

CFGP 2.0: a versatile web-based platform for supporting comparative and evolutionary genomics of fungi and Oomycetes

Jaeyoung Choi^{1,2}, Kyeongchae Cheong^{1,2}, Kyongyong Jung^{1,2}, Jongbum Jeon^{1,2},
Gir-Won Lee^{1,3}, Seogchan Kang⁴, Sangsoo Kim³, Yin-Won Lee^{2,5} and
Yong-Hwan Lee^{1,2,5,6,7,*}

¹Fungal Bioinformatics Laboratory, ²Department of Agricultural Biotechnology, Seoul National University, Seoul 151-742, ³Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743, Korea, ⁴Department of Plant Pathology and Environmental Microbiology, The Pennsylvania State University, University Park, PA 16802, USA, ⁵Center for Fungal Pathogenesis, ⁶Center for Fungal Genetic Resources and ⁷Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Korea

Received September 12, 2012; Revised October 21, 2012; Accepted October 27, 2012

ABSTRACT

In 2007, Comparative Fungal Genomics Platform (CFGP; <http://cfgp.snu.ac.kr/>) was publicly open with 65 genomes corresponding to 58 fungal and Oomycete species. The CFGP provided six bioinformatics tools, including a novel tool entitled BLASTMatrix that enables search homologous genes to queries in multiple species simultaneously. CFGP also introduced Favorite, a personalized virtual space for data storage and analysis with these six tools. Since 2007, CFGP has grown to archive 283 genomes corresponding to 152 fungal and Oomycete species as well as 201 genomes that correspond to seven bacteria, 39 plants and 105 animals. In addition, the number of tools in Favorite increased to 27. The Taxonomy Browser of CFGP 2.0 allows users to interactively navigate through a large number of genomes according to their taxonomic positions. The user interface of BLASTMatrix was also improved to facilitate subsequent analyses of retrieved data. A newly developed genome browser, Seoul National University Genome Browser (SNUGB), was integrated into CFGP 2.0 to support graphical presentation of diverse genomic contexts. Based on the standardized genome warehouse of CFGP 2.0, several systematic platforms designed to support studies on selected gene families have been developed. Most of them are connected through Favorite to allow of sharing data across the platforms.

INTRODUCTION

Fungal genome sequencing has rapidly increased since the release of the genome sequences of *Saccharomyces cerevisiae* in 1996 (1). With the current and anticipated advances in sequencing technology (2,3), the rate of fungal genome sequencing will continue to accelerate. Currently, there exist more than 300 fully sequenced fungal genomes in the public domain (4,5), with many species and several isolates of previously sequenced species being sequenced (6,7). In addition, the 1000 Fungal Genome project (F1000; <http://1000.fungalgenomes.org/>) will greatly help us to uncover genomic underpinnings of fungal evolution and life styles via large-scale comparative genomics studies. In combination with genomes from plants and animals as well as fungi, in-depth comparative genomics across multiple eukaryotic kingdoms will be facilitated (8–10). To efficiently support such large-scale, genome-based inquiries, it is critical to archive the available genome sequences and annotation information in a standardized format so that they can be easily and efficiently retrieved and analyzed.

To address this need, in 2007, the first version of Comparative Fungal Genomics Platform (CFGP) was released with 65 fungal and Oomycete genomes (11). The CFGP was founded on a new user interface (UI) called Data-driven User Interface (DUI), which made use of its bioinformatics tools and the management of task histories easy and efficient. Since then, the number of genomes archived and bioinformatics tools have grown substantially. (Supplementary Table S1 and Table 1) Furthermore, the standardized genome data warehouse of CFGP has supported the development of multiple platforms that are specialized for supporting the archiving and

*To whom correspondence should be addressed. Tel: +82 2 880 4674; Fax: +82 2 873 2317; Email: yonglee@snu.ac.kr

Table 1. List of bioinformatics tools available in CFGP 2.0

Category	Name	Input data	References
Sequence search	BLAST	Sequences	(12)
	BLAST2	Sequences	(12)
	BLASTMatrix	Sequences	(11)
ID of functional domains	InterPro Scan	Sequences	(13)
Phylogenetic analysis	ClustalW	Sequences	(14)
	DNAML	Alignment	(15)
	PROML	Alignment	(15)
	DNAPARS	Alignment	(15)
	PROTPARS	Alignment	(15)
	PHYML	Alignment	(16)
ID of secreted proteins	SignalP 3.0	Sequences	(17)
	SigCleave	Sequences	(18)
	SigPred	Sequences	(19)
	RPSP	Sequences	(20)
	SecretomeP	Sequences	(21)
Subcellular localization	PSortII	Sequences	(22)
	predictNLS	Sequences	(23)
	ChloroP	Sequences	(24)
	TargetP	Sequences	(25)
Prediction of trans-membrane helices	TMHMM2	Sequences	(26)
Prediction of RNA secondary structure	tRNAScan-SE	Sequences	(27)
	mFold 3.2	Sequences	(28)
Post-translational modification	NetCGlyc	Sequences	(29)
	NetNGlyc	Sequences	–
	NetOGlyc	Sequences	(30)
	NetPhos	Sequences	(31)
Conserved motif search	MEME	Sequences	(32)

A total of 27 tools in nine categories are available via the Favorite browser of CFGP 2.0.

analysis of specific gene families and functional groups (Table 2). Some of these platforms share the Favorite of CFGP to provide an efficient mechanism for sharing data with CFGP and to enable the use of its bioinformatics tools for a variety of analyses. In this article, we outline the improvements made in CFGP 2.0 and how its standardized genome warehouse has been exploited in development of other comparative genomics platforms.

METHODS

System design

The structure and core databases of CFGP 2.0 are basically identical to those used for the first version. The system consists of databases including a standardized genome warehouse, wrapper programs written by the Perl and C languages and DUI. To balance the server load so as to ensure a more efficient operation of the system, its core databases were distributed in multiple servers, and more web servers were added. The MySQL relational database management system was used to manage and curate the data. Its web interfaces were written in PHP with javascript, and analysis functions in Favorite Browser

Table 2. List of online platforms and tools supporting studies on specific gene families or functional groups^a

Name	URL	Reference
Cyber-infrastructure for <i>Fusarium</i>	http://www.fusariumdb.org/	(33)
Fungal Transcription Factor Database	http://ftfd.snu.ac.kr/	(34)
Fungal Cytochrome P450 Database	http://p450.riceblast.snu.ac.kr/	(35)
Seoul National University Genome Browser	http://genomebrowser.snu.ac.kr/	(36)
The Systematic Platform for Identifying Mutated Proteins	http://pimp.starflr.info/	(37)
Insect Mitochondrial Genome Database	http://www.imgd.org/	(38)
Fungal Secretome Database	http://fsd.snu.ac.kr/	(39)
Eukaryotic DNAJ/K Database	http://edd.snu.ac.kr/	–
Cell Wall-degrading Enzyme Database	http://www.cwde.org/	–

^aThese platforms were developed using the standardized genome warehouse and the web template engine of CFGP 2.0.

were relayed by Perl scripts and automatically coordinated by the system monitoring servers.

Mining orthologs

The source code of InParanoid 4.1 was used for the identification of orthologs in the archived proteomes. First, the genomes of 35 species that are frequently utilized were subjected to ortholog identification (Supplementary Table S1). All pairwise comparisons of data from these 35 species were carried out. The latest version of InParanoid 7 provides orthologs from 100 eukaryotic genomes. However, some genomes that we have used were not included in the latest version; data for those species were downloaded from CFGP 2.0 and subsequently subjected to ortholog identification.

EXPANDED GENOME DATA WAREHOUSE

The standardized genome warehouse of CFGP has been substantially expanded in both the number of species and taxonomic coverage. In addition to 283 genomes corresponding to 152 fungal and Oomycete species, 39 plant and 105 animal genomes have also been archived (Figure 1 and Supplementary Table S2). The animal and plant genomes were incorporated to enable comparative evolutionary genomics studies across multiple eukaryotic kingdoms.

ENHANCED UTILITY AND NEW FEATURES

Improved UI

The UI of CFGP 2.0 was greatly improved to provide better user experience. All modifications followed the HTML5 and CSS3 standards, on which most widely used web browsers are based. Three main application/utility

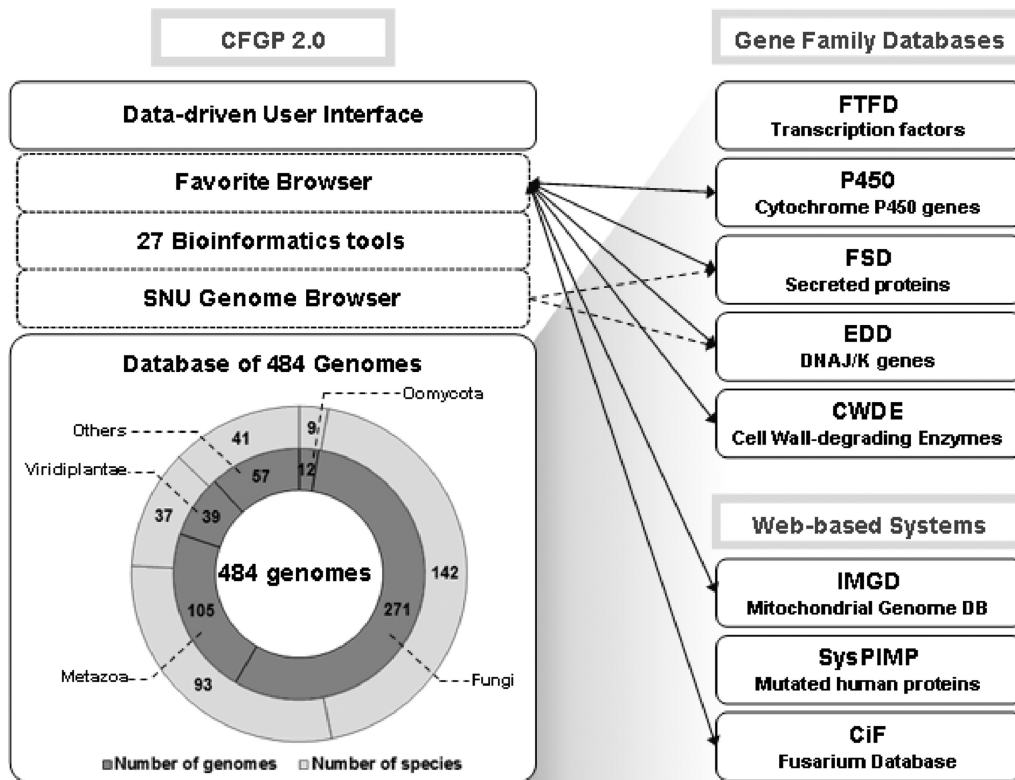


Figure 1. A diagram illustrating the system architecture and the content of the genomes archived in CFGP 2.0. Key features of CFGP 2.0 were depicted on the left. The web-based platforms that have been developed based on the standardized genome warehouse of CFGP 2.0 are listed on the right. Bidirectional arrows indicate that they support the Favorite Browser, which synchronizes with CFGP 2.0. Dashed arrows denote that SNUGB was integrated in two platforms, FSD and EDD. In the pie graph, the inner and outer circles represent the number of genomes and species for each taxon, respectively.

CFGP Comparative Fungal Genomics Platform

283 genomes from 152 fungal species

amethys1(jeeyoung Choi) is logged in. Modify MyInfo | Logout | Withdrawal
About | Notice | Tasks | Publications | Guide | User Scenario | Sitemap | State | Contact

SEQUENCE HISTORIES DOMAINS ANALYSIS FAVORITE SNUGB

Favorite frame

12377
1 objects Updated @ Jan 16, 2012

Saccharo10
5 objects Updated @ Jan 10, 2012

Zygo15
4 objects Updated @ Jan 10, 2012

Peizo17
4 objects Updated @ Jan 10, 2012

Zygo14
5 objects Updated @ Jan 10, 2012

B202
6 objects Updated @ Jan 10, 2012

Saccharo9
6 objects Updated @ Jan 10, 2012

Saccharo8
7 objects Updated @ Jan 10, 2012

Saccharo7
5 objects Updated @ Jan 10, 2012

Ventriillum01
5 objects Updated @ Jan 10, 2012

Basidi02
4 objects Updated @ Jan 10, 2012

Fusarium08
4 objects Updated @ Jan 10, 2012

Fusarium07
6 objects Updated @ Jan 10, 2012

Fusarium06
4 objects Updated @ Jan 10, 2012

Fusarium04
5 objects Updated @ Jan 10, 2012

Presentation frame

Magnaporthe oryzae 8
supercont8.5 (Total Length : 4,490,059bp)

Magnifying Option Zoom In [+2x] +4x [Select Range] Zoom Out [-1.5x] -3x [View All] [Im]

Range Option Start 333,678 bp End 354,544 bp [Refresh]

Viewing Option Contig Region Squish Transcript [Mask] View ORF Option Display View Exon/Intron Option Display [Refresh]

Additional Viewing Option [Refresh]

Frame toggle buttons

EST Region None FSD Ratio None
rRNA None FSD Pattern None
SPSP None SigCleave None
TRNRM None

SignaIP None SigPred [Refresh]

Application frame

Contig Browser

List of proteins

This page displays the list of proteins belonging to the selected contig dataset.

Dataset Name : Magnaporthe oryzae 8

Search by Protein name acetylxyylan esterase Search View All

Chk	Locus name	Gene	Protein name	Length(aa)	Contig	Ch#	Function
<input checked="" type="checkbox"/>	MGG_011467D	acetylxyylan esterase		229	5	0	[SNUGB]
<input type="checkbox"/>	MGG_098407D	acetylxyylan esterase 2		299	6	0	[SNUGB]

There are 2 proteins in total.

Page 1 of 1

10 items per page

Copyright © 2009-2011 Fungal Bioinformatics Laboratory / Fungal Plant Pathology Laboratory / Supported by SPS

Last Update : 2012-07-20 15:27:36 Total Access : 43,393
This site is compatible to at 1024x768.

Figure 2. Outline of the improved DUI of CFGP 2.0. The frames outlined in blue, orange and purple correspond to the Favorite, presentation and application frame, respectively. The buttons boxed in red are to hide or show Favorite and presentation frames, respectively.

frames were rearranged to make the switch from one frame to another more intuitive. The Favorite and presentation frames can be toggled for more flexible browsing (Figure 2). All web pages have been thoroughly tested with multiple web browsers, including Chrome, Firefox, Internet Explorer and Safari.

Taxonomy browser

As the number of species representing diverse taxa continues to increase, a browsing tool based on simple text search (e.g. species name) was not sufficient to help users to efficiently explore available genomes. The Taxonomy Browser implemented in CFGP 2.0 provides predictive text feature as well as a hierarchical tree-based taxon

structure to show the taxonomic position of a chosen species. Once a specific species is selected, the number of genomes and available sequence types are listed with direct links to corresponding sequences.

Seoul National University Genome Browser

The first version of CFGP did not offer a graphical interface to present the genomic context and notable features such as GC content, functional domains and the signal peptide. This new genome browser implemented in CFGP 2.0 enabled users to view such information in the chosen region. The target region of viewing can be selected by assigning the start and end positions with a mouse or by typing in its genome coordinate (Figure 3).

Magnaporthe oryzae 8

supercont8.6 (Total Length : 4,133,993bp)



Figure 3. A screenshot of SNUGB implemented in CFGP 2.0. SNUGB allows users to view 13 biological features and the gene structures in the selected genome region. Those features include GC contents, functional domains, nuclear localization signals, signal peptides and trans-membrane helices.

New bioinformatics tools added to the Favorite Browser

Compared with the first version that only provided six tools, CFGP 2.0 is equipped with 27 tools covering nine categories of data analysis or viewing (Table 1). This addition enables users to perform more analyses without leaving CFGP.

Ortholog browsing function

Finding orthologs for a specific gene in multiple species often requires numerous BLAST searches and validation processes. In the first version of CFGP, we tried to eliminate copy-and-paste of sequences by incorporating the DUI (11). In CFGP 2.0 we simplified the identification and collection of orthologs by offering pre-computed data. There are several ortholog identification programs such as InParanoid (40), Ortholog, MSOAR (41) and THOR (42). We adopted InParanoid to identify orthologs via pairwise comparisons among 35 frequently accessed genomes. For every protein sequence encoded by each of these 35 genomes, orthologous genes in the other 34 genomes are provided to allow a quick overview of its distribution among these species and also to support their further analyses by saving them into a Favorite on the fly.

FUNCTIONAL/EVOLUTIONARY GENOMICS PLATFORMS DEVELOPED BASED ON THE STANDARDIZED GENOME WAREHOUSE OF CFGP 2.0

Via the use of the standardized genome warehouse of CFGP 2.0, a number of platforms that aim to support comparative analyses of specific gene families and/or functional groups have been developed: (i) Cyber-infrastructure for *Fusarium* (CiF; <http://www.fusariumdb.org/>) (33), (ii) Fungal Transcription Factor Database (FTFD; <http://ftfd.snu.ac.kr/>) (34), (iii) Fungal Cytochrome P450 Database (FCPD; <http://p450.riceblast.snu.ac.kr/>) (35), (iv) Fungal Secretome Database (FSD; <http://fsd.snu.ac.kr/>) (39), (v) Eukaryotic DNAJ and DNAK Database (EDD; <http://edd.snu.ac.kr/>) (Cheong *et al.*, manuscript in preparation) and (vi) Cell Wall-degrading Enzymes Database (CWDE; <http://www.cwde.org/>) (Choi *et al.*, manuscript in preparation). The Seoul National University Genome Browser (SNUGB) (<http://genomebrowser.snu.ac.kr/>) (36) was also implemented in FSD and EDD. The Insect Mitochondrial Genome Database (IMGD; <http://www.imgd.org/>) (38) employs the Species-driven UI, which enables intuitive and fast taxonomical browsing with multiple add-on analysis functions. Finally, the Systematic Platform for Identifying Mutated Proteins (SysPIMP; <http://pimp.starflr.info/>) (37) was developed to support the identification of mutations related to human diseases. The Favorite Browser of CFGP 2.0 is connected with many of those platforms to efficiently support data exchange and sharing across multiple platforms. All the data saved in the Favorite Browser are synchronized in real-time so that users can fully exploit data and functions provided by these platforms.

FUTURE DIRECTIONS

To keep up with the rapidly released and updated eukaryotic genomes, CFGP 2.0 will be updated on a regular basis. We will integrate more useful modules, software or interface scheme to continuously improve the environment for users conducting comparative and evolutionary genomics studies. In order to support efforts to uncover possible functions of many hypothetical genes, the ortholog information database will be expanded by adding the corresponding information from more species.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2.

FUNDING

National Research Foundation of Korea grant funded by the Korea government [2012-0001149 and 2012-0000141]; TDPAF [309015-04-SB020]; Next-Generation BioGreen 21 Program of Rural Development Administration in Korea [PJ00821201]; a graduate fellowship through the Brain Korea 21 Program (to J.C., K.C. and J.J.). Funding for open access charge: Seoul National University.

Conflict of interest statement. None declared.

REFERENCES

- Goffeau,A., Barrell,B.G., Bussey,H., Davis,R.W., Dujon,B., Feldmann,H., Galibert,F., Hoheisel,J.D., Jacq,C., Johnston,M. *et al.* (1996) Life with 6000 genes. *Science*, **274**, 546, 563–567.
- Metzker,M.L. (2010) Sequencing technologies — the next generation. *Nat. Rev. Genet.*, **11**, 31–46.
- Hawkins,R.D., Hon,G.C. and Ren,B. (2010) Next-generation genomics: an integrative approach. *Nat. Rev. Genet.*, **11**, 476–486.
- Grigoriev,I.V., Nordberg,H., Shabalov,I., Aerts,A., Cantor,M., Goodstein,D., Kuo,A., Minovitsky,S., Nikitin,R., Ohm,R.A. *et al.* (2012) The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res.*, **40**, D26–D32.
- Keyhani,N.O. (2011) Fungal genomes and beyond. *Fungal. Genom. Biol.*, **1**, e101.
- Liti,G., Carter,D.M., Moses,A.M., Warringer,J., Parts,L., James,S.A., Davey,R.P., Roberts,I.N., Burt,A., Koufopanou,V. *et al.* (2009) Population genomics of domestic and wild yeasts. *Nature*, **458**, 337–341.
- Gan,X., Stegle,O., Behr,J., Steffen,J.G., Drewe,P., Hildebrand,K.L., Lyngsoe,R., Schultheiss,S.J., Osborne,E.J., Sreedharan,V.T. *et al.* (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature*, **477**, 419–423.
- Cornell,M.J., Alam,I., Soanes,D.M., Wong,H.M., Hedeler,C., Paton,N.W., Rattray,M., Hubbard,S.J., Talbot,N.J. and Oliver,S.G. (2007) Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi. *Genome Res.*, **17**, 1809–1822.
- Richards,T.A., Soanes,D.M., Foster,P.G., Leonard,G., Thornton,C.R. and Talbot,N.J. (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell*, **21**, 1897–1911.
- van Dam,T.J., Rehmann,H., Bos,J.L. and Snel,B. (2009) Phylogeny of the CDC25 homology domain reveals rapid differentiation of Ras pathways between early animals and fungi. *Cell. Signal.*, **21**, 1579–1585.
- Park,J., Park,B., Jung,K., Jang,S., Yu,K., Choi,J., Kong,S., Kim,S., Kim,H., Kim,J.F. *et al.* (2008) CFGP: a web-based,

- comparative fungal genomics platform. *Nucleic Acids Res.*, **36**, D562–D571.
12. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezukh, Y., McGinnis, S. and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.
 13. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D211–D215.
 14. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
 15. Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
 16. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
 17. Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, **340**, 783–795.
 18. Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
 19. Bradford, J.R. (2001) Protein design for biopharmaceutical development at glaxosmithkline: in silico methods for prediction of signal peptides and their cleavage sites, and linear epitopes. *Thesis (MRes)*. University of Leeds (School of Biochemistry and Molecular Biology), Leeds, UK.
 20. Plewczynski, D., Slabinski, L., Ginalski, K. and Rychlewski, L. (2008) Prediction of signal peptides in protein sequences by neural networks. *Acta Biochim. Pol.*, **55**, 261–267.
 21. Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G. and Brunak, S. (2004) Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.*, **17**, 349–356.
 22. Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, **35**, W585–W587.
 23. Cokol, M., Nair, R. and Rost, B. (2000) Finding nuclear localization signals. *EMBO Rep.*, **1**, 411–415.
 24. Emanuelsson, O., Nielsen, H. and Von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.*, **8**, 978–984.
 25. Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.*, **300**, 1005–1016.
 26. Sonnhammer, E.L., von Heijne, G. and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **6**, 175–182.
 27. Schattner, P., Brooks, A.N. and Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.*, **33**, W686–W689.
 28. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
 29. Julenius, K. (2007) NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. *Glycobiology*, **17**, 868–876.
 30. Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.*, **15**, 115–130.
 31. Blom, N., Gammeltoft, S. and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.*, **294**, 1351–1362.
 32. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
 33. Park, B., Park, J., Cheong, K.C., Choi, J., Jung, K., Kim, D., Lee, Y.H., Ward, T.J., O'Donnell, K., Geiser, D.M. *et al.* (2011) Cyber infrastructure for *Fusarium*: three integrated platforms supporting strain identification, phylogenetics, comparative genomics and knowledge sharing. *Nucleic Acids Res.*, **39**, D640–D646.
 34. Park, J., Jang, S., Kim, S., Kong, S., Choi, J., Ahn, K., Kim, J., Lee, S., Park, B., Jung, K. *et al.* (2008) FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics*, **24**, 1024–1025.
 35. Park, J., Lee, S., Choi, J., Ahn, K., Park, B., Kang, S. and Lee, Y.H. (2008) Fungal cytochrome P450 database. *BMC Genomics*, **9**, 402.
 36. Jung, K., Park, J., Choi, J., Park, B., Kim, S., Ahn, K., Choi, D., Kang, S. and Lee, Y.H. (2008) SNUGB: a versatile genome browser supporting comparative and functional fungal genomics. *BMC Genomics*, **9**, 586.
 37. Xi, H., Park, J., Ding, G., Lee, Y.H. and Li, Y. (2009) SysPIMP: the web-based systematical platform for identifying human disease-related mutated sequences from mass spectrometry. *Nucleic Acids Res.*, **37**, D913–D920.
 38. Lee, W., Park, J., Choi, J., Jung, K., Park, B., Kim, D., Lee, J., Ahn, K., Song, W., Kang, S. *et al.* (2009) IMGD: an integrated platform supporting comparative genomics and phylogenetics of insect mitochondrial genomes. *BMC Genomics*, **10**, 148.
 39. Choi, J., Park, J., Kim, D., Jung, K., Kang, S. and Lee, Y.H. (2010) Fungal secretome database: integrated platform for annotation of fungal secretomes. *BMC Genomics*, **11**, 105.
 40. Ostlund, G., Schmitt, T., Forslund, K., Kostler, T., Messina, D.N., Roopra, S., Frings, O. and Sonnhammer, E.L. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.*, **38**, D196–D203.
 41. Fu, Z., Chen, X., Vacic, V., Nan, P., Zhong, Y. and Jiang, T. (2007) MSOAR: a high-throughput ortholog assignment system based on genome rearrangement. *J. Comput. Biol.*, **14**, 1160–1175.
 42. Bainbridge, M.N., Warren, R.L., He, A., Bilenky, M., Robertson, A.G. and Jones, S.J. (2007) THOR: targeted high-throughput ortholog reconstructor. *Bioinformatics*, **23**, 2622–2624.