program and inherent features of each algorithm.

The phylum Basidiomycota had a larger proportion of secretory proteins (17.90%) than other fungal taxonomy such as the subphylum Mucoromycotina (11.99%) and the phyla Ascomycota (12.87%) and Microsporidia (15.10%). Within the phylum Ascomycota, the subphylum Pezizomycotina showed a higher portion of class SP (7.82%) than the subphyla Saccharomycotina and Taphrinomycotina (4.57% and 3.74%, respectively). When considered that subphylum Pezizomycotina contains many pathogenic fungi (47 of 59) compared with subphylum Saccharomycotina (11 of 65), the abundance of secretory proteins in the subphylum Pezizomycotina suggests that pathogens may have larger secretome than saprophytes in general. In fact, *Magnaporthe oryzae* and *Neurospora crassa*, a closely related pair of pathogen and non-pathogen supported by

|    | Sclerotinia sclerotiorum | Sporotrichum thermophile | Stagonospora nodorum | Talaromyces stipitatus | Thielavia terrestris | Trichoderma atroviride | Trichoderma reesei | Trichoderma virens Gv29-8 | Trichophyton equinum | Uncinocarpus reesii | Verticillium albo-atrum VaMs. 102 | Verticillium dahliae Vdls. 17 | Total |
|----|-------------------------|--------------------------|---------------------|------------------------|---------------------|-----------------------|-------------------|------------------------|-------------------|-------------------|--------------------------|-------------------|--------|
|    | 38.3 | 14,522 | 971 | 1,109 | 147 | 2,227 | N |
|    | 38.7 | 8,806 | 697 | 658 | 66 | 1,421 | N |
|    | 37.2 | 15,983 | 1,511 | 1,309 | 142 | 2,962 | [67] |
|    | 35.7 | 13,252 | 748 | 1,116 | 114 | 1,978 | N |
|    | 37.0 | 9,815 | 877 | 855 | 67 | 1,799 | N |
|    | 36.1 | 11,100 | 907 | 935 | 86 | 1,928 | N |
|    | 33.5 | 9,129 | 738 | 766 | 70 | 1,574 | [68] |
|    | 38.8 | 11,643 | 933 | 1,009 | 93 | 2,035 | N |
|    | 24.2 | 8,576 | 571 | 699 | 69 | 1,339 | N |
|    | 22.3 | 7,798 | 485 | 626 | 64 | 1,175 | [60] |
|    | 32.9 | 10,239 | 1,074 | 815 | 73 | 1,962 | N |
|    | 33.9 | 10,575 | 1,157 | 861 | 77 | 2,095 | N |
| Total | 2,059.4 | 641,257 | 50,164 | 52,111 | 5,578 | 107,853 | - |

* Taxonomy based on [69]
  b Insufficient exon/intron information
  c Incomplete coverage of genome information

---

**Table 4: List and distribution of secretion-associated proteins of the fungal genomes belonging to the subphylum Pezizomycotina archived in FSD (Continued)**
| Species | Fungi (Kingdom) | Saccharomycotina (Phylum) | Size (Mb) | # of ORFs | Class SP | Class SP longevity | Class SL | Putative Secretome | Ref |
|---------|----------------|----------------------------|-----------|-----------|-----------|-------------------|-----------|-------------------|-----|
| Candida albicans SC5314 | Ascomycota (Phylum) | Saccharomycotina (Subphylum) | 14.3 | 6,185 | 321 | 405 | 87 | 813 | [70,71] |
| Candida albicans WO-1 | | | 14.5 | 6,160 | 310 | 385 | 78 | 773 | [72] |
| Candida dubliniensis | | | 12.3 | 5,165 | 308 | 321 | 83 | 712 | [72] |
| Candida guilliermondii | | | 10.6 | 5,920 | 279 | 400 | 63 | 742 | [72] |
| Candida lusitaniae | | | 10.6 | 5,941 | 310 | 482 | 50 | 842 | [72] |
| Candida parapsilosis | | | 13.1 | 5,733 | 308 | 321 | 83 | 712 | [72] |
| Candida tropicalis | | | 14.6 | 6,258 | 360 | 373 | 76 | 809 | [72,74] |
| Debaryomyces hansenii | | | 12.2 | 6,354 | 254 | 357 | 74 | 685 | [73] |
| Eremothecium gossypii | | | 8.8 | 4,717 | 204 | 333 | 35 | 572 | [75] |
| Kluyveromyces lactis | | | 10.7 | 5,327 | 248 | 304 | 60 | 612 | [73] |
| Kluyveromyces polysporus | | | 14.7 | 5,367 | 219 | 276 | 58 | 553 | [76] |
| Lodderomyces elongisporus | | | 15.5 | 5,802 | 253 | 351 | 50 | 654 | [72] |
| Pichia stipitis | | | 15.4 | 5,839 | 263 | 374 | 58 | 695 | [78] |
| Saccharomyces bayanus 623-6C YM4911 | | | 11.9 | 4,966 | 200 | 275 | 44 | 519 | [79] |
| Saccharomyces bayanus MCYC 623 | | | 11.5 | 9,385 | 663 | 767 | 141 | 1571 | [80] |
| Saccharomyces castellii | | | 11.4 | 4,677 | 177 | 240 | 46 | 463 | [79] |
| Saccharomyces cerevisiae 273614N | | | 12.5 | 5,354 | 223 | 261 | 51 | 535 | [81] |
| Saccharomyces cerevisiae 322134S | | | 12.6 | 5,382 | 224 | 290 | 53 | 567 | [81] |
| Saccharomyces cerevisiae 378604X | | | 12.6 | 5,400 | 232 | 267 | 53 | 552 | [81] |
| Saccharomyces cerevisiae AWRI1631 | | | 11.2 | 5,451 | 220 | 364 | 63 | 647 | N |
| Saccharomyces cerevisiae BC187 | | | 12.5 | 5,332 | 226 | 263 | 47 | 536 | [81] |
| Saccharomyces cerevisiae DBVPG1106 | | | 12.5 | 5,318 | 225 | 253 | 52 | 530 | [81] |
| Saccharomyces cerevisiae DBVPG1373 | | | 12.4 | 5,349 | 229 | 260 | 48 | 537 | [81] |
| Saccharomyces cerevisiae DBVPG1788 | | | 12.4 | 5,347 | 227 | 263 | 46 | 536 | [81] |
| Saccharomyces cerevisiae DBVPG1853 | | | 12.5 | 5,359 | 224 | 265 | 51 | 540 | [81] |
| Saccharomyces cerevisiae DBVPG6040 | | | 12.6 | 5,364 | 221 | 271 | 50 | 542 | [81] |
| Saccharomyces cerevisiae DBVPG6044 | | | 12.5 | 5,890 | 224 | 268 | 48 | 540 | [81] |
| Saccharomyces cerevisiae DBVPG6765 | | | 12.2 | 5,377 | 230 | 263 | 48 | 541 | [81] |
| Saccharomyces cerevisiae K11 | | | 12.5 | 5,375 | 228 | 270 | 52 | 550 | [81] |
| Saccharomyces cerevisiae L_1374 | | | 12.4 | 5,346 | 225 | 264 | 55 | 544 | [81] |
| Saccharomyces cerevisiae L_1528 | | | 12.4 | 5,346 | 227 | 258 | 48 | 533 | [81] |
| Saccharomyces cerevisiae M22 | | | 10.8 | 6,755 | 249 | 399 | 62 | 710 | [82] |
| Saccharomyces cerevisiae NCYC110 | | | 12.5 | 5,408 | 226 | 264 | 57 | 547 | [81] |
| Saccharomyces cerevisiae NCYC361 | | | 12.6 | 5,360 | 228 | 261 | 49 | 538 | [81] |
| Saccharomyces cerevisiae RM11-1a | | | 11.7 | 5,696 | 264 | 283 | 63 | 610 | N |
| Saccharomyces cerevisiae S288C | | | 12.2 | 6,692 | 394 | 425 | 99 | 918 | [83] |
| Saccharomyces cerevisiae SK1 | | | 12.4 | 5,433 | 233 | 269 | 55 | 557 | [81] |
| Saccharomyces cerevisiae UWOP03_461_4 | | | 12.6 | 5,329 | 218 | 268 | 51 | 537 | [81] |
| Saccharomyces cerevisiae UWOP05_217_3 | | | 12.6 | 5,350 | 217 | 264 | 47 | 528 | [81] |
| Saccharomyces cerevisiae UWOP05_227_2 | | | 12.6 | 5,334 | 220 | 266 | 51 | 537 | [81] |
| Saccharomyces cerevisiae UWOP03_787_3 | | | 12.6 | 5,392 | 225 | 269 | 51 | 545 | [81] |
| Saccharomyces cerevisiae UWOP07_2421 | | | 12.6 | 5,368 | 226 | 266 | 56 | 548 | [81] |
| Saccharomyces cerevisiae W303 | | | 12.4 | 5,467 | 237 | 271 | 52 | 560 | [81] |
| Saccharomyces cerevisiae Y12 | | | 12.6 | 5,370 | 223 | 268 | 57 | 548 | [81] |
| Saccharomyces cerevisiae Y55 | | | 12.3 | 5,415 | 239 | 262 | 60 | 561 | [81] |
recent phylogenomic studies [37-39], contain 22.31% and 16.93% of secretory proteins, respectively. Moreover, the same tendency was found in comparison with 158 fungal/oomycete genomes archived in the FSD (pathogens and saprophytes showed 14.06% and 11.70%, respectively).

**Effectors encoded by fungal/oomycete and Plasmodium genomes**

*Phytophthora* species, a group that includes many important plant pathogens, uses a RXLR signal to secrete effectors to host cells [40]. RXLR effectors were tightly co-located with signal peptides predicted by the SignalP 3.0 with high confidence values (HMM and NN for 0.93 and 0.65, respectively) [41]. With the same conditions, we identified 734 putative RXLR effectors from three *Phytophthora* species, similar to a previous study [42]. However, 153 fungal genomes showed that only 0.04% of the total proteome contained this motif, suggesting that the use of RXLR for secretion is oomycete-specific.

The motivation of finding the RXLR pattern in oomycetes was the RXLX [EDQ] motif of the VTS in the malaria pathogen, *Plasmodium falciparum*. Once *P. falciparum* invades the human erythrocyte, it secretes the proteins that carry the pentameric VTS of the RXLX [EDQ] motif from the parasitophorous vacuole to the host cytoplasm [12,13]. To determine how many VTSs could be detected by our pipeline, we investigated 217 proteins of *P. falciparum* [13]. Of these, 115 proteins (53.00%) were classified as secretory proteins, defined in the FSD by the RXLX [EDQ] motif. Comparing our result to that predicted by SignalP 3.0 alone (41 out of 217), we found that our pipeline demonstrated high fidelity in detecting proteins containing VTSs.

In class SP, the proportions of proteins possessing the RXLX [EDQ] motif but not the RXLR motif were 96.75%, 56.18%, and 93.21% in fungi, oomycetes, and *Plasmodium* species, respectively (Figure 3A). There were similar proportions of the RXLX [EDQ] motif in classes SP3 and SL across the three groups (Figure 3B and 3C). Taken together, these data show that the RXLR motif, with signal peptides predicted by SignalP 3.0, is oomycete-specific [41]. It is interesting that fungal genomes have significantly higher numbers of the RXLX [EDQ] motif than *Plasmodium* species (t-test based on amino acid frequency in each genome; $P = 2.2e^{-16}$), suggesting

### Table 5: List and distribution of secretion-associated proteins of the fungal genomes belonging to the subphylum Saccharomycotina and Taphrinomycotina archived in FSD (Continued)

| Genome Name                  | Accession | RPKM | Proportions (%) | Total Proteins | Proteins with VTS |
|------------------------------|-----------|------|-----------------|---------------|------------------|
| Saccharomyces cerevisiae Y9   | 12.6      | 5,377 | 12.6            | 223           | 271              |
| Saccharomyces cerevisiae Yle17_E5 | 12.5      | 5,376 | 12.5            | 227           | 265              |
| Saccharomyces cerevisiae YLM789 | 12.0     | 5,903 | 12.0            | 293           | 303              |
| Saccharomyces cerevisiae YLM975 | 12.4     | 5,341 | 12.4            | 223           | 255              |
| Saccharomyces cerevisiae YLM978 | 12.4     | 5,353 | 12.4            | 224           | 258              |
| Saccharomyces cerevisiae YLM981 | 12.5     | 5,351 | 12.5            | 224           | 256              |
| Saccharomyces cerevisiae YPS128 | 12.4     | 5,364 | 12.4            | 230           | 269              |
| Saccharomyces cerevisiae YPS163 | 10.7     | 6,648 | 10.7            | 229           | 368              |
| Saccharomyces cerevisiae YPS606 | 12.5     | 5,354 | 12.5            | 224           | 270              |
| Saccharomyces cerevisiae YS2   | 12.6      | 5,383 | 12.6            | 221           | 254              |
| Saccharomyces cerevisiae YS4   | 12.5      | 5,398 | 12.5            | 215           | 267              |
| Saccharomyces cerevisiae YS9   | 12.6      | 5,373 | 12.6            | 226           | 265              |
| Saccharomyces kloveni          | 11.0      | 2,968 | 11.0            | 120           | 180              |
| Saccharomyces kudriavzevii     | 11.2      | 3,768 | 11.2            | 187           | 195              |
| Saccharomyces mikatae          | 11.5      | 9,016 | 11.5            | 575           | 630              |
| Saccharomyces mikatae WashU    | 10.8      | 3,100 | 10.8            | 161           | 154              |
| Saccharomyces paradoxus        | 11.9      | 8,999 | 11.9            | 581           | 615              |
| Yarrowia lipolytica            | 20.5      | 6,524 | 20.5            | 409           | 464              |
| Taphrinomycotina (Subphylum)   |           |      |                 |              |                  |
| Pneumocystis carinii a b c     | 6.3       | 4,020 | 6.3             | 129           | 333              |
| Schizosaccharomyces japonicus  | 11.3      | 5,172 | 11.3            | 207           | 312              |
| Schizosaccharomyces octosporus | 11.2      | 4,925 | 11.2            | 190           | 263              |
| Schizosaccharomyces pombe      | 12.6      | 5,058 | 12.6            | 192           | 288              |
| Total                        | 853.1     | 383,828 | 853.1         | 17,389        | 21,403           |

a Taxonomy based on [69]
b Insufficient exon/intron information
c Incomplete coverage of genome information
Table 6: List and distribution of secretion-associated proteins of the fungal genomes belonging to the phyla Basidiomycota, Chytridiomycota, and Microsporidia, the subphylum Mucoromycotina, and the phylum Peronosporomycota (oomycetes) archived in FSD

| Species | Size (Mb) | # of ORFs | Class SP | Class SP<sup>1</sup> | Class SL | Putative Secretome | Ref |
|---------|-----------|-----------|----------|----------------------|----------|--------------------|-----|
| **Fungi (Kingdom)** | | | | | | | |
| **Basidiomycota (Phylum)** | | | | | | | |
| Coprinus cinereus | 36.3 | 13,410 | 1,189 | 1,032 | 119 | 2,340 | N |
| Cryptococcus neoformans Serotype A | 18.9 | 6,980 | 377 | 549 | 56 | 982 | N |
| Cryptococcus neoformans Serotype B | 19.0 | 6,870 | 331 | 529 | 44 | 904 | N |
| Cryptococcus neoformans Serotype D B-3501A | 18.5 | 6,431 | 342 | 523 | 39 | 904 | [86] |
| Cryptococcus neoformans Serotype D JEC21 | 19.1 | 6,475 | 344 | 541 | 38 | 923 | [86] |
| Laccaria bicolor | 64.9 | 20,614 | 1,190 | 2,024 | 256 | 3,470 | N |
| **Pucciniomycotina (Subphylum)** | | | | | | | |
| Melampsora laricis-populina | 21.9 | 16,694 | 1,035 | 1,483 | 233 | 3,021 | N |
| Puccinia graminis | 88.7 | 20,569 | 1,931 | 2,020 | 230 | 4,181 | N |
| Sporobolomyces roseus | 21.2 | 5,536 | 187 | 592 | 43 | 822 | N |
| **Ustilaginomycotina (Subphylum)** | | | | | | | |
| Malassezia globosa | 9.0 | 4,286 | 211 | 378 | 37 | 626 | [50] |
| Ustilago maydis 521 | 19.7 | 6,689 | 789 | 583 | 10 | 1,382 | [89] |
| Ustilago maydis FB1 | 19.3 | 6,950 | 481 | 717 | 34 | 1,323 | [89] |
| Ustilago maydis MIPS | 19.7 | 6,787 | 574 | 687 | 34 | 1,295 | N |
| **Chytridiomycota (Phylum)** | | | | | | | |
| Batrachochytrium dendrobatidis JAM81 | 24.3 | 8,732 | 806 | 750 | 108 | 1,664 | N |
| Batrachochytrium dendrobatidis JEL423 | 23.9 | 8,818 | 650 | 785 | 91 | 1,526 | N |
| **Mucoromycotina (Subphylum incertae sedis)** | | | | | | | |
| *Mucor circinelloides* | 36.6 | 10,930 | 580 | 623 | 83 | 1,286 | N |
| Phycymyces blakesleeanus | 55.9 | 14,792 | 642 | 1,085 | 221 | 1,948 | N |
| Rhizopus oryzae | 46.1 | 17,482 | 750 | 994 | 202 | 1,946 | [90] |
| **Microsporidia (Phylum)** | | | | | | | |
| Antonospora locustae<sup>b</sup> | 6.1 | 2,606 | 166 | 208 | 62 | 436 | N |
| Encephalitozoon cuniculi | 2.5 | 1,996 | 90 | 135 | 34 | 259 | [91] |
| **Alveolata (Kingdom)** | | | | | | | |
| Apicomplexa (Phylum) | | | | | | | |
| Plasmodium berghei | 18.0 | 12,175 | 844 | 554 | 569 | 1,967 | N |
| Plasmodium chabaudi | 16.9 | 15,007 | 1,027 | 643 | 661 | 2,331 | N |
| Plasmodium falciparum 3D7 | 21.0 | 5,387 | 212 | 283 | 267 | 762 | [92] |
| Plasmodium knowlesi | 23.5 | 5,103 | 305 | 280 | 81 | 666 | N |
| **Stramenopila (Kingdom)** | | | | | | | |
| Peronosporomycota (Phylum) | | | | | | | |
| Hyaloperonospora parasitica | 83.6 | 14,789 | 868 | 1,235 | 132 | 2,235 | N |
| Phytophthora capsici | 107.8 | 17,414 | 1,485 | 1,759 | 136 | 2,800 | N |
| Phytophthora infestans<sup>c</sup> | 228.5 | 22,658 | 1,668 | 1,923 | 153 | 3,744 | [93] |
| Phytophthora ramorum | 66.7 | 15,743 | 1,670 | 1,372 | 91 | 3,133 | [94] |
| Phytophthora sojae | 86.0 | 19,027 | 2,040 | 1,662 | 96 | 3,798 | [94] |
| **Total** | 1,449.1 | 386,513 | 27,761 | 31,470 | 4,796 | 64,027 | - |

<sup>a</sup> Taxonomy based on [69]  
<sup>b</sup> Insufficient exon/intron information  
<sup>c</sup> Incomplete coverage of genome information
that the RXLX [EDQ] motif may be one of fungal-specific signatures of effectors.

Utility

**FSD web interfaces**

To support the browsing of the global patterns of archived data, the FSD prepares diverse charts and tables. For example, intersections of prediction results are summarized in a chart for each genome (Figure 4). Despite of the many programs, all prediction results for each protein are displayed on one page, allowing users to browse them easily (Figure 5).

The SNUGB interface (http://genomebrowser.snu.ac.kr/[15] provides several fields: i) signal peptides predicted by four different programs; ii) effector patterns, such as RXLR and RXLX [EDQ]; iii) nucleotide localization signals predicted by predictNLS; iv) transmembrane helixes predicted by TMHMM 2.0c; and v) hydropathy plots (Figure 6). The users can readily compare secretome-related information with diverse genomic contexts. The personalized virtual space, Favorite, supports in-depth analyses in the FSD

The FSD allows users to collect proteins of interest and save them into the Favorite, which provides thirteen...
functions: i) classes distribution of proteins; ii) comparisons of predicted signal peptides generated by the four programs; iii) distributions and lists of proteins with predicted signal peptide cleavage sites; iv) compositions of amino acids near the cleavage sites; v) analyses of subcellular localization predictions; vi) lists and ratios of proteins that have chloroplast transit peptides, as determined by ChloroP 1.1; vii) analyses of proteins detected by SecretomeP 1.0f; viii) lists and distribution charts of proteins with trans-membrane helices, as predicted by TMHMM 2.0c; ix) hydropathy plots for proteins; x) analyses of proteins believed to be targeted to the nucleus of a host cell supported by predictNLS; xi) distributions and lists of proteins with a specific amino acid patterns; xii) lists of functional domains predicted by InterProScan; xiii) domain architecture of InterPro Scan (Figure 5 One page summary for a protein).
7). From these result pages, users can collect and store proteins in Favorite again, for further analyses. Additionally, Favorites created in the FSD can be shared with the CFGP (http://cfgp.snu.ac.kr/)[16], permitting users to use the 22 bioinformatics tools provided in the CFGP web site.

Conclusions

Given the availability of large number of fungal genomes and diverse prediction programs for secretory proteins, a three-layer classification rule was established and implemented in a web-based database, the FSD. With the aid of an automated pipeline, the FSD classifies putative secretory proteins from 158 fungal/oomycetes genomes into four different classes, three of which are defined as the putative secretome. The proportion of fungal secretory proteins and host targeting signals varies considerably by species. It is interesting that fungal genomes have high proportions of the RXLX [EDQ] motif, characterized as host targeting signal in Plasmodium species. Summaries of the complex prediction results from twelve programs help users to readily access to the information provided by the FSD. Favorite, a personalized virtual space in the CFGP, serves thirteen different analysis tools for further in-depth analyses. Moreover, 22 bioinformatics tools
provided by the CFGP can be utilized via the Favorite. Given these features, the FSD can serve as an integrated environment for studying secretory proteins in the fungal kingdom.

### Availability and requirements

All data and functions described in this paper can be freely accessed through the FSD web site at http://fsd.snu.ac.kr/.

### Acknowledgements

This work was supported by the National Research Foundation of Korea grants (2009-0063340 and 2009-0080161) and grants from the Biogreen21 (2009001-034-044-009-01-00), the TDPAF (30015-04-SB020), and the Crop Functional Genomics Center (2009K001198). JC is grateful for the graduate fellowship through the Brain Korea 21 Program.

### References

1. Lippincott-Schwartz J, Roberts TH, Hirschberg K: Secretory protein trafficking and organelle dynamics in living cells. Annu Rev Cell Dev Biol 2000, 16:557-589.
2. Abbas KA, Lichtman HA, Pillai S: Cellular and Molecular Immunology Saunders, 6 2006.
3. Cho WJ, Jeremic A, Rognlien KT, Zhvania MG, Lazrishvili I, Tamar B, Jena BP: Structure, isolation, composition and reconstitution of the neuronal fusion pore. Cell Biol Int 2004, 28(10):1699-708.
4. Cho SJ, Jeffinia K, Glavaski A, Jeffinia S, Jena BP, Anderson LL: Structure and dynamics of the fusion pores in live GH-secreting cells revealed using atomic force microscopy. Endotechnology 2002, 143(3):1144-1148.
5. Suarez MB, Sanz L, Chamorro MI, Rey M, Gonzalez FJ, Llobell A, Monte E: Proteomic analysis of secreted proteins from Trichoderma harzianum: Identification of a fungal cell wall-induced aspartic protease. Fungal Genet Biol 2005, 42(11):924-934.
6. van den Wymelenberg A, Minges P, Sabat G, Martinez D, Aerts A, Salamov A, Grigoriev I, Shapiro H, Putnam N, Belinky P, Dosoretz C, Gaskell J, Kersten P, Cullen D: Computational analysis of the Phanerochaete chrysosporium v2.0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveal complex mixtures of secreted proteins. Fungal Genet Biol 2006, 43(3):343-356.
7. Vinzant TB, Adney WS, Decker NR, Baker JO, Kinert MT, Sherman NE, Fox JW, Himmel ME: Fingerprinting Trichoderma reesei hydrolases in a commercial cellulase preparation. Appl Biochem Biotechnol 2001, 91-93:99-107.
8. Lesage G, Bussey H: Cell wall assembly in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 2006, 70(2):317-343.
9. Waters CM, Bassler BL: Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 2005, 21:319-346.
10. Cornelis GR, Van Gijsegem F: Assembly and function of type III secretory systems. Annu Rev Microbiol 2000, 54:735-774.
11. Kamoun S: A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol 2006, 44:41-60.
12. Hiller NL, Bhattacharjee S, van Ooj C, Liokos K, Harrison T, Lopez-Estrano C, Halder N: A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 2004, 306(5703):1994-1997.
13. Manti M, Good RT, Bug M, Kruepler E, Cowman AF: Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 2004, 306(5703):1930-1933.
14. Bouvhs W, Wattenberg A, Zorn H: Fungal secretomes—nature’s toolbox for white biotechnology. Appl Microbial Biotechnol 2008, 80(3):381-388.

15. Jung K, Park J, Choi J, Park B, Kim S, Ahn K, Choi J, Choi D, Kang S, Lee Y-H: SNUGa: a versatile genome browser supporting comparative and functional genomic analyses. BMC Genomics 2008, 9:588.

16. Park J, Park B, Jung K, Jang S, Yu K, Choi J, Kong S, Kim S, Kim H, Kim JF, Blair JE, Lee K, Kang S, Lee YH: CFGa: a web-based, comparative fungal genomics platform. Nucleic Acids Res 2008, 36:D562-571.

17. Bendtsen JD, Nielsen H, von Heijne G, Krogh A: Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004, 340(4):783-795.

18. Choi KH, Tan TW, Ranganathan S: SPoD—a signal peptide database. BMC Bioinformatics 2005, 6:249.

19. Rice P, Longden I, Bleasby A: EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000, 16(6):276-277.

20. Bradford JR: Protein Design for Biopharmaceutical Development at GlaxoSmithKline: in silico Methods for Prediction of Signal Peptides and their Cleavage Sites, and Linear Epitopes. Ph.D thesis, The University of Leeds, Department of Biological Sciences 2001.

21. Ploegczynski D, Slabinski L, Ginalski K, Rychlewski L: Prediction of signal peptides in protein sequences by neural networks. Acta Biochim Pol 2008, 55(2):261-267.

22. Hutton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Bojiller CJ, Nakai K: WoLF PSORT: Protein localization predictor. Nucleic Acids Res 2003, 31(Web Server issue):W585-S87.

23. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 2000, 300(4):1005-1016.

24. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umbrasyk L, Drawd A, Jansen R, Liu Y, Cheung HK, Miller P, Genest M, Roeder GS, Snyder M: Subcellular localization of the yeast proteome. Genes Dev 2002, 16(6):707-719.

25. Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S: Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 2004, 17(4):349-356.

26. Sonnhammer EL, von Heijne G, Krogh A: A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 1998, 6:175-182.

27. Raykohl I, Alanen H, Salo K, Juvansuu A, Nguyen DV, Lahta-Ranta M, Ruddock L: A molecular specificity code for the three mammalian KDEL receptor subunits. FEBS Lett 2000, 474(1-2):411-415.

28. Park J, Park J, Jang S, Kim S, Kong S, Choi J, Ahn K, Kim J, Lee K, Sim S, Park B, Jung K, Kim S, Kang S, Lee YH: FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. BMC Genomics 2008, 9(Suppl 1):W585-587.

29. Colom M, Nair R, Rost B: A hidden Markov model for predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 2000, 300(4):1005-1016.

30. Cokol M, Nair R, Rost B: A simple method for displaying the hydropathic character of a protein. Mol Cell Proteomics 2005, 4:1005-1016.

31. Jiang RH, Tripathy S, Gover et al: TyperB: A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol 2006, 6:199.

32. Wang H, Xu Z, Gao L, Hao B: A fungal phylogeny based on 82 complete genomes using the composition vector method. BMC Evol Biol 2009, 9:195.

33. Soanes DM, Alam I, Connell M, Wong YM, Hedeler C, Paton NW, Rattay M, Hubbard SJ, Oliver SG, Talbot NJ: Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One 2008, 3(6):e2300.

34. Whisson SC, Boeving PC, Moleleki L, Avrava AO, Morales JG, Gilroy EM, Armstrong MR, Grootlaad SW, van West P, Chapman S, Heim I, Toth IK, Pithchard L, Birch PR: A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature 2007, 450(7166):115-118.

35. Wymelenberg AV, Sabat G, Martinez D, Rajangam AS, Teeri TT, Gaskell J, Kersten PJ, Cullen D: The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. J Biotechnol 2005, 118(1):17-34.

36. Yi S, Womack S, Kamoun S, Lee AF, Kaiser J, Womack B: An analysis of the Candida albicans genome database for soluble secreted proteins using computer-based prediction algorithms. Yeast 2003, 20(7):595-610.

37. Mueller O, Kahmann R, Aguilar G, Trejo-Aguilar B, Wu A, de Vries RP: The secretome of the maize pathogen Ustilago maydis. Fungal Genet Biol 2008, 45(Suppl 1):S63-70.

38. Maitz A, Mau T, Conrath R, Schmitz C, Durbin R: Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus Fusarium graminearum. Proteomics 2007, 7(17):3171-3183.

39. Satto S, Liu F, Koc H, Tien M: Expression analysis of extracellular proteins from Phanerochaete chrysosporium grown on oak substrate.Curr Genomics 2007, 8(4):399-404.

40. Wymelenberg AV, Sabat G, Martinez D, Rajangam AS, Teeri TT, Gaskell J, Kersten PJ, Cullen D: The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. J Biotechnol 2005, 118(1):17-34.

41. Lee SA, Warnsmy S, Kamoun S, Lee AF, Kaiser J, Womack B: An analysis of the Candida albicans genome database for soluble secreted proteins using computer-based prediction algorithms. Yeast 2003, 20(7):595-610.

42. Abbas A, Kac H, Liu F, Tien M: Fungal degradation of wood: initial proteomic analysis of extracellular proteins of Phanerochaete chrysosporium grown on oak substrate.Curr Genomics 2007, 8(4):399-404.

43. Wymelenberg AV, Sabat G, Martinez D, Rajangam AS, Teeri TT, Gaskell J, Kersten PJ, Cullen D: The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. J Biotechnol 2005, 118(1):17-34.

44. Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, Ferreira P, Ruiz-Duenas FJ, Martinez AT, Kersten P, Hammel KE, Kubicek CP, Ferreira P, Ruiz-Duenas FJ, Martinez AT, Kersten P, Hammel KE: Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus Fusarium graminearum. Proteomics 2007, 7(17):3171-3183.

45. Wang H, Xu Z, Gao L, Hao B: A fungal phylogeny based on 82 complete genomes using the composition vector method. BMC Evol Biol 2009, 9:195.

46. Soanes DM, Alam I, Connell M, Wong YM, Hedeler C, Paton NW, Rattay M, Hubbard SJ, Oliver SG, Talbot NJ: Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One 2008, 3(6):e2300.
Page 14 of 15

http://www.biomedcentral.com/1471-2164/11/105

54. Fedorova ND, Khalidi N, Joardar VS, Maiti R, Amended P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albaranga A, Angiulli S, Bussey H, Boywer P, Coty PJ, Dyer PS, Egan A, Galens K, Fraser-Liggett CM, Haas BJ, Inman JM, Kent N, Lemese SA, Malavais I, Órlios J, Roemer T, Rommig CM, Sundaram JP, Sutton G, Turner G, Vieter JC, et al. Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. PLoS Genet 2008, 4(6):e100046.

55. Payne GA, Nieman WC, Wortman JR, Pritchard BL, Brown D, Dean RA, Bhatnagar D, Cleveland TE, Machida M, Yu J. Whole genome comparison of A. flavus and A. oryzae. Med Mycol 2006, 44(6):9-11.

56. Nieman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Benimam A, Abe K, Archer DB, Bermejo C, Bennett J, Boywer P, Chen D, Collins M, Coulson R, Davies R, Dyer PS, Farman M, Fedorova N, Fedorova N, Feldblum TV, Fischer R, Fosler N, Fraser A, Garcia JL, Garcia MJ, Goble A, Goldman GH, Gomki M, Griffith-Jones S, et al. Genome sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature 2005, 438(7071):1151-1156.

57. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Sczaztocci C, Farman B, Butler J, Purcell S, Harris S, Braus GH, Draft O, Busch S, D’Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffith-Jones S, Doonan JH, Yu J, Venken K, Pain A, Freitag M, et al. Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature 2004, 438(7071):1157-1161.

58. Pel HJ, de Winde J, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albargh R, Albaranga M, Andersen MR, Bendtson JD, Benen JA, van den Berg M, Breestraat S, Caddick MX, Contreras R, Cornell M, Coutsinho PM, Coutsinho EG, Debes AJ, Dekker P, van Dijk PW, Dijkstra L, Driessen AJ, D’enfert C, Geyserns S, Gossens C, Groot GS, et al. Genome sequence and analysis of the versatile cell factory Aspergillus niger CBS 513.88. Nat Biotechnol 2007, 25(2):221-231.

59. Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kumasuto K, Arima T, Akita O, Kashikawa Y, Abe K, Gomi K, Homichi H, Kitamoto K, Kobayashi T, Takeuchi M, Denning DW, Galagan JE, Nieman WC, Yu J, Archer DB, Bennett JW, Bhatnagar D, Cleveland TE, Fedorova ND, Gotoh O, Horikawa H, Hosoya M, Ishimori Y, Ishioka T, Isogai S, et al. Genomic sequence and analysis of Aspergillus oryzae. Nature 2005, 438(7071):1157-1161.

60. Sharpton TJ, Wortman MJ, Hillier LW, Jaffe DB, Moller CA, Noonan VH, Venter JC, et al. The genomic sequence of the filamentous fungus Magnaporthe grisea. Mgnaporthe grisea 2005, 422(6934):980-986.

61. van der Berg MA, Albargh R, Albaranga M, Badger JH, Danan JM, Dreesen AJ, Garcia-Estrada C, Fedorova ND, Harris DM, Heijne WH, Joardar V, Kiel JA, Kovalchuk A, Martin JF, Nieman WC, Nijland JG, Pronk JT, Roubos JA, van der Klei IJ, van Peij NN, Veenhuis M, van Dohren H, Winger J, Wortman JR, et al. Genome sequencing and analysis of the filamentous fungus Penicillium chrysogenum. Nat Biotechnol 2008, 26(10):1161-1168.

62. Espagne E, Lepiniet O, Malagnac F, Da Silva C, Jaillon O, Porcelli B, Coulovou A, Aouj AM, Segurens B, Lethein J, Poultan M, Roerendaal V, Grossetete S, Khall H, Coppen E, Descloward-Chablat M, Picard M, Contamine V, Arnisse S, Bourdais A, Bertheau-Leveiller V, Gauherottet D, de Vries RP, Battaglia E, Coutsinho PM, Dancin EG, Henriksen B, Kohuye RE, Sainsard-Chanet A, Bovin A, Pinan-Lucare B, et al. Genome sequence of the model ascomycete fungus Podospora anserina. Genom Biol 2008, 9(5):R77.

63. Han JK, Lowe RG, Solomon PS, Tan KC, Schoch CL, Spatafora JW, Crous PW, Kodira C, Birren BW, Galagan JE, Torns SF, McDonald BA, Oliver RP. Dothideomycete Plant Interactions Illuminated by Genome Sequencing and EST Analysis of the Wheat Pathogen Stagonospora nodorum. Plant Cell 2007, 19(3):3347-3368.

64. Martinez D, Berk GA, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutsinho PM, Cullen D, Danchin EG, Gregoriev V, Harris P, Jones T, Kobilka CP, Hans C, Ho I, Lomando LF, de Leon AL, Magnusson J, Merino S, Mills N, Nelson C, Putnam N, Robertset B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerham-Pavlinen A, et al. Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nat Biotechnol 2008, 26(5):533-560.

65. Hibbett DS, Binder M, Biechler JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Luking R, Thonston Lumhsech V, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch CL, Spatafora JW, Stalsper JS, Vilgalys R, Aime MC, Aptoot A, Baur R, Begerow D, Benny GL, Castlebury LA, Crous PW, Day VC, Gams W, Geiser DM, et al. A higher level phylogenetic classification of the Fungi. Mycologia 2007, 111(5):599-617.

66. Jones T, Federspiel NA, Crous PW, Dungan J, Kalman S, Magee BB, Newpot G, Thorstenor YR, Agbanan N, Magee PT, Davis RW, Scherer S. The diploid genome sequence of Candida albicans. PLoS Natl Acad Sci USA 2004, 101(19):7329-7334.

67. van het Hoog M, Rasmussen MD, Lin MF, Santos MA, Saktihkumar S, Munro CA, Rhenbya E, Grabher M, Forche A, Reedyt J, Agrapiotif A, Innaud MB, Bates S, Brown AJ, Brunke S, Costanzo MC, Fitzpatrick DA, de Groot PW, Harris D, Heyer LS, Billion KL, Fredman M, Lindahl N, Lugue MC, Martin R, Neiman AM, Nikolaou E, Quinn J, et al. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 2007, 459(7247):657-662.

68. Ducy B, Sherman D, Fischer G, Durens P, Casaregola S, Lafontaine I, De Montiggi J, Murc C, Neugvelig C, Tallia E, Goffard N, Frangeul L, Aigle M, Arndt v, Berulb A, Barbe V, Banyu S, Blanchis C, Becker K, Becken J, Bekkayst A, Besrani E, Boyer J, Cattolli L, Cofrol D, de Danarau A, Despou L, Fabre E, Fairhead C, Femmy-Dumazet H, et al. Genome evolution in yeasts. Nature 2004, 430(6998):35-44.

69. Blandin G, Oztar-Kalstromopolous O, Wincker P, Antiquenove F, Dujon B. Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis. FEMS Lett 2000, 481(1-9):91-94.

70. Dietrich FS, Voegeli S, Brachet S, Lerdh A, Gates K, Steiner S, Mohr C, Anthoovard B, Babour A, Barbe V, Batter V, Banyu S, Blanchin C, Becker K, Beyne J, Blicxken K, Boisn E, Boyer J, Cattolli L, Cofrol D, de Danarau A, Despou L, Fabre E, Fairhead C, Femmy-Dumazet H, et al. Genome evolution in yeasts. Nature 2004, 430(6998):35-44.

71. Kopriou M, Oztar-Kalstromopolous O, Wincker P, Antiquenove F, Dujon B. Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis. FEMS Lett 2000, 481(1-9):91-94.
Genome sequence of the lignocellulose-bioconverting fungus *P. chrysosporium* provides insights into mycorrhizal symbiosis. Nature Biotechnol 2007, 25(3):319-326.

Clifton P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. Science 2003, 301(5629):71-76.

Kells M, Patterson N, Endrizzi M, Brenn B, Lander ES. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 2003, 421(6937):241-254.

Lig G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai II, Bergman CM, Benassai D, O’Kelle MJ, van Oudenaarden A, Barton DB, Ballew E, Nguyen AN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ. Population genomics of domestic and wild yeasts. Nature 2009, 458(7236):337-341.

Doniger SW, Kim HS, Swain D, Corcuela D, Williams M, Yang SP, Fay JC. A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet 2008, 4(8):e1000183.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SC. Life with 6000 genes. Science 1996, 274(5287):563-547.

Gu Z, David L, Petrov D, Jones T, Davis RW, Steinetz BM. Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 2005, 102(4):1092-1097.

Wood V, Gwilliam R, Jourdjian AM, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayler J, Baker S, Bamshad D, Bowman M, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gerties S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, et al. The genome sequence of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Nature 2005, 438(7069):97-101.

Feldbrugge M, Basse CW, Steinberg G, Ibeas JI, Holloman W, Guzman P, Farman M, Stajich JE, Sentandreu R, Gonzalez-Prieto JM, Kennell JC, Feldmann H, Perlin MH, Wosten HA, de Vries R, Ruiz-Herrera J, Peyrin P, Aerts A, Ahren D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Martin DM, et al. Genomic compaction of the eukaryote parasite *Entamoeba histolytica* cunincki. Nature 2001, 414(6862):450-453.

Gardner MJ, Hall N, Fung E, White O, Benimana M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallow SJ, Suh B, Peterson J, Angiuoli S, Perna M, Allen J, Selengut J, Haft D, Matzer MW, Vaidya AB, Martin DM, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 2002, 419(6906):499-510.

Haas BJ, Kamoun S, Zody MC, Jiang RH, Handlaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Tonto-Alaibto T, Bozkurt TO, Ah-Fong AM, Alvarado L, Anderson VL, Armstrong MR, Avwara A, Baxter L, Beynon J, Boeving PC, Bollman SR, Bos J, Bulone V, Cai G, Calik C, Carlington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahigren N, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 2009, 461(7262):393-398.

Tyler BM, Tripathy S, Zhang X, Deyal P, Jiang RH, Aerts A, Amededo FD, Baxter L, Benassai D, Beynon JI, Chapman J, Damasceno CM, Dorrance AE, Dou D, Dickenman AW, Dubchak IL, Garbelotto M, GiJen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Kamps K, Lamour KH, Lee MK, McDonald WH, Medina M, et al. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 2006, 313(5791):1261-1266.

doi:10.1186/1471-2164-11-105

Cite this as: Choi et al.: *Fungal Secretome Database: Integrated platform for annotation of fungal secretomes*. *BMC Genomics* 2010 11:105.