Genomics & Informatics Vol. 4(4) 147-160, December 2006

A Comparative Genome-Wide Analysis of GATA Transcription Factors in Fungi

Jongsun Park[1,2], Hyojeong Kim[2], Soonok Kim[3,4], Sunghyung Kong[1,2], Jaejin Park[1], Seryun Kim[1], Hyea-young Han[1], Bongsoo Park[1], Kyongyong Jung[1,2] and Yong-Hwan Lee[1,2,3,4]*

[1]Fungal Bioinformatics Laboratory, [2]Department of Agricultural Biotechnology, [3]Center for Fungal Genetic Resources, and [4]Center for Agricultural Biomaterials, Seoul National University, Seoul 151-921, Korea

Abstract

GATA transcription factors are widespread eukaryotic regulators whose DNA-binding domain is a class IV zinc finger motif in the form $CX_2CX_{17-20}CX_2C$ followed by a basic region. In fungi, they act as transcriptional activators or repressors in several different processes, ranging from nitrogen source utilization to mating-type switching. Using an in-house bioinformatics portal system, we surveyed 50 fungal and 9 out-group genomes and identified 396 putative fungal GATA transcription factors. The proportion of GATA transcription factors within a genome varied among taxonomic lineages. Subsequent analyses of phylogenetic relationships among the fungal GATA transcription factors, as well as a study of their domain architecture and gene structure, demonstrated high degrees of conservation in type IVa and type IVb zinc finger motifs and the existence of distinctive clusters at least at the level of subphylum. The SFH1 subgroup with a 20-residue loop was newly identified, in addition to six well-defined subgroups in the subphylum Pezizomycotina. Furthermore, a novel GATA motif with a 21-residue loop ($CX_2CX_{21}CX_2C$, designated 'zinc finger type IVc') was discovered within the phylum Basidiomycota. Our results suggest that fungal GATA factors might have undergone multiple distinct modes of evolution resulting in diversified cellular modulation in fungi.

Keywords: GATA transcription factor, Fungi, Comparative analysis, Phylogenetics, Type IVc zinc finger

*Corresponding author: E-mail yonglee@snu.ac.kr, Tel +82-2-880-4674, Fax +82-2-873-2317
Accepted 5 Dec 2006

Introduction

GATA transcription factors (TFs) are a family of regulatory proteins widespread in eukaryotes. They contain one or two highly conserved type IV zinc finger motifs ($CX_2CX_{17-20}CX_2C$) followed by a basic region that recognize and bind to a consensus DNA sequence, WGATAR (W=T or A; R=G or A), of target genes (Scazzocchio, 2000). Based on the spacing between the cysteine pairs at the zinc finger loop, the zinc finger motifs have been categorized into two groups: those with 17-residue loops ($CX_2CX_{17}CX_2C$ zinc finger type IVa) and those with 18-residue loops ($CX_2CX_{18}CX_2C$ zinc finger type IVb Teakle and Gilmartin, 1998). The DNA-binding domains of characterized metazoan GATA TFs always have type IVa zinc fingers with a leucine in the seventh position of the loop. In vertebrates, six distinct GATA TFs are well-defined and they play critical roles in development, differentiation, and control of cell proliferation (for review, see Lowry and Atchley, 2000). In contrast, all the characterized GATA factors in plants contain type IVb zinc finger motifs with either leucine or glutamine in the seventh position of the loop and an additional residue lysine in position 16. Plant GATA TFs have been implicated in light-dependent and nitrate-dependent control of transcription, although their *in vivo* functions remain very poorly understood (for review, see Reyes *et al.*, 2004).

In fungi, all GATA TFs contain a highly conserved zinc finger DNA-binding domain and the majority of the fungal GATA TFs fall into two different categories: one is 'animal-like' and the second is 'plant-like.' The residue at position seven of plant-like fungal GATA factors is glutamic acid. Fungal GATA TFs are involved in diverse functions such as nitrogen control, siderophore biosynthesis, light-regulated photomorphogenesis, circadian regulation, and mating-type switching (Scazzocchio, 2000). Recently, 20- and 19-residue zinc finger loops were found in *Saccharomyces cerevisiae* (Maxon and Herskowitz, 2001) and *Candida albicans* (Munchow *et al.*, 2002), respectively. Fungal and plant GATA TFs typically contain a single zinc finger, and animal GATA TFs typically have two or more zinc fingers, although exceptions occur (Teakle and Gilmartin, 1998; Scazzocchio, 2000; Reyes *et al.*, 2004).

Since the completion of the *S. cerevisiae* genome sequencing project in 1996, over 50 other fungal genome sequences including *Neurospora crassa* (Borkovich *et al.*, 2004), *Magnaporthe grisea* (Dean *et al.*, 2005), *Aspergillus*

nidulans (Galagan *et al.*, 2005), and *Ustilago maydis* (Kamper *et al.*, 2006) have been completed at a phenomenal rate, and another 20 projects are underway. These projects have generated considerable numbers of hypothetical proteins with as-yet-undefined biological functions, indicating that accurate automated annotation techniques and functional prediction are critical to fungal genome annotation and function analysis.

In this study, we constructed a pipeline that catalogs and annotates predicted functions of hypothetical proteins through an in-house bioinformatics portal system (CROSSFPP; Park *et al.*, unpublished). The CROSSFPP system provides a common gateway to access genome databases and bioinformatics tools like InterPro, ClustalW, and BLAST. Using our pipeline system, we conducted a genome-wide survey of GATA TF-related sequences in 50 fungal species and 9 out-group species, and identified 396 putative fungal GATA TFs. Subsequent analyses of evolutionary relationships among the fungal GATA TFs demonstrated that type IVa and type IVb zinc finger motifs are highly conserved, and GATA TFs in the subphylum Pezizomycotina can be classified into seven distinctive phylogenetic subgroups, including a newly identified SFH1subgroup with a 20-residue loop. In addition, a novel GATA motif with a 21-residue loop ($CX_2CX_{21}CX_2C$, designated 'zinc finger type IVc') was discovered within the phylum Basidiomycota. Our results suggest that fungal GATA TFs underwent multiple distinct modes of evolution such as selective loss or duplication, or sequence divergence. Furthermore, we provide a new strategy to predict the putative roles of GATA TFs by comparative genomics analysis.

Materials and Methods

Identification of GATA TFs

Genome sequences of 50 fungi and 9 out-groups were used for analysis of fungal GATA TFs and their phylogenetic relationships. Species and database origins of the 59 genome sequences are listed in Table 1. For efficient analyses, the diversely formatted 59 genome datasets were standardized to the same format and integrated into a dedicated database (MySQL database) in which each protein sequence of an annotated gene was treated as an object in a single pipeline (Fig. 1). The standardization process was applied to chromosome (or contig) sequences and nucleotide sequences of annotated genes.

Putative GATA TFs were identified using InterProScan (Zdobnov and Apweiler, 2001) installed in the CROSSFPP system (Fig. 1; Park *et al.*, unpublished). In total, 740,984 annotated protein sequences retrieved from 59 genome databases (470,617 from 50 fungal genomes) were

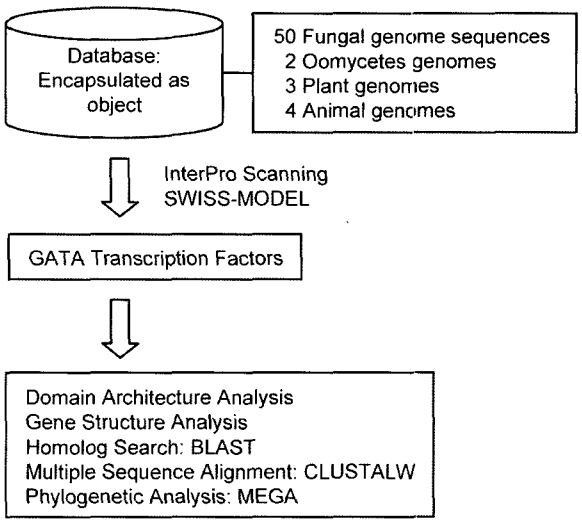

Fig. 1. Schematic diagram of genome-wide identification and phylogenetic analysis of GATA transcription factors from 50 fungal and 9 out-group species. A dedicated database was constructed by integrating all annotated genes from 59 genome sequences and scanned using InterPro to identify putative GATA transcription factors. Selected transcription factors were subjected to domain and gene structures and phylogenetic analysis.

scanned for domains denoted "GATA-type zinc finger" with accession number IPR000679. The results were stored in the InterPro annotation database in the CROSSFPP system for further analysis and all data collected in this study are available at http://ftfd.snu.ac.kr/GATA/. Selected proteins were verified as true GATA TFs by several criteria, for example, the length of GATA motifs (>40 amino acids), the three-dimensional structure of zinc fingers using SWISS-MODEL (Schwede *et al.*, 2003), and the existence of similar motifs like the Arf GTPase activating domain (Vitale *et al.*, 1998).

Multiple sequence alignment and phylogenetic analysis

Amino acid sequences of GATA TFs were aligned using ClustalW 1.83 (Thompson *et al.*, 1994) and MEGA3 (Kumar *et al.*, 2004). Phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei, 1987) and viewed in TreeViewX (http://darwin.zoology.gla. ac.uk/~rpage/treeviewx/index.html).All sequence alignments were tested with a bootstrap method using 10,000 repetitions.

BLAST search for further analyses of homologous genes

BLAST searches (McGinnis and Madden, 2004) were carried out to find genes/proteins with sequence similarity

to the identified GATA TFs. Fifty-nine individual species datasets and one combined dataset were created and used in various BLAST searches. The e-value threshold was set at 1e-10 due to the innate similarity of GATA motifs around 1e-7.

Analysis of domain architecture and exon–intron structure of GATA TFs

Interspecies comparison of identified GATA TFs was performed at the domain architecture and gene structure levels. Domain architecture information on predicted GATA TFs was obtained from InterProScan. Annotated DNA-binding domains and other functional domains were graphically displayed over full-length proteins. Gene structures of the GATA TFs were also graphically visualized using exon and intron information provided from genome databases.

Results and Discussion

Identification of GATA transcription factors among 50 fungal species

Fifty-nine publicly available genome sequences, including 50 fungal species, were incorporated into a dedicated database in a standardized format to manipulate different forms of data with the same pipeline. Genes with IPR000679 denoted as "GATA-type zinc finger" were retrieved and their putative structures were analyzed via SWISS-MODEL (Fig. 1). Through this process, 396 genes from 50 fungal species, 159 genes from 3 plant and 4 animal species, and 1 gene from two oomycetes were identified as GATA TFs (Table 1). The number of fungal GATA TFs in each species ranged from 3 to 16, which is relatively fewer than in plant and metazoan species. The average number of GATA TFs in seven species belonging to the phylum Basidiomycota was about 9.9, and that of *Rhizopus oryzae* (the phylum Zygomycota) was 37, the largest number of GATA TFs in a single fungus species. Species belonging to the phylum Ascomycota contained about 6.9 GATA TFs on average. The average number of GATA TFs among the three subphyla within the phylum Ascomycota was similar; i.e., the subphyla Taphrinomycotina, Saccaromycotina, and Pezizomycotina contained on average 4 (1 species), 7.7 (22 species), and 6.2 (19 species) GATA TFs, respectively. However, the proportion of GATA TFs within the genome in each species differed between the subphyla Saccharomycotina and Pezizomycotina (Fig. 2). GATA TFs in Saccharomycotina represented 0.12% of the genome content on average, ranging from 0.08 to 0.21%. The percentage in Pezizomycotina ranged from 0.04 to 0.08%, with an average of 0.05%, which is one-third of

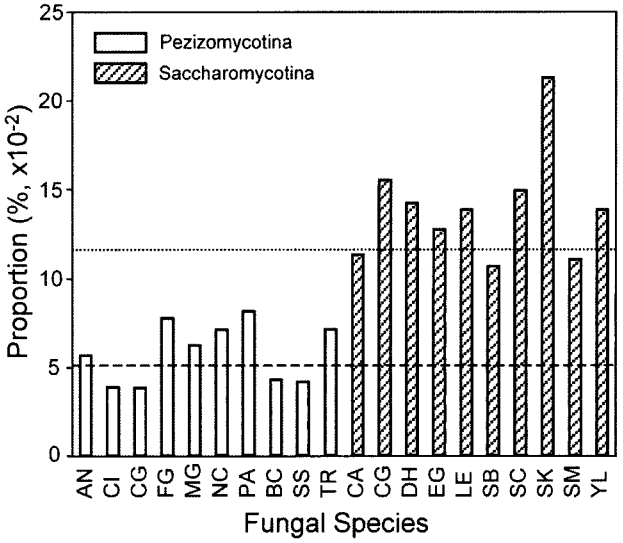

Fig. 2. Comparative analysis of GATA transcription factors in the genomes of Pezizomycotina and Saccharomycotina. The proportion expresses the number of GATA transcription factors relative to the total number of annotated genes in the genome. Dashed lines indicate the averages of GATA TF proportions in the genomes of Pezizomycotina (upper) and Saccharomycotina (lower). Species names are abbreviated as follows: AN, *Aspergillus nidulans*; CI, *Coccidioides immitis*; CG, *Chaetomium globosum*; FG, *Fusarium graminearum*; MG, *Magnaporthe grisea*; NC, *Neurospora crassa*; PA, *Podospora anserina*; BC, *Botrytis cinerea*; SS, *Sclerotinia sclerotiorum*; TR, *Trichoderma reesei*; CA, *Candida albicans*; CG, *Candida glabrata*; DH, *Debaryomyces hansenii*; EG, *Eremothecium gossypii*; LE, *Lodderomyces elongisporus*; SB, *Saccharomyces bayanus*; SC, *Saccharomyces cerevisiae*; SK, *Saccharomyces kudriavzevii*; SM, *Saccharomyces mikatae*; YL, *Yarrowia lipolytica*.

that in Saccharomycotina. This result shows that GATA TFs in Pezizomycotina might have increased their complexity rather than number in order to modulate more diversified cellular processes.

Phylogenetic analysis of fungal GATA factors

A phylogenetic tree of all 556 GATA factors identified in this study was constructed in an attempt to reveal the relationships among putative GATA TFs, in particular the phylogenetic position of fungal GATA TFs relative to those of animals and plants. In agreement with previous studies (Teakle and Gilmartin, 1998 Scazzocchio, 2000), the phylogenetic positions of fungal GATA TFs were different from those of animal and plant GATA TFs. Most fell into independent clusters. When their domain arrangement was examined, animal GATA TFs from four species contained additional domains for DNA binding, like the C_2H_2 zinc finger and Myb DNA-binding domain, but not

additional domains for different functions. However, plant GATA TFs from three species included additional domains for different functions, like a CCT domain involved in light signal transduction (Hayama and Coupland, 2003), but none for DNA binding. Similar to plant GATA TFs, fungal GATA TFs also contained extra domains with various functions, such as PAS, which acts as a sensor in many signaling proteins (Hefti *et al.*, 2004), and were devoid of other DNA-binding domains. Interestingly, only one GATA TF was identified in *Phytophthora sojae* and none in *Phytophthora infestans* (Table 1). A further analysis indicated that these two oomycetes also lacked fungal-specific transcription factors like APSES proteins (data not shown; Sheppard *et al.*, 2005), supporting the idea that the oomycetous *Phytophthora* species, which were once classified as fungi, are the most distantly related organisms from true fungi (James *et al.*, 2006).

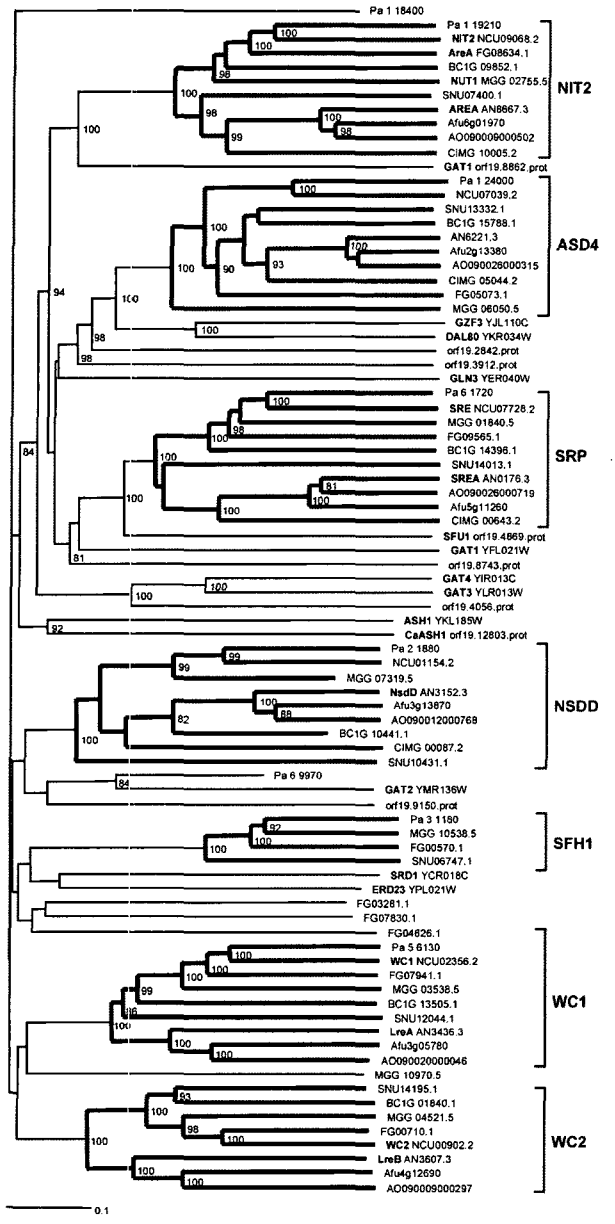

Fig. 3. Simplified phylogenetic tree of GATA transcription factors from thesubphylum Pezizomycotina, with *S. cerevisiae* and *C. albicans*. Full-length protein sequences were used to construct a phylogenetic tree bootstrap values from 10,000 replicates are shown. Members of Pezizomycotina in sevensubgroups are indicated with bold lines. The scale bar corresponds to 0.1 estimated amino acid substitutions per site.

Within the kingdom Fungi, GATA TFs exhibited different properties in different phyla. In the phylum Zygomycota, which is thought to be basal to Ascomycota and Basidiomycota (Cavailer-Smith, 1987), *R. oryzae* appeared to possess the largest number of GATA TFs among fungi (Table 1) and showed a different evolutionary preference from others. Since only one Zygomycota has been sequenced to date, most GATA TFs failed to be grouped with any other fungal GATA TFs; only 12 of 37 were grouped in three clusters. Further analyses of their domain architecture and gene structures suggest that *R. oryzae* might have undergone several duplication events at a relatively early stage.

In the phylum Basidiomycota, 69 GATA TFs were identified from seven genome sequences (Table 1), among which only Urbs1 of *U. maydis* and Cwc2 of *Cryptococcus neoformans* have been characterized (Voisard *et al.*, 1993, Lu *et al.*, 2005). Those GATA TFs formed seven clusters: two including GATA TFs from other phyla and with a known function, and five including GATA TFs only within this phylum and without any previously defined GATA TFs. Consistent with its function in the regulation of ferrichrome-type siderophore biosynthesis (Voisard *et al.*, 1993), *U. maydis* Urbs1 was loosely clustered with other fungal siderophore-regulating TFs (the SRP group in Fig. 3). *C. neoformans* Cwc2, homologous to *N. crassa* WC-2 (Lu *et al.*, 2005; discussed below), assembled with other WC-2-like GATA TFs (the WC2 group in Fig. 3).

The GATA TFs of 42 species belonging to the phylum Ascomycota were assembled into distinctive phylogenetic clusters, which were well supported by high bootstrap values at the level of subphylum. While most clusters included GATA TFs from both subphyla Pezizomycotina (filamentous ascomycetes) and Saccharomycotina (yeast form ascomycetes), several GATA TFs were observed only in a specific subphylum. In Pezizomycotina, a new group of GATA transcription factor, SFH1, was identified in addition to six subgroups of well-defined GATA TFs all seven subgroups are extensively discussed below. In Saccharomycotina,

Table 1. Genome-wide distribution of GATA transcription factors in 50 fungal and 9 other species

Species	# of GATA factors	# of GATA motifs	Motif Type[a]				Source[b]	Reference[c]
			IVa	IVb	IVc	IVe		
Fungi (Kingdom)								
Ascomycota (Phylum)								
Saccharomycotina (Subphylum)								
Candida albicans	16	18	12	4	2	0	SGTC	(Jones *et al.*, 2004)
Candida dubliniensis	6	7	5	2	0	0	SI	-
Candida glabrata	8	8	4	3	1	0	CBS	(Dujon *et al.*, 2004)
Candida guilliermondii	5	6	6	0	0	0	BI	-
Candida lusitaniae	7	8	6	2	0	0	BI	-
Candida tropicalis	7	8	6	2	0	0	BI	-
Debaryomyces hansenii	9	10	6	3	1	0	CBS	(Dujon *et al.*, 2004)
Eremothecium gossypii	6	6	3	2	1	0	NCBI	(Dietrich *et al.*, 2004)
Kluyveromyces lactis	4	4	3	0	1	0	Genoscope	-
Kluyveromyces waltii	5	5	3	2	0	0	BI	(Dujon *et al.*, 2004)
Lodderomyces elongisporus	8	9	6	2	1	0	BI	-
Saccharomyces cerevisiae 288C	10	10	4	3	3	0	SGD	(Goffeau *et al.*, 1996)
Saccharomyces cerevisiae RM11a	7	7	4	1	2	0	BI	-
Saccharomyces cerevisiae YJM789	7	7	4	1	2	0	SI	-
Saccharomyces bayanus	10	10	3	3	4	0	BI	(Kellis *et al.*, 2003)
Saccharomyces castellii	5	5	4	1	0	0	WUGSC	(Cliften *et al.*, 2003)
Saccharomyces kudriavzevii	8	8	2	3	3	0	WUGSC	-
Saccharomyces kluyveri	2	2	1	1	0	0	WUGSC	(Cliften *et al.*, 2003)
Saccharomyces mikatae	10	10	4	2	4	0	BI	(Kellis *et al.*, 2003)
Saccharomyces paradoxus	11	11	4	3	4	0	BI	(Kellis *et al.*, 2003)
Pichia stipitis	9	10	6	3	1	0	JGI	-
Yarrowia lipolytica	9	10	5	2	1	2	CBS	(Dujon *et al.*, 2004)
Taphrinomycotina (Subphylum)								
Schizosacharomyces pombe	4	6	4	1	1	0	NCBI	(Wood *et al.*, 2002)
Pezizomycotina (Subphylum)								
Botrytis cinerea	7	8	4	3	1	0	BI	-
Sclerotinia sclerotiorum	6	7	4	2	1	0	BI	-
Aspergillus fumigatus	6	7	4	3	0	0	TIGR	(Nierman *et al.*, 2005)
Aspergillus nidulans	6	7	4	3	0	0	BI	(Galagan *et al.*, 2005)
Aspergillus oryzae	6	7	4	3	0	0	DOGAN	(Machida *et al.*, 2005)
Aspergillus terreus	5	6	3	3	0	0	BI	-
Aspergillus niger	6	7	4	3	0	0	JGI	-
Coccidioides immitis	4	5	4	1	0	0	BI	-
Histoplasma capsulatum	5	6	3	3	0	0	BI	-
Uncinocarpus reesii	3	4	4	0	0	0	BI	-
Chaetomium globosum	4	4	2	1	1	0	BI	-
Fusarium graminearum	9	10	5	4	1	0	BI	-
Fusarium verticillioides	6	7	4	2	1	0	BI	-
Fusarium solani	7	8	5	2	1	0	JGI	-
Magnaporthe grisea	8	9	4	4	1	0	BI	(Dean *et al.*, 2005)
Neurospora crassa	7	8	4	3	1	0	BI	(Borkovich *et al.*, 2004)
Podospora anserina	8	9	4	4	1	0	IGM	-
Trichoderma reesei	7	8	4	3	1	0	JGI	-
Stagonospora nodorum	7	8	4	3	1	0	BI	-
Basidiomycota (Phylum)								
Ustilaginomycotina (Subphylum)								
Ustilago maydis 521	10	12	5	6	1	0	BI	(Kamper *et al.*, 2006)
Agricomycotina (Subphylum)								
Phanerochaete chrysosporium	6	6	2	2	2	0	JGI	(Martinez *et al.*, 2004)
Coprinus cinereus	9	10	4	3	3	0	BI	-
Laccaria bicolor	11	13	6	5	2	0	JGI	-
Cryptococcus neoformans Serotype A	12	12	2	5	3	2	BI	-
Cryptococcus neoformans Serotype B	10	10	2	4	3	1	NCBI	-
Cryptococcus neoformans Serotype D	11	11	2	4	4	1	SGTC	(Loftus *et al.*, 2005)
Zygomycota (Phylum)								
Rhizopus oryzae	37	48	31	14	1	2	BI	-
Chromista (Kingdom)								
Oomycota (Phylum)								
Phytophthora sojae	1	1	1	0	0	0	JGI	(Tyler *et al.*, 2006)
Phytophthora ramorum	0	0	0	0	0	0	JGI	(Tyler *et al.*, 2006)
Viridiplantae (Kingdom)								
Streptophyta (Phylum)								
Arabidopsis thaliana	32	32	0	27	5	0	TAIR	(AGI, 2000)
Oryza sativa japonica	26	30	0	22	7	1	IRGSP	(IRGSP, 2005)
Populus trichocarpa	38	38	0	28	10	0	JGI	(Tuskan *et al.*, 2006)
Metazoa (Kingdom)								
Nematoda (Phylum)								
Caenohabditis elegans	19	20	13	2	4	1	NCBI	(CSC, 1998)
Antropoda (Phylum)								
Drosophila melanogaster	10	14	12	0	2	0	BDGP	(Kornberg and Krasnow, 2000)
Coradata (Phylum)								
Mus musculus	16	23	27	0	0	0	Ensembl	(Waterston *et al.*, 2002)
Homo sapiens	18	27	22	1	0	0	Ensembl	(Lander *et al.*, 2001)
Total	556	627	308	219	90	10		

[a]IVe indicates motifs with less than 17 or more than 21 amino acid residues at the loop.
[b]SGTC, Stanford Genome Technology Center; SI, Sanger's Institute; CBS, Center For Biological Sequences; BI, Broad Institute; WGSC, WashU Genome Sequencing Center; JGI, DOE Joint Genomic Institute; DOGAN, Database Of the Genome; IGM, Analyze Instituté de Génétique et Microbiologie; TAIR, The Arabidopsis Information Resource; IRGSP, International Rice Genome Sequencing Project; BDGP, Berkeley Drosophila Genome Project.
[c]AGI, *Arabidopsis* Genome Initiative; CSC, C. *elegans* Sequencing Consortium.

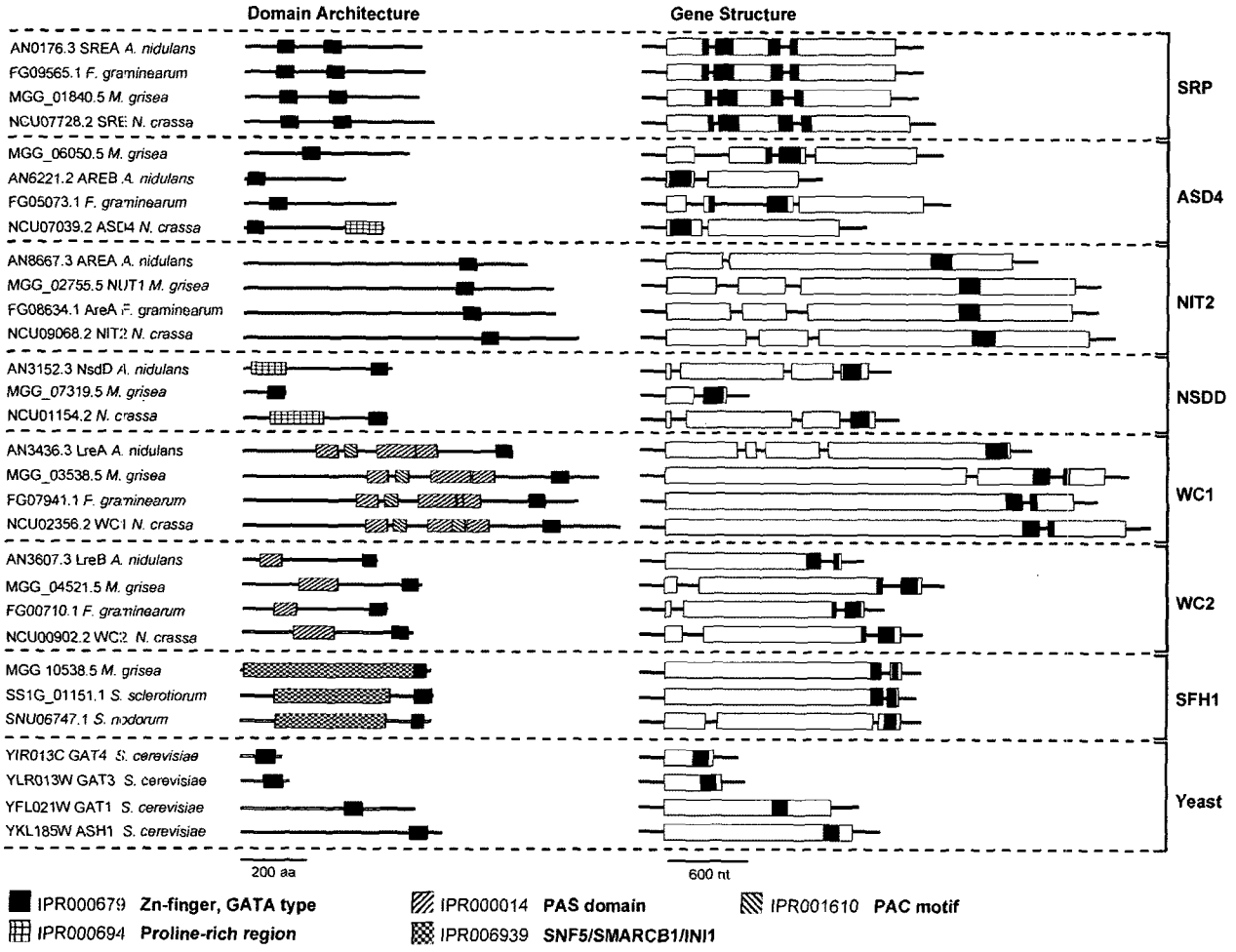

Fig. 4. Domain architecture and gene structure of selected GATA factors representing seven phylogenetic subgroups in subphylum Pezizomycotina and yeast. The domain architecture and gene structure of 27 GATA transcription factors in subphylum Pezizomycotina are shown with those of *S. cerevisiae*. Locations of GATA-type zinc finger domains, PAS domains, PAC motifs, and/or proline-rich regions are illustratedwith different patterns. In the gene structure section, an open box represents an exon and a black box indicates a GATA motif.

two subgroups comprising SRD1 and ECM23, and GAT3-and GAT4-like GATA TFs were found only in *Saccharomyces* species, while a subgroup of GATA TFs (tentatively named Y1), including orf19.1150 of *C. albicans* and CTRG03475.3 of *C. tropicalis*, appeared to be present only in *Candida* species. As described below in detail, the biological functions of Srd1p, Ecm23p, Gat3p, and Gat4p have been defined in *S. cerevisiae*, but those of the Y1 *Candida*-specific GATA TFs have not been characterized to date. *Schizosaccharomyces pombe* of the subphylum Taphrinomycotina contained only four GATA TFs. They barely clustered with other fungal GATA TFs, suggesting their distant phylogenetic relationship within the fungi.

GATA TFs of Pezizomycotina

The GATA TFs of 19 species in the subphylum Pezizomycotina (filamentous ascomycetes) encompassing four classes were analyzed further. These included two species (*Botrytis cinerea* and *Sclerotinia sclerotiorum*) belonging to the class Leotiomycetes; five species of *Aspergillus*, *Coccidioides immitis*, *Histoplasma capsulatum*, and *Uncinocarpus reesii* belonging to the class Eurotiomycetes; three species of *Fusarium*, *Chaetomium globosum*, *M. grisea*, *N. crassa*, *Podospora anserina*, and *Trichoderma reesei* belonging to the class Sordariomycetes; and *Stagnospora nodorum* belonging to the class Dothideomycetes (Table 1). Two representative species (*S. cerevisiae* and

C. albicans) in the subphylum Saccharomycotina (yeast form ascomycetes) were also included for comparison (Fig. 3). A total of 117 GATA TFs was retrieved from 19 species, with an average of 6 GATA TFs per species, ranging from 3 to 9. The distribution of GATA TFs in Sordariomycetes and Eurotiomycetes was different, where an average of 7 (ranging from 4 to 9) and 5 (ranging from 3 to 6) GATA TFs were identified, respectively. Seven phylogenetic subgroups were identified from a phylogenetic tree with full-length sequences and named SRP, NIT2, ASD4, WC1, WC2, NSDD, and SFH1. A similar tree structure was obtained when the domain sequences were used. Most GATA TFs in Pezizomycotina were distributed evenly to the first six subgroups, enabling prediction of their functions in other species. Two clades were found within each subgroup, reflecting already constructed taxonomic structures at the class level. These data suggest that these clades of GATA TFs emerged before divergence into classes, and then evolved in different ways. The SFH1 subgroup with 11 GATA TFs was newly identified in this comprehensive analysis and named after Sfh1p, a new Snf5p paralog (Cao *et al.*, 1997). GATA TFs in this subgroup contained a SNF5/SMARCB1/INI1 domain constituting the Swi/Snf family of chromatin-remodeling complex in *S. cerevisiae* (Klochendler-Yeivin and Yaniv, 2001). The biological functions of SFH1 GATA factors in filamentous ascomycetes have not yet been characterized.

The GATA TFs in each subgroup were highly conserved not only in amino acid sequence but also in domain arrangement and gene structure (Fig. 4). Most subgroups had one zinc finger motif, except the SRP subgroup with two type IVa zinc finger motifs. The location of the zinc finger motif was conserved within each subgroup. Zinc finger motifs were mostly located near the C-terminus, with those of SRP slightly biased to the N-terminal and those of ASD4 varying among species. Four subgroups had an additional domain other than a GATA-type zinc finger, a proline-rich region in NsdD, three PAS domains in WC-1, one PAS domain in WC-2, and SNF5/SMARCB1/INI1 in Sfh1p. The ASD4 group had a leucine zipper, implicated in protein–protein interactions not yet compiled into the InterPro database. In contrast, additional domains were rarely identified in GATA TFs from Saccharomycotina. Additional domains with functions other than DNA binding were also identified in plant GATA TFs (Reyes *et al.*, 2004). These data seem to support modular evolution of different groups of GATA TFs that is, GATA zinc finger domains might be integrated in different templates with other domains (Riechmann *et al.*, 2000). GATA TFs in the same subgroup also shared similar gene structure, which was clearly demonstrated in the SRP subgroup. Three exons and two introns comprised this subgroup of GATA TFs in each species, in which two zinc finger motifs spanning two adjacent exons interfered with each intron.

SRP

Seventeen GATA TFs were grouped with the SRE of *N. crassa*, suggesting a role in siderophore biosynthesis and regulation, while *C. globosum* did not have this subgroup of GATA TFs. SreA acts as a repressor of siderophore biosynthesis (Eisendle *et al.*, 2003) and SRE is known to act as a negative regulator of iron transport. SidA (L-orhithine-N^5-oxygenase) and SidC (peptide synthetase), the other two enzymes implicated in siderophore biosynthesis in *A. nidulans* (Haas, 2003), were also well distributed among the fungal species (data not shown).

NIT2 and ASD4

Four GATA TFs are known to regulate nitrogen metabolism in *S. cerevisiae*, two of which act positively (Gln3p and Gat1p), and the other two negatively (Dal80p and Deh1p; Cunningham *et al.*, 2000; Oliveira *et al.*, 2003). Homologs of these two subgroups were identified and extensively characterized in *N. crassa* and *A. nidulans*. NIT2 and AREA, homologs of Gln3p, functioned as major nitrogen regulatory proteins in *N. crassa* and *A. nidulans*, respectively (Tao and Marzluf, 1999; Morozov *et al.*, 2001). NIT2-like GATA factors were present in 18 species, except *H.capsulatum*. NIT2 is known to interact with the Zn(II)Cys6 binuclear transcription factor NIT4 (NCU08294.2) to regulate expression of *nit-3*, encoding nitrate reductase (Feng and Marzluf, 1998; Mo and Marzluf, 2003). NIT4 homologs were also identified in 19 fungal species, indicating that the regulatory machinery of nitrogen metabolism may be conserved in Pezizomycotina (data not shown).

The involvement of Dal80p homologs in nitrogen metabolism is controversial. NREB of *Penicillium chrysogenum*was reported to be involved in nitrogen metabolism, as was NRE, a Gln3p homolog (Haas *et al.*, 1997). AREB also partially complemented the *areA* phenotype in *A. nidulans* (Conlon *et al.*, 2001), but ASD4 in *N. crassa* was not implicated in this cellular process (Feng *et al.*, 2000). However, ASD4-like GATA factors were conserved among the Pezizomycotina. Only *Aspergillus terreus* did not have this subgroup of GATA TFs. All ASD4 orthologs, except that in *H. capsulatum*, were predicted to contain the conserved leucine zipper motif, which may be involved in protein–protein interactions, leaving the possibility of homo- or heterodimer formation for functioning. Homo- or heterodimerization of Dal80p and Deh1p of *S. cerevisiae* has been reported by

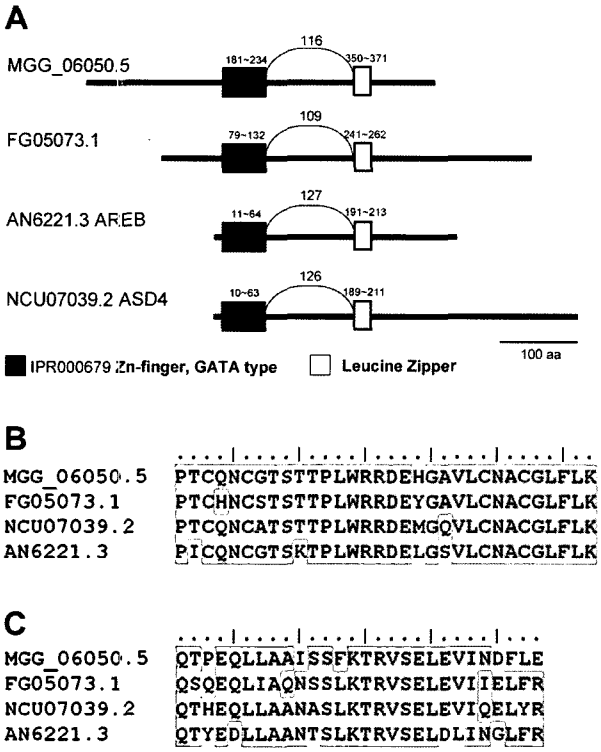

Fig. 5. Domain architecture of GATA transcription factors in the ASD4 subgroup. (A) Domain architecture of four representative proteins in the ASD4 subgroup is shown with GATA motifs (black boxes) and leucine zippers (white boxes). Numbers on each box indicate the sizeof each motif and spacing is shown between two motifs. (B) and (C) Multiple sequence alignment of GATA motifs (B) and leucine zippers (C) in the ASD4 subgroup are shown. Shaded amino acids indicate they are identical.

two-hybrid analysis (Svetlov and Cooper, 1998). The GATA and leucine zipper motifs consisted of 53 and 21 amino acids (aa), respectively, and were arranged in a conserved manner with similar spacing (109–126 aa) among species, irrespective of their location (Fig. 5A). The amino acid sequences of zinc finger and leucine zipper regions were also highly conserved among species (Fig. 5B and 5C).

WC1 and WC2

Fifteen and 16 GATA TFs were grouped together with WHITE COLLAR-1 (WC-1) and WC-2 of *N. crassa*, respectively. *C. globosum* contained a WC-1 ortholog and a WC-2-like protein missing the GATA motif. *C. immitis* and *U. reesii* possessed neither WC-1 nor WC-2 orthologs, while *P. anserina* appeared to lack WC-2. WC-1 and WC-2 are essential components for light responses, including entrainment of the circadian clock in *N. crassa*, which are mediated by the Per–Arnt–Sim (PAS) domains (Linden *et al.*, 1997; Lee *et al.*, 2000). WC-1 contains three PAS domains and WC-2 has one. The number and structure of domains were conserved among the WC-1 and WC-2 orthologs identified, except those in *Aspergillus oryzae* (Fig. 4). The most extreme N-terminal PAS domain of WC-1 belongs to a specialized class known as LOV (light, oxygen, or voltage) domains and acts as blue-light photoreceptor (reviewed in Cheng *et al.*, 2003; Crosson *et al.*, 2003). Eleven flavin-contacting residues found in the LOV domain of phototropins, the plant blue-light photoreceptors for phototrophism, were also well conserved among WC-1 homologs from 14 species. Recently, MgWC-1 was also reported to act as blue-light receptor in *M. grisea*, regulating asexual spore development and release (Lee *et al.*, 2006). The existence of a circadian clock in *Aspergillus* was reported, along with the homologs LreA and LreB (Greene *et al.*, 2003). Overall sequence similarity was higher among members within the same class, i.e., Eurotiomycetes in which *Aspergillus* resides and Sordariomycetes in which *Neurospora* and *Magnaporthe* reside (data not shown). In addition, FRQ orthologs serving as circadian oscillators in *N. crassa* (Loros *et al.*, 1986 Lee *et al.*, 2000) were only found in nine Sordariomycetes species. These results suggest that fungal species within Sordariomycetes may be subject to a similar light regulation as that in *N. crassa*, while supporting previous reports that the *Aspergillus* circadian clock has different properties than the well described *N. crassa* clock (Greene *et al.*, 2003).

NSDD

Fifteen GATA TFs from 14 species were grouped with NsdD of *A. nidulans*, which is involved in activating sexual development (Han *et al.*, 2001). NsdD homologs were not identified in five species: *C. globosum*, *Fusarium graminearum*, *Fusarium solani*, *S. sclerotiorum*, and *U. reesii*. The NsdD of *A. nidulans* consisted of 461 amino acids with one type IVb GATA zinc finger domain near the C-terminus, and was rich in proline (11.3%) and serine (13.4%). GATA TFs in this subgroup shared a similar primary structure, except those in *B. cinerea* (BC1G_10441.1) and *M. grisea* (MGG_07319.5). These two GATA factors consisted of about 130 amino acids, suggesting the possibility of an incorrect gene model. However, the GATA TFs retrieved from the same class were more closely related, in spite of the length variation. This subgroup of GATA TFs was distantly related to GAT2 (YMR136W) of *S. cerevisiae*, whose function remains unknown.

SFH1 as a novel subgroup in filamentous fungi

Several GATA TFs comprised a new highly conserved cluster. Eleven members belong to this subgroup. All of them have an evolutionarily highly conserved SNF5/SMARCB1/INI1 domain (IPR006939), a component of the yeast

SWI/SNF complex (Klochendler-Yeivin and Yaniv, 2001). Because the SNF5 domain contained most of the amino acids of these GATA TFs, we searched for proteins with IPR006939 in the fungal genomes. Of 37 genes with an SNF5 domain from filamentous ascomycetes, 11 had a GATA type zinc finger motif near the C-terminus after subsequent domain analysis, and these resided in the same clade with SFH1 of *S. cerevisiae*, one of two yeast homologs of human SNF5/INI1. SFH1 consists of the chromatin remodeling RSC complex, which is known to interact with kinetochore components involved in chromosome segregation during the G2–M transition and to play a role in reconfiguring cetromeric and flanking nucleosomes (Cao *et al.*, 1997). Interestingly, a GATA motif along with the adjacent highly basic region was not found in SFH1 homologs from Eurotiomycetes or in SFH1 itself. These results again support the modular evolution of GATA domains during evolution, conferring functions in addition to chromatin remodeling. The spacing between two cysteine pairs at zinc finger loops in all members was 20 amino acid residues, forming a different type of zinc finger motif (type IVc).

GATA factors of Saccharomycotina

A total of 173 putative GATA TFs were identified from 23 yeast genome sequences, including nine *Saccharomyces* and six *Candida* genomes (Table 1). Each species of *Saccharomyces* possesses ten GATA TFs (Table 1) and most have been well defined in *S. cerevisiae*. *Saccharomyces kluyveri* and two *S. cerevisiae* strains, RM11a and YJM789, contained only two and seven annotated GATA TFs in their genomes, respectively (Table 1). However, this outcome was likely due to the limitation of automated gene prediction algorithms rather than to strain divergence. When explored using BLAST searches, for example, the genomes of the two *S. cerevisiae* variants contained unannotated GATA TFs whose protein sequences were almost identical to the counterparts in the *S. cerevisiae* 288C genome.

Our phylogenetic analysis indicated that 70 *Saccharomyces* GATA TFs were categorized into six subgroups: NIT2 (Gln3p and Gat1p of *S. cerevisiae*), ASD4 (Gzf3p and Dal80p of *S. cerevisiae*), NSDD (Gat2p of *S. cerevisiae*), ASH1 (Ash1p of *S. cerevisiae*), SRD1/ECM23 (Srd1p and Ecm23p of *S. cerevisiae*), and GAT3/GAT4 (Gat3p and Gat4p of *S. cerevisiae*). While the first three subgroups and the ASH1 subgroup were shared by filamentous fungal and *Candida* species, respectively, the SRD1/ECM23 and GAT3/GAT4 subgroups were found only in *Saccharomyces* species. Although their precise biological functions remain undefined, Srd1p of *S. cerevisiae* is related to pre-rRNA processing (Fabian *et al.*, 1990), and Ecm23p, previously known as Srd2p, is associated with pseudohyphal growth (Canizares *et al.*, 2002). Gat3p of *S. cerevisiae* in the GAT3 subgroup is responsible for the selective use of good over poor nitrogen sources (Banerjee and Zhang, 2003), and the role of Gat4p needs to be characterized. Interestingly, the SRP-type GATA TF, well conserved throughout the fungi, was lacking in this genus and the closely related *Kluyveromyces*.

Candida albicans has a relatively large number of GATA TFs compared to other *Candida* species (Table 1). Forty-nine GATA TFs from six *Candia* species were grouped into six subgroups: the *Candida*-specific Y1, SRP, NIT2, ASD4, NSDD, and ASH1 groups. Although ASH1 GATA TFs are conserved in *Saccharomyces* and *Candida* species, they have evolved different functions. Even with the sequence similarity, Ash1p of *S. cerevisiae* represses *HO* expression in the daughter cell responsible for mating-type switching (Cosma, 2004), but CaASH1 of *C. albicans* controls filamentous growth and virulence (Inglis and Johnson, 2002). The Y1 *Candida*-specific GATA TFs have not been characterized.

In addition, except those of *S. pombe*, which belongs to Taphrinomycotina, GATA TFs from other saprophytic yeasts have a tendency to assemble with *Saccharomyces* GATA TFs. In *S. pombe*, four GATA TFs were identified, among which only Fep1, which encodes an iron-sensing transcription factor (Pelletier *et al.*, 2003), was distantly related to SFU1, the SRP of *C. albicans* (Lan*et al.*, 2004).

GATA motif analysis

GATA motifs are typically classified into two groups based on the spacing between two cysteine pairs at the zinc finger loop: type IVa ($CX_2CX_{17}CX_2C$) and type IVb ($CX_2CX_{18}CX_2C$). Fungal species contain both types of zinc finger loops, while animals have only type IVa and plants have exclusively type IVb (Teakle and Gilmartin, 1998). To decipher the distribution and relationship of GATA motifs in fungi, 442 motifs from 396 GATA TFs were retrieved from 50 fungal genomes. In addition, 184 motifs from 159 GATA TFs were also retrieved from four animal and three plant genomes for comparative analysis. Forty-six fungal GATA TFs, including siderophore regulatory proteins in *A. nidulans* (SREA) and *N. crassa* (SRE), had two motifs. Of the fungal GATA TFs, 233 (52.7%) were type IVa and 139 (31.4%) were type IVb.

When type IVa GATA motifs were phylogenetically analyzed, three clear clades were identified: one each specific to fungi or animals, and one for both fungal and animal GATA TFs (Fig. 6A). NIT2, ASD4, and SRP C-terminal GATA motifs were grouped in the fungal-

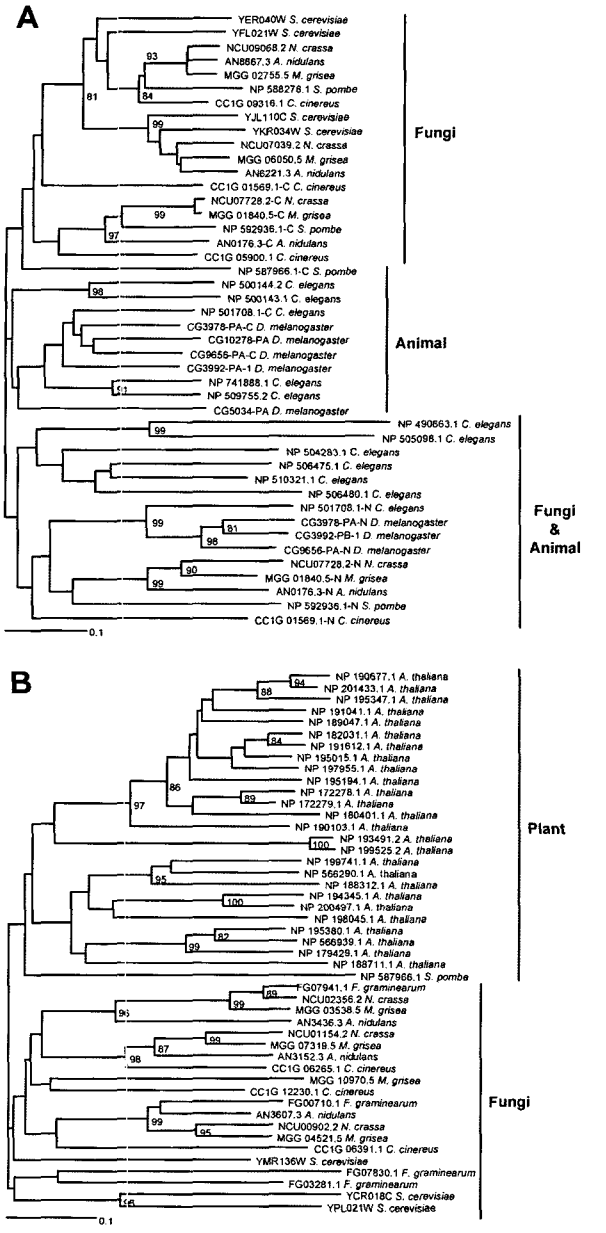

Fig. 6. Phylogenetic analysis of GATA motifs, types IVa and IVb. (A) Phylogenetic tree of 44 type IVa GATA motifs. (B) Phylogenetic tree of 47 type IVb GATA motifs. Names of GATA motifs were derived from gene names and the direction (i.e., -N (N-terminal) and -C (C-terminal)) was added only if the ORF contained two GATA motifs. Phylogenetic trees were drawn by the neighbor-joining method with 10,000 bootstrap replicates. Bootstrap values over 80% are shown.

specific clade. However, SRP N-terminal GATA motifs were grouped in the fungal and animal clade. It is interesting to note that each GATA motif at different positions in the SRP subgroup was placed in a different clade, suggesting that they might have evolved separately after a duplication event (Caddick and Arst, 1990).

GATA motifs of proteins belonging to the WC1, WC2, and NSDD subgroups were identified as type IVb and grouped separately from plant type IVb (Fig. 6B). NP_587966.1 of *S. pombe* had the type IVb motif, but was not grouped in the fungal clade. It remains to be determined whether the type IVb motif in Taphrinomycotina is independently placed and/or is the bridge between fungal and plant clades.

New GATA motifs of type IVc

In addition to type IVa and IVb, several GATA TFs with 19- or 20-residue motifs have been found in plants (Reyes *et al.*, 2004) and fungi belonging to the subphylum Saccharomycotina (Munchow *et al.*, 2002). ASH1 in *S. cerevisiae* has a 20-residue motif and its DNA-binding ability was shown by a DNA footprinting experiment (Maxon and Herskowitz, 2001). CaASH1 in *C. albicans* has a 19-residue motif that also binds DNA (Munchow *et al.*, 2002).

Fifty-six GATA motifs from 50 fungal genomes were found to have 19- or 20-residues. In addition, six motifs having 21-residues were also identified in fungal GATA TFs, but not in plants and animals. Based on our findings, the 19- to 21-residue type GATA motif, $CX_{2-4}CX_{19-21}CX_2C$, was collectively named type IVc. Type IVc motifs were also found in *Drosophila* and *Caenorhabditis elegans*, but not in humans or mice (Table 1). In fungi, type IVc motifs were found in GATA TFs from the phyla Zygomycota, Ascomycota, and Basidiomycota. However, the frequency varied among fungal taxonomic groups. Basidiomycetes showed the highest proportion of type IVc GATA motifs (24.3%), followed by Sacchromycotina (17.3%) and Pezizomycotina (7.5%).

When type IVc GATA motifs were phylogenetically analyzed, fungal type IVc motifs were grouped apart from plant GATA factors (bootstrap value, 99.45; Fig. 7). GATA motifs of subphylum Saccharomycotina had 19 and 20 residues. Multiple sequence alignment of these motifs suggests the gain or loss of one amino acid at position 14 between the two cysteine pairs in the zinc finger loop. In Basidiomycota, six unique 21-residue GATA motifs were found in three strains of *C. neoformans*, and these grouped together with 20-residue motifs in the same clade. The 19- and 20-residue motifs of yeast also grouped together and have been characterized as transcription factors with DNA-binding ability (Maxon and Herskowitz, 2001; Munchow *et al.*, 2002). These results strongly support the hypothesis that 21-residue type IVc motifs found in this study could be classified as transcription factors.

In Pezizomycotina, all GATA TFs with type IVc motifs had an additional domain of SNF5/SMARCB1/INI1. With

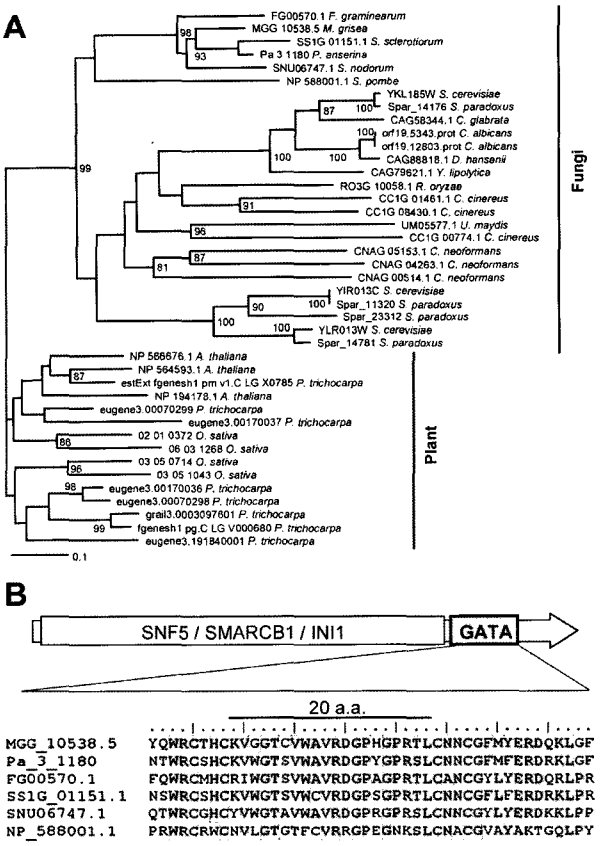

Fig. 7. Phylogenetic analysis of GATA motif type IVc. (A) Phylogenetic tree of 47 type IVc GATA motifs. Tree was drawn by the neighbor-joining method with 10,000 bootstrap replicates. Bootstrap values over 80% are shown. (B) Domain architecture and amino acid alignment of GATA motifs from subphylum Pezizomycotina. Multiple sequence alignment of GATA motifs belonging to the SFH1 group are displayed with the 20-amino-acid-residue loop.

highly conserved domain architecture, sequences of GATA motifs were also well conserved (Fig. 7B). Except for NP_588001.1 in *S. pombe*, 75% of the loop sequences were highly conserved among the 11 GATA motifs. *Schizosaccharomyces japonica,* for which annotation data are not publicly available, has similar GATA factors, with the SNF5/SMARCB1/INI1 domain showing 66.9% amino acid identity. These facts strongly suggest that Pezizomycotina-specific type IVc motifs share a common ancestor with Taphrinomycotina.

With an exponential increase in the number of fully sequenced genomes, determining the biological functions of hypothetical proteins from genomic sequences has become a central goal of bioinformatics. Based on the assumption that proteins that function in a similar pathway are likely to evolve in a correlated fashion, our phylogenetic profiling and comparative analysis of 396 putativeGATA factors from 50 complete fungal genome sequences strongly support comparative phylogenomics as an alternative strategy for *in silico* functional prediction of uncharacterized genes, using evolutionary analysis at the genome-wide level. Thus, we have provided not only an inventory of fungal GATA transcription factors, but also a platform for comparative genomic analyses of other transcription factor families.

Acknowledgements

This study was supported by grants from Microbial Genomics and Applications Center (0462-20060021) and Crop Functional Genomics Center (CG1421) of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology, by a grant from Biogreen21 project funded by Rural Development Administration, and by Korean Research Foundation Grant (KRF-2004-005-F00013) to Y.-H. Lee.

References

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796-815.

Banerjee, N. and Zhang, M. Q. (2003). Identifying cooperativity among transcription factors controlling the cell cycle in yeast. *Nucleic Acids Res.* 31, 7024-7031.

Borkovich, K. A., Alex, L. A., Yarden, O., Freitag, M., Turner, G. E. *et al.* (2004). Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol. Mol. Biol. Rev.* 68, 1-108, table of contents.

Caddick, M. X., Arst, H. N Jr. (1990). Nitrogen regulation in *Aspergillus*: are two fingers better than one? *Gene* 95, 123-127.

Canizares, J. V., Pallotti, C., Sainz-Pardo, I., Iranzo, M., and Mormeneo, S. (2002). The SRD2 gene is involved in *Saccharomyces cerevisiae* morphogenesis. *Arch. Microbiol.* 177, 352-357.

Cao, Y., Cairns, B. R., Kornberg, R. D., and Laurent, B. C. (1997). Sfh1p, a component of a novel chromatin-remodeling complex, is required for cell cycle progression. *Mol. Cell Biol.* 17, 3323-3334.

Cavailer-Smith, T. (1987). The origin of fungi and pseudofungi. Cambridge Univ. Press.

Cheng, P., He, Q., Yang, Y., Wang, L.,and Liu, Y. (2003). Functional conservation of light, oxygen, or voltage domains in light sensing. *Proc. Natl. Acad. Sci. USA.* 100, 5938-5943.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B. *et*

al. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science* 301, 71-76.

Conlon, H., Zadra, I., Haas, H., Arst, H. N., Jr., Jones, M. G. *et al.* (2001). The *Aspergillus nidulans* GATA transcription factor gene areB encodes at least three proteins and features three classes of mutation. *Mol. Microbiol.* 40, 361-375.

Cosma, M. P. (2004). Daughter-specific repression of *Saccharomyces cerevisiae* HO: Ash1 is the commander. *EMBO Rep.* 5, 953-957.

Crosson, S., Rajagopal, S. and Moffat, K. (2003). The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. *Biochemistry* 42, 2-10.

C. elegans Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012-2018.

Cunningham, T. S., Rai, R., and Cooper, T. G. (2000). The level of DAL80 expression down-regulates GATA factor-mediated transcription in *Saccharomyces cerevisiae*. *J. Bacteriol.*182, 6584-6591.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K. *et al.* (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434, 980-986.

Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K. *et al.* (2004). The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304, 304-307.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S. *et al.* (2004). Genome evolution in yeasts. *Nature* 430, 35-44.

Eisendle, M., Oberegger, H., Zadra, I., and Haas, H. (2003). The siderophore system is essential for viability of *Aspergillus nidulans*: functional analysis of two genes encoding l-ornithine N 5-monooxygenase (sidA) and a non-ribosomal peptide synthetase (sidC). *Mol. Microbiol.* 49, 359-375.

Fabian, G. R., Hess, S. M., and Hopper, A. K. (1990). *srd1*, a *Saccharomyces cerevisiae* suppressor of the temperature-sensitive pre-rRNA processing defect of *rrp1-1*. *Genetics* 124, 497-504.

Feng, B., Haas, H., and Marzluf, G. A. (2000). ASD4, a new GATA factor of *Neurospora crassa*, displays sequence-specific DNA binding and functions in ascus and ascospore development. *Biochemistry* 39, 11065-11073.

Feng, B. and Marzluf, G. A. (1998). Interaction between major nitrogen regulatory protein NIT2 and pathway-specific regulatory factor NIT4 is required for their synergistic activation of gene expression in *Neurospora crassa*. *Mol. Cell Biol.* 18, 3983-3990.

Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R. *et al.* (2005). Sequencing of *Aspergillus nidulans*and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438, 1105-1115.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B. *et al.* (1996). Life with 6000 genes. *Science* 274, 546, 563-547.

Greene, A. V., Keller, N., Haas, H., and Bell-Pedersen, D. (2003). A circadian oscillator in *Aspergillus spp.* regulates daily development and gene expression. *Eukaryot. Cell* 2, 231-237.

Haas, H. (2003). Molecular genetics of fungal siderophore biosynthesis and uptake: the role of siderophores in iron uptake and storage. *Appl. Microbiol. Biotechnol.* 62, 316-330.

Haas, H., Angermayr, K., Zadra, I., and Stoffler, G. (1997). Overexpression of nreB, a new GATA factor-encoding gene of *Penicillium chrysogenum*, leads to repression of the nitrate assimilatory gene cluster. *J. Biol. Chem.* 272, 22576-22582.

Haas, H., Zadra, I., Stoffler, G., and Angermayr, K. (1999). The *Aspergillus nidulans*GATA factor SREA is involved in regulation of siderophore biosynthesis and control of iron uptake. *J. Biol. Chem.* 274, 4613-4619.

Han, K. H., Han, K. Y., Yu, J. H., Chae, K. S., Jahng, K. Y. *et al.* (2001). The nsdD gene encodes a putative GATA-type transcription factor necessary for sexual development of *Aspergillus nidulans*. *Mol. Microbiol.* 41, 299-309.

Hayama, R. and Coupland, G. (2003). Shedding light on the circadian clock and the photoperiodic control of flowering. *Curr. Opin. Plant Biol.* 6, 13-19.

Hefti, M. H., Francoijs, K. J., de Vries, S. C., Dixon, R., and Vervoort, J. (2004). The PAS fold. A redefinition of the PAS domain based upon structural prediction. *Eur. J. Biochem.* 271, 1198-1208.

Horowitz, N. H., Charlang, G., Horn, G., and Williams, N. P. (1976). Isolation and identification of the conidial germination factor of *Neurospora crassa*. *J. Bacteriol.*127, 135-140.

Inglis, D. O. and Johnson, A. D. (2002). Ash1 protein, an asymmetrically localized transcriptional regulator, controls filamentous growth and virulence of *Candida albicans*. *Mol. Cell Biol.* 22, 8669-8680.

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature* 436, 793-800.

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V. *et al.* (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443, 818-822.

Jones, T., Federspiel, N. A., Chibana, H., Dungan, J.,

Kalman, S. *et al.* (2004). The diploid genome sequence of *Candida albicans*. *Proc. Natl. Acad. Sci. USA.* 101, 7329-7334.

Kamper, J., Kahmann, R., Bolker, M., Ma, L. J., Brefort, T. *et al.* (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444, 97-101.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241-254.

Klochendler-Yeivin, A. and Yaniv, M. (2001). Chromatin modifiers and tumor suppression. *Biochim. Biophys. Acta.* 1551, M1-10.

Kornberg, T. B. and Krasnow, M. A. (2000). The Drosophila genome sequence: implications for biology and medicine. *Science* 287, 2218-2220.

Kumar, S., Tamura, K., and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* 5, 150-163.

Lan, C. Y., Rodarte, G., Murillo, L. A., Jones, T., Davis, R. W. *et al.* (2004). Regulatory networks affected by iron availability in *Candida albicans*. *Mol. Microbiol.* 53, 1451-1469.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C. *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Lee, K., Loros, J. J., and Dunlap, J. C. (2000). Interconnected feedback loops in the Neurospora circadian system. *Science* 289, 107-110.

Lee, K., Singh, P., Chung, W. C., Ash, J., Kim, T. S. *et al.* (2006). Light regulation of asexual development in the rice blast fungus, *Magnaporthe oryzae*. *Fungal Genet. Biol.* 43, 694-706.

Linden, H., Ballario, P., and Macino, G. (1997). Blue light regulation in *Neurospora crassa*. *Fungal Genet. Biol.*22, 141-150.

Loftus, B. J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P. *et al.* (2005). The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307, 1321-1324.

Loros, J. J., Richman, A., and Feldman, J. F. (1986). A recessive circadian clock mutation at the frq locus of *Neurospora crassa*. *Genetics* 114, 1095-1110.

Lowry, J. A. and Atchley, W. R. (2000). Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J. Mol. Evol.* 50, 103-115.

Lu, Y. K., Sun, K. H., and Shen, W. C. (2005). Blue light negatively regulates the sexual filamentation via the Cwc1 and Cwc2 proteins in *Cryptococcus neoformans*. *Mol. Microbiol.* 56, 480-491.

Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T. *et al.* (2005). Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 438, 1157-1161.

Martinez, D., Larrondo, L. F., Putnam, N., Gelpke, M. D., Huang, K. *et al.*(2004). Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.* 22, 695-700.

Matzanke, B. F., Bill, E., Trautwein, A. X., and Winkelmann, G. (1987). Role of siderophores in iron storage in spores of *Neurospora crassa* and *Aspergillus ochraceus*. *J. Bacteriol.*169, 5873-5876.

Maxon, M. E. and Herskowitz, I. (2001). Ash1p is a site-specific DNA-binding protein that actively represses transcription. *Proc. Natl. Acad. Sci. USA.* 98, 1495-1500.

McGinnis, S. and Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 32, W20-25.

Mo, X. and Marzluf, G. A. (2003). Cooperative action of the NIT2 and NIT4 transcription factors upon gene expression in *Neurospora crassa*. *Curr. Genet.* 42, 260-267.

Morozov, I. Y., Galbis-Martinez, M., Jones, M. G., and Caddick, M. X. (2001). Characterization of nitrogen metabolite signalling in Aspergillus via the regulated degradation of areA mRNA. *Mol. Microbiol.* 42, 269-277.

Munchow, S., Ferring, D., Kahlina, K., and Jansen, R. P. (2002). Characterization of *Candida albicans* ASH1 in *Saccharomyces cerevisiae*. *Curr. Genet.* 41, 73-81.

Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S. *et al.* (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 438, 1151-1156.

Oliveira, E. M., Martins, A. S., Carvajal, E., and Bon, E. P. (2003). The role of the GATA factors Gln3p, Nil1p, Dal80p and the Ure2p on ASP3 regulation in *Saccharomyces cerevisiae*. *Yeast* 20, 31-37.

Pelletier, B., Beaudoin, J., Philpott, C. C., and Labbe, S. (2003). Fep1 represses expression of the fission yeast *Schizosaccharomyces pombe* siderophore-iron transport system. *Nucleic Acids Res.* 31, 4332-4344.

Reyes, J. C., Muro-Pastor, M. I., and Florencio, F. J. (2004). The GATA family of transcription factors in Arabidopsis and rice. *Plant Physiol.* 134, 1718-1732.

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. *et al.*(2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105-2110.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.

Scazzocchio, C. (2000). The fungal GATA factors. *Curr. Opin. Microbiol.* 3, 126-131.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003).

SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381-3385.

Sheppard, D. C., Doedt, T., Chiang, L. Y., Kim, H. S., Chen, D. *et al.* (2005). The *Aspergillus fumigatus*StuA protein governs the up-regulation of a discrete transcriptional program during the acquisition of developmental competence. *Mol. Biol. Cell* 16, 5866-5879.

Svetlov, V. V. and Cooper, T. G. (1998). The *Saccharomyces cerevisiae*GATA factors Dal80p and Deh1p can form homo- and heterodimeric complexes. *J. Bacteriol.* 180, 5682-5688.

Tao, Y. and Marzluf, G. A. (1999). The NIT2 nitrogen regulatory protein of Neurospora: expression and stability of nit-2 mRNA and protein. *Curr. Genet.* 36, 153-158.

Teakle, G. R. and Gilmartin, P. M. (1998). Two forms of type IV zinc-finger motif and their kingdom-specific distribution between the flora, fauna and fungi. *Trends Biochem. Sci.* 23, 100-102.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I. *et al.* (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596-1604.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. *et al.* (2006). Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313, 1261-1266.

Vitale, N., Moss, J., and Vaughan, M. (1998). Molecular characterization of the GTPase-activating domain of ADP-ribosylation factor domain protein 1 (ARD1). *J. Biol. Chem.* 273, 2553-2560.

Voisard, C., Wang, J., McEvoy, J. L., Xu, P., and Leong, S. A. (1993). urbs1, a gene regulating siderophore biosynthesis in *Ustilago maydis*, encodes a protein similar to the erythroid transcription factor GATA-1. *Mol. Cell Biol.* 13, 7091-7100.

Waterston, R. H.Lindblad-Toh, K.Birney, E.Rogers, J.Abril, J. F. *et al.*(2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520-562.

Wood, V.Gwilliam, R.Rajandream, M. A.Lyne, M.Lyne, R. *et al.* (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415, 871-880.

Zdobnov, E. M. and Apweiler, R. (2001). InterProScan-an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847-848.