

Comparative Evolutionary Histories of the Fungal Chitinase Gene Family Reveal Non-Random Size Expansions and Contractions due to Adaptive Natural Selection

Magnus Karlsson and Jan Stenlid

Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, P.O. 7026, SE-75007, Uppsala, Sweden.

Abstract: Gene duplication and loss play an important role in the evolution of novel functions and for shaping an organism's gene content. Recently, it was suggested that stress-related genes frequently are exposed to duplications and losses, while growth-related genes show selection against change in copy number. The fungal chitinase gene family constitutes an interesting case study of gene duplication and loss, as their biological roles include growth and development as well as more stress-responsive functions. We used genome sequence data to analyze the size of the chitinase gene family in different fungal taxa, which range from 1 in *Batrachochytrium dendrobatidis* and *Schizosaccharomyces pombe* to 20 in *Hypocrea jecorina* and *Emmericella nidulans*, and to infer their phylogenetic relationships. Novel chitinase subgroups are identified and their phylogenetic relationships with previously known chitinases are discussed. We also employ a stochastic birth and death model to show that the fungal chitinase gene family indeed evolves non-randomly, and we identify six fungal lineages where larger-than-expected expansions (*Pezizomycotina*, *H. jecorina*, *Gibberella zeae*, *Uncinocarpus reesii*, *E. nidulans* and *Rhizopus oryzae*), and two contractions (*Coccidioides immitis* and *S. pombe*) potentially indicate the action of adaptive natural selection. The results indicate that antagonistic fungal-fungal interactions are an important process for soil borne ascomycetes, but not for fungal species that are pathogenic in humans. Unicellular growth is correlated with a reduction of chitinase gene copy numbers which emphasizes the requirement of the combined action of several chitinases for filamentous growth.

Keywords: chitinases, gene family, fungi, evolution, phylogeny

Introduction

Chitin is a polymer which consists of N-acetylglucosamine monomers (GlcNAc), linked by β -1,4-glucosidic bonds. It is widely distributed in nature and it is a constituent of the exoskeleton of invertebrates, of zooplankton and of fungal cell walls. Chitinases (EC 3.2.1.14) hydrolyze the bonds between GlcNAc residues releasing oligomeric, dimeric (chitobiose) or monomeric (GlcNAc) products. Chitobiose can be further cleaved by N-acetylhexosaminidases (EC 3.2.1.52) into GlcNAc (Keyhani and Roseman, 1999). Chitinases are divided into two different glycoside hydrolase families (18 and 19) based on amino acid sequence similarity (Henrissat, 1991; Henrissat and Bairoch, 1993). These two families share limited similarity at the amino acid level and have different three-dimensional structures and modes of action (Iseli et al. 1996). These enzymes can display either exo- or endoactivity, depending on the structure of the catalytic site (Terwisscha van Scheltinga et al. 1994; van Aalten et al. 2000; van Aalten et al. 2001).

Growth and morphological development of fungi makes cell wall remodelling a necessity. Cell expansion and division, spore germination, hyphal branching and septum formation all depend on the activities of hydrolytic enzymes intimately associated with the fungal cell wall, among them chitinases (Adams, 2004). Chitinases are also implied in autolysis and recycling of older parts of the fungal mycelia (Duo-Chuan, 2006). Chitinases also have aggressive roles as fungal pathogenicity factors during infection of other fungi (mycoparasitism), insects and nematodes (Wattanalai et al. 2004; Duo-Chuan, 2006; Gan et al. 2007a). Furthermore, chitinases are involved in degradation of chitin for nutritional needs (Duo-Chuan, 2006). Lysis of the host cell wall and degradation of nematode egg shells are shown to be important steps in the mycoparasitic and nematophagous attack (Howell, 2003; Benitez et al. 2004; Gan et al. 2007a), and hence chitinases from various fungi used as biocontrol agents have been

Correspondence: Magnus Karlsson, Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, P.O. 7026, SE-75007, Uppsala, Sweden. Tel: +46 18 671875; Fax: +46 18 673599; Email: magnus.karlsson@mykopat.slu.se



Copyright in this article, its metadata, and any supplementary data is held by its author or authors. It is published under the Creative Commons Attribution By licence. For further information go to: <http://creativecommons.org/licenses/by/3.0/>.

cloned and characterised (Felse and Panda, 1999; Hoell et al. 2005; Klemsdal et al. 2006; Gan et al. 2007a; Gan et al. 2007b; Dong et al. 2007).

The diversity of chitinase function during the fungal life cycle raises interesting questions regarding the evolution of this important gene family. Fungal chitinases belong to glycoside hydrolase family 18 (GH18) and they consist of discrete domains, which are variously arranged in different orders in different proteins (Gilkes et al. 1991; Warren, 1996; Henrissat and Davies, 2000). Besides the catalytic domain there is very often a substrate-binding domain present. These substrate-binding domains are not necessary for chitinolytic activity, although they seem to enhance the efficiency of the enzymes (Suzuki et al. 1999; Limon et al. 2001).

There is a large variation in the number of GH18 genes present in different fungal genomes, from 1 in *Schizosaccharomyces pombe* to 20 in *Hypocrea jecorina* (Seidl et al. 2005). This implies that the size of the fungal GH18 gene family has been highly dynamic throughout evolution. Gene duplication is an important process that can contribute to the evolution of novel functions. However, the mechanisms that govern the fate of duplicated genes are not very well understood. Recent progress suggests that stress-related genes frequently are exposed to duplications and losses, while growth-related genes show selection against change in copy number (Wapinski et al. 2007). High copy-numbers of stress-responsive genes may be beneficial by allowing adaptations to diverse ecological niches. Recent paralogues diversify most frequently at the level of regulation, and more rarely at the level of biochemical function (Wapinski et al. 2007).

Identification of expansions as well as contractions of protein families in fungi with diverse ecological roles can aid in understanding relationships between function and phylogeny. The fungal GH18 gene family constitutes an interesting case study of gene duplication and loss, as their biological roles includes growth and development as well as more stress-responsive functions. Hence it is possible to test the hypothesis that growth-related genes display selection against changes in copy number while stress-related genes tolerate more duplications and losses, within a single gene family. Here we present a study where we use genome sequence data to analyze the size of the GH18 gene family in different fungal taxa and to infer their phylogenetic relationships. We also employ a stochastic birth and death model to test for non-random evolution of the

fungal GH18 gene family. We show that the fungal GH18 gene family indeed evolves non-randomly and we identify fungal lineages where larger-than-expected expansions or contractions potentially indicate the action of adaptive natural selection.

Materials and Methods

Biomining of genome sequences

In order to avoid sampling bias our study was restricted to fungal species where genome sequence information and estimates of divergence times were available. *In silico* translated gene products from individual fungal genome sequences were screened for the presence of GH18s using BLASTP (Altschul et al. 1997) in an iterative process. Fungal genomes were available at the homepages of the DOE Joint Genome Institute (<http://www.jgi.doe.gov/>), the Fungal Genome Initiative at the BROAD Institute (<http://www.broad.mit.edu/annotation/fgi/>) or at Génolevures (<http://cbi.labri.u-bordeaux.fr/Genolevures/blast/index.php>). The 18 published GH18s from *H. jecorina*, Chi18-1 through Chi18-18, and two additional proteins annotated in the genome, Chi18-rel1 (Chi18-20) and Chi18-rel2 (Chi18-19), protein ID 65162 and 121355 in version 2.0 were used as starting material, as this is so far the highest number of GH18s from a single fungal species (Seidl et al. 2005). Later on the number of proteins used was reduced to Chi18-1, Chi18-2, Chi18-12 and Chi18-19 from *H. jecorina* as these representatives provided the same information as the larger set. In addition, the GH18s that were identified by the first round of similarity searches in a target genome was iteratively used in a second round of BLAST searches against the same genome. The protein identifiers from the respective genome sequencing projects were used during subsequent analyses, except for when the protein was characterized and named.

Phylogenetic analysis

Amino acid sequences of GH18 catalytic domains were determined by InterProScan (Zdobnov and Apweiler, 2001) or Conserved Domain Database searches (Marchler-Bauer et al. 2005). Sequences were manually trimmed and aligned with Clustal X (Thompson et al. 1997) and inspected using BioEdit (Hall, 1999). Amino-acid similarity between sequences was calculated using MegAlign, implemented in the DNASTAR program package

(DNASTAR, Madison, WI). Phylogenetic analysis of catalytic domains was performed using maximum likelihood methods implemented in PhyML-aLRT 1.1 (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006). The JTT amino-acid substitution model (Jones et al. 1992) was used, the proportion of invariable sites was set to 0, one category of substitution rate was used and gaps were treated as unknown characters. The starting tree to be refined by the maximum likelihood algorithm was a distance-based BIONJ tree estimated by the program (Guindon and Gascuel, 2003). Statistical support for phylogenetic grouping was assessed by approximate likelihood-ratio tests based on a Shimodaira-Hasegawa-like procedure (SH-aLRT) (Anisimova and Gascuel, 2006) and by bootstrap analysis (500 resamplings).

Likelihood analysis of gene gain and loss

In order to statistically test whether the size of the fungal GH18 gene family is compatible with a

stochastic birth and death model we used the program CAFE (Computational Analysis of gene Family Evolution) (De Bie et al. 2006), which is based on the probabilistic framework developed by Hahn et al. (2005). From a specified phylogenetic tree and the gene family size in extant species, we inferred the most likely gene family size at internal nodes, tested for accelerated rates of gene family expansions or contractions and identified the branches that are responsible for the non-random evolution.

The fungal GH18 gene family data in extant species that were used in the current analysis are found in Table 1. Fungal GH18s can be divided into three major phylogenetic clusters, A, B and C, and further subdivisions within these clusters are made (Seidl et al. 2005). Our phylogenetic analysis shows that cluster C GH18s can be merged with cluster A and therefore we analysed the data in three ways; cluster A GH18s separately, cluster B GH18s separately and all GH18s merged. CAFE assumes that the gene family under study

Table 1. Number of chitinase genes in different fungal species.

Species	Class	Cluster A GH18 genes	Cluster B GH18 genes	Total no. of GH18 genes
<i>Batrachochytrium dendrobatidis</i>	Chytridiomycetes	1	0	1
<i>Rhizopus oryzae</i>	Mucormycotina	9	6	15
<i>Schizosaccharomyces pombe</i>	Schizosaccharomycetes	0	1	1
<i>Yarrowia lipolytica</i>	Saccharomycetes	1	2	3
<i>Candida albicans</i>	Saccharomycetes	2	3	5
<i>Saccharomyces cerevisiae</i>	Saccharomycetes	1	1	2
<i>Emericella nidulans</i>	Eurotiomycetes	17	3	20
<i>Ajellomyces capsulatus</i>	Eurotiomycetes	7	2	9
<i>Uncinocarpus reesii</i>	Eurotiomycetes	11	3	14
<i>Coccidioides immitis</i>	Eurotiomycetes	6	3	9
<i>Gibberella zeae</i>	Sordariomycetes	16	3	19
<i>Hypocrea jecorina</i>	Sordariomycetes	11	9	20
<i>Magnaporthe grisea</i>	Sordariomycetes	11	4	15
<i>Neurospora crassa</i>	Sordariomycetes	8	4	12
<i>Ustilago maydis</i>	Ustilaginomycetes	2	1	3
<i>Filobasidiella neoformans</i>	Tremellomycetes	3	1	4
<i>Coprinopsis cinerea</i>	Agaricomycetes	7	1	8
<i>Phanerochaete chrysosporium</i>	Agaricomycetes	7	3	10

is present in the most recent common ancestor of all taxa included in the analysis. Therefore *Batrachochytrium dendrobatidis* was excluded from the analysis of cluster B GH18s. The phylogenetic relationships between the species that were included in the analysis are shown in Figure 1, with branch lengths in millions of years. Phylogenetic relationships and estimations of divergence times were taken from previous publications (Bowman et al. 1996; Kasuga et al. 2002; Padovan et al. 2005; Taylor and Berbee, 2006), assuming that the Devonian ascomycete *Paleopyrenomycites devonicus* (Taylor et al. 2005) represents Pezizomycotina (Taylor and Berbee, 2006) which gives an estimated age of 923 millions of years for the fungal phylum.

Alternative estimates of divergence times can be made by assuming that *P. devonicus* represents

Sordariomycetes (estimated age of the fungal phylum at 1630 millions of years) or Ascomycota (estimated age of the fungal phylum at 495 millions of years) as outlined in Taylor and Berbee, (2006), although these alternative estimates resulted in more improbable age estimates when compared with age estimates in other phyla. However, these alternative estimates were included in the analysis although *Coccidioides immitis*, *Uncinocarpus reesii*, *Ajellomyces capsulatus* and *B. dendrobatidis* were excluded because of incompatibility of divergence estimates (Bowman et al. 1996; Kasuga et al. 2002; Padovan et al. 2005; Taylor and Berbee, 2006). The four different phylogenetic trees used in the current study including branch lengths are found in Supplemental information (S1).

The birth and death parameter (λ) was estimated from the data (De Bie et al. 2006) and was 0.001

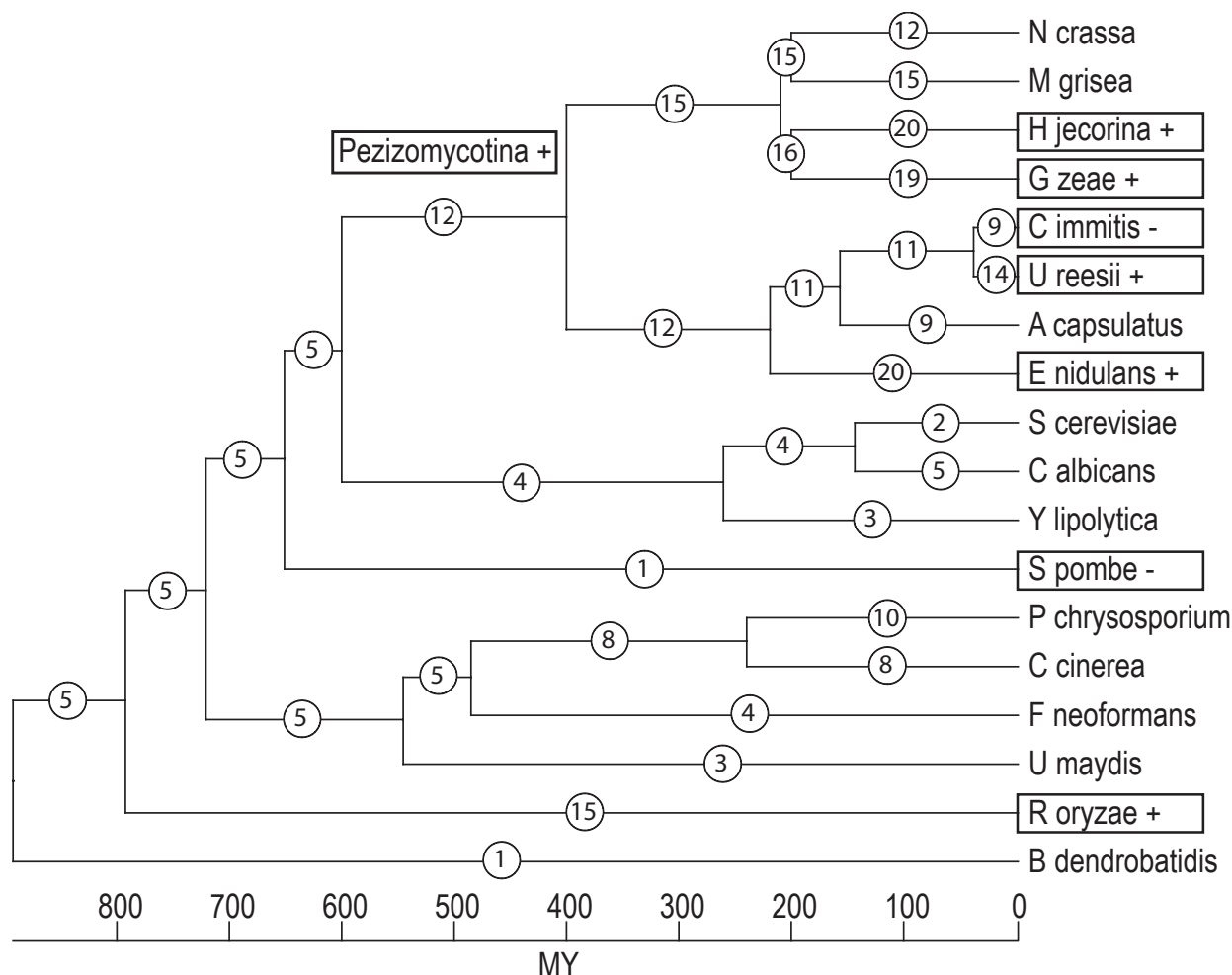


Figure 1. Distribution of GH18 gain and loss among fungal lineages. Phylogenetic relationships among the fungal species used in the current study are shown, including divergence dates in millions of years (Taylor and Berbee, 2006). Circled numbers represent total number of GH18 genes in extant species and estimates of total number of GH18 genes for ancestral species. Boxed taxon names indicates a significant (p -values ≤ 0.05 or Likelihood ratios >50) expansion (+), or a significant contraction (-) of the GH18 gene family size.

for all datasets. *p*-values were computed using 1000 re-samplings and identification of the branch that was the most likely cause of deviations from a random model was determined by Viterbi, Branch-cutting and Likelihood ratio test procedures (De Bie et al. 2006). We considered *p*-values ≤ 0.05 or likelihood ratios above 50 to be significant for branch identification.

Results

There was considerable variation in the number of GH18 genes between different fungal species, ranging from 1 in *B. dendrobatidis* and *S. pombe* to 20 in *H. jecorina* and *Emericella nidulans* (Table 1). Because of the large variation in length and domain structure of fungal GH18s (Seidl et al. 2005), only the GH18 catalytic domains were used for phylogenetic analysis. Fungal GH18s were divided into two major clusters, A and B (Figs. 2 and 3), each subdivided into groups. Cluster A sequences formed six separate groups that were equivalent with the A-II, A-III, A-IV, A-V, C-I and C-II groups described by Seidl et al. (2005) (Fig. 2). Cluster B sequences formed five separate groups, here referred to as B-I through B-V (Fig. 3). Groups B-I and B-II were described previously (Seidl et al. 2005). The average number of cluster A GH18 genes in the included species was 6.6, ranging from 0 (*S. pombe*) and 17 (*E. nidulans*). This average was higher than for cluster B GH18 genes where the average was 2.8 genes per species, ranging from 0 (*B. dendrobatidis*) and 9 (*H. jecorina*). The different GH18 subgroups were characterized by considerable differences in levels of amino-acid conservation. Subgroups A-II, A-IV, A-V and B-V showed the highest levels of mean interspecific amino-acid identity between species in Eurotiales (Table 2). Members in subgroups C-I and C-II displayed the lowest levels of conservation while levels in B-I and B-II were intermediate to the other groups (Table 2).

To investigate the evolutionary change of the number of GH18 genes in fungi, we estimated the number of genes in ancestral species and the number of gene gains and losses for each branch of the phylogenetic tree of the fungal species (Fig. 1). The analysis showed that the fungal GH18 gene family, analysing both cluster A and B together, as well as cluster A GH18s alone have evolved non-randomly ($p < 0.001$, Table 3). Six branches were identified as contributing to this

non-random pattern, including 5 expansions and 1 contraction (Table 3). These branches included the ancestor to the Pezizomycotina clade, as well as extant species *C. immitis*, *U. reesii*, *E. nidulans*, *Gibberella zeae* and *Rhizopus oryzae*. Analysis of gene phylogenies of GH18 subgroups, identified subgroups C-I and C-II as the likely targets responsible for the non-random expansion seen in *E. nidulans* and *U. reesii* (Fig. 2). The contraction in *C. immitis* probably took place in subgroup C-II (Fig. 2). It was not possible to identify the target subgroup in *G. zeae*, although C-I and C-II appeared to have expanded compared to other Sordariomycetes (Fig. 2, Seidl et al. 2005). It should be noted though, that an additional gene was also present in subgroups A-III and A-V, as compared with the closely related *H. jecorina* (Fig. 2, Seidl et al. 2005). The expansion seen in cluster A GH18s in *R. oryzae* took place in subgroup A-III (Fig. 2). Although cluster B GH18s did not display a non-random pattern of evolution as a whole ($p = 0.568$), two branches were still identified where significant expansions took place (Table 3). For both *H. jecorina* and *R. oryzae* this expansion took place in the large B-I/II/III/IV cluster (Fig. 3). Analyzing all GH18s together using *R. oryzae* as the most recent common ancestor of all taxa instead of *B. dendrobatidis* resulted in a significant contraction in *S. pombe* (Table S2).

Taylor and Berbee, (2006) also published two alternative estimates of divergence times of fungal taxa, although these alternative estimates resulted in more improbable age estimates when compared with age estimates in other phyla. Analyzing the evolution of the GH18 gene family using these alternative estimates was performed to assess the robustness of the analysis to differences in divergence dates. The two more recent estimates of divergence dates (estimated age of the fungal phylum at 495 or 923 millions of years) both showed that the GH18 family have evolved non-randomly ($p < 0.001$, Table S2), and identified the same branches as contributing to this non-random pattern (Table S2). The oldest estimate (estimated age of the fungal phylum at 1630 millions of years) gave no significant changes in size of the GH18 gene family (Table S2).

The involvement of GH18 subgroups C-I and C-II in both expansions and contractions in ascomycetes qualified them for further study. Based on the genome sequence, C-I members in *E. nidulans*

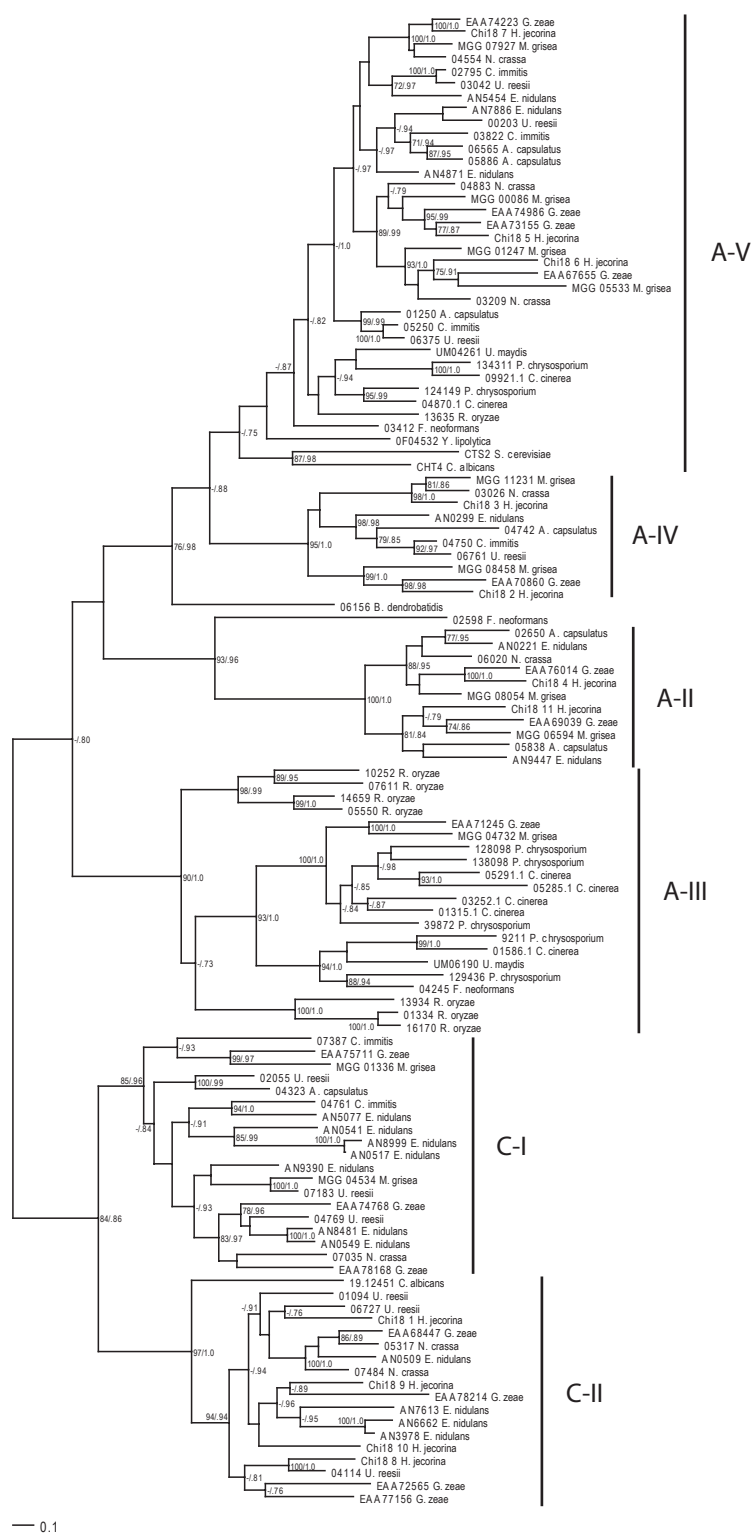


Figure 2. Phylogenetic relationships of fungal cluster A family 18 glycoside hydrolase catalytic domains. Phylogenetic analyses were performed using maximum likelihood methods as implemented in PhyML-aLRT, based on an alignment of family 18 glycoside hydrolase catalytic domain amino acid sequences. Branch support values (bootstrap proportions/approximate likelihood-ratio test probabilities) are associated with nodes, with a dash indicating that the support was <70%/0.70. The bar marker indicates the number of amino-acid substitutions. Protein identifiers include protein name, GenBank accession nos. or locus/protein ID from the respective genome projects. Group names are indicated, see text for reference. Cluster B GH18 Chi18-12 from *H. jecorina* was used as outgroup (not shown).

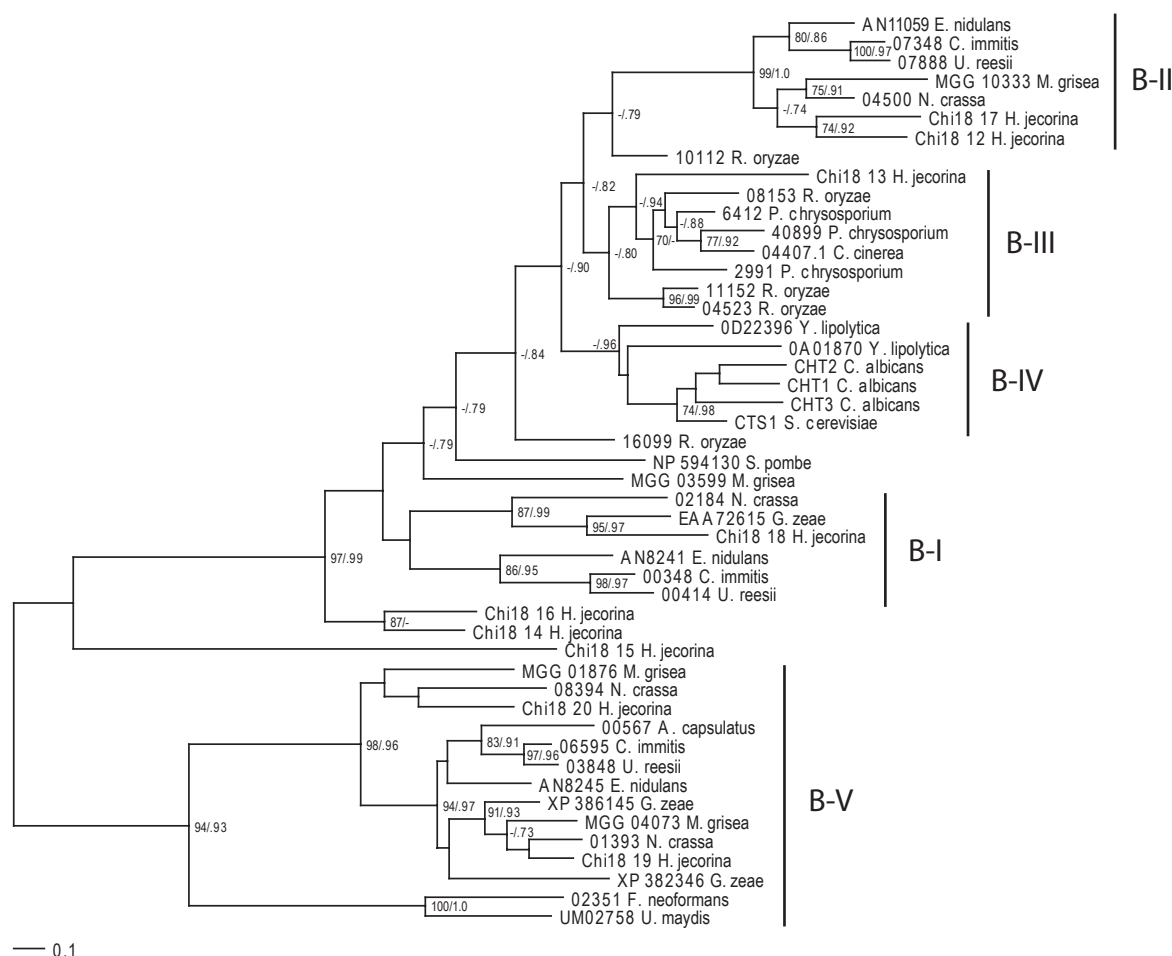


Figure 3. Phylogenetic relationships of fungal cluster B family 18 glycoside hydrolase catalytic domains. Phylogenetic analyses were performed using maximum likelihood methods as implemented in PhyML-aLRT, based on an alignment of family 18 glycoside hydrolase catalytic domain amino acid sequences. Branch support values (bootstrap proportions/approximate likelihood-ratio test probabilities) are associated with nodes, with a dash indicating that the support was <70%/0.70. The bar marker indicates the number of amino-acid substitutions. Protein identifiers include protein name, GenBank accession nos. or locus/protein ID from the respective genome projects. Group names are indicated, see text for reference. Cluster A GH18 06156 from *B. dendrobatidis* was used as outgroup (not shown).

were characterised by Chitin-binding type 1 domains (InterPro acc. no. IPR001002), while C-II members in *E. nidulans* contained LysM peptidoglycan binding domains (InterPro acc. no. IPR002482) in addition to Chitin-binding type 1 domains. These proteins showed considerable similarity with the α -subunit of the yeast killer toxin from *Kluyveromyces lactis* (Fig. 4). The A-III members in *R. oryzae* were short and contained no other domains than the GH18 catalytic domain. The domain-structure of cluster B GH18s in *H. jecorina* has been described before (Seidl et al. 2005) and were characterized by fungal cellulose binding domains (InterPro acc. no. IPR000254). The domain-structure of *R. oryzae* cluster B members included both carbohydrate-binding family V/XII domains (InterPro acc. no. IPR003610) and GH18 carbohydrate binding domains (InterPro acc. no. IPR005089).

Discussion

Accurate estimates of divergence times for fungi are notoriously hard to obtain due to a very limited fossil record. In the current study we have used three different time estimates on our GH18 gene family data. The differences relates to whether the Devonian ascomycete *P. devonicus* (Taylor et al. 2005) represents Sordariomycetes, Pezizomycotina or Ascomycota (Taylor and Berbee, 2006). Using the Sordariomycete calibration result in no significant changes in the GH18 gene family due to the estimated ancient origin of the fungal phylum (1630 millions of years), more than three times the age of the first fossil evidence of land plants (Taylor and Berbee, 2006). The results when using the two divergence estimates based on *P. devonicus* as a representative for Pezizomycotina (origin of fungal phylum at 923 millions of years) or Ascomycota (origin of fungal

Table 2. Mean intraspecific levels of % amino-acid identity for GH18 subgroups in *E. nidulans* and mean interspecific levels of % amino-acid identity for GH18 subgroups between *E. nidulans* and other members of Eurotiales.

GH18 subgroup	<i>E. nidulans</i>	<i>A. capsulatus</i>	<i>U. reesii</i>	<i>C. immitis</i>
A-II	42.4	45.9	-	-
A-IV	*	52.5	56.8	57.1
A-V	58.3	56.4	52.4	54.0
C-I	22.8	12.0	15.8	19.3
C-II	25.3	-	27.1	-
B-I	*	-	35.7	35.3
B-II	*	-	29.2	30.2
B-V	*	57.5	54.6	56.2

Abbreviations: *: Only one GH18 member in *E. nidulans*; -: no GH18 members present.

phylum at 495 millions of years) are very similar, which show a certain level of robustness of the analysis, even for large differences in divergence estimates. Furthermore, the Pezizomycotina estimate has been shown to result in least number of improbable time estimates when compared with other phyla (Taylor and Berbee, 2006).

The number of GH18 genes in different fungal species varies considerably. An expansion in size of a particular gene family or subgroup within a gene family, such as GH18s, suggests that this gene family or subgroup has been important for the fitness of the species during evolution. The observed variation could possibly be attributed to

differences in morphology, growth patterns, nutrient acquisition or antagonistic ability between species. Our approach can be used to establish links between phylogeny of the GH18 gene family with the ecological role of the species and to identify specific subgroups as important evolutionary targets in specific fungal lineages.

Filamentous ascomycetes generally possess larger number of GH18 genes as compared with other fungal groups. This larger GH18 gene family size can possibly be attributed to a larger gene copy-number in certain subgroups, but more importantly to the presence of several GH18 subgroups that appear to be unique for filamentous

Table 3. Non-randomly evolving branches in the fungal GH18 gene family.

Data set	Branch ID	<i>p</i> -value ¹	Likelihood ratio ¹	Change ²
All GH18 genes		<0.001		
	Pezizomycotina	0.005	36	7
	Coccidioides	0.032	48	-2
	Uncinocarpus	0.009	31	3
	Emericella	0.002	14	8
	Rhizopus	0.010	1	10
Cluster A GH18 genes		<0.001		
	Pezizomycotina	0.002	34	6
	Gibberella	0.026	5	5
	Coccidioides	0.002	161	-3
	Uncinocarpus	0.053	57	2
	Emericella	0.004	24	8
	Rhizopus	0.043	1	6
Cluster B GH18 genes		0.568		
	Hypocrea	0.002	14	5
	Rhizopus	0.018	1	5

¹See De Bie et al. (2006) for reference.

²Gene family size change as compared with the most recent ancestor.

Figure 4. Partial alignment of *K. lactis* zymocin with subgroups C-I and C-II members in *E. nidulans*. Identical residues in a column are indicated in white and boxed in gray, two different residues in a column are indicated by gray boxes, gaps are indicated by dashes. The conserved GH18 gene family active site residues are indicated by asterisks. AN10838 and AN0509 represents the C-II GH18 subgroup, AN5077, AN0517 and AN8481 represents the C-I subgroup. KT = Killer Toxin α -subunit.

the pathogenic *C. immitis*, although these species are very closely related (Bowman et al. 1996; Kasuga et al. 2002). This difference should be related to the different life-styles of the two fungi, and we hypothesize that the expansion of C-II seen in *U. reesii* is a consequence of the need for antagonistic interactions with other soil-dwelling fungi. The contraction in *C. immitis* may be reflecting adaptation to a pathogenic lifestyle, where antagonistic interactions with other fungi are minimized. Another human pathogen, *A. capsulatus*, also contains the same number of GH18 genes (9) as *C. immitis* and also lacks subgroup C-II members completely, which indicate that the function performed by C-II members of the GH18 protein family is dispensable for the human pathogenic lifestyle.

Another intriguing result is the different evolutionary trajectories of the GH18 gene family between the human pathogen *C. immitis* and the closely related *U. reesii*. An expansion of GH18 subgroup C-II in saprotrophic *U. reesii* is in contrast to a contraction in the same subgroup in

obscure. Although very speculative, the emergence of C-I and C-II GH18s coincide with the estimated time of colonization of terrestrial environments by plants (Sanderson, 2003), which suggests the possibility that the emergence of terrestrial plants created new ecological niches where filamentous ascomycetes could expand into to compete for space and nutrients.

Another expansion took place in cluster B GH18s in *H. jecorina*, which is closely related to mycoparasitic fungi such as *H. lixii*, *H. virens* and *H. atroviridis*. Seidl et al. (2005) reported that certain cluster B GH18 genes from *H. jecorina* have high similarity to GH18s from entomopathogens, such as *Metarhizium anisopliae*, which suggests an aggressive role of these proteins in chitin degradation. Again, there is expression data that support this role; *chi18-13* from *H. atroviridis* is up-regulated during growth on fungal cell walls and during plate confrontation assays (Seidl et al. 2005). The fact that two GH18 subgroups that are implied in aggressive fungal-fungal interactions have expanded significantly during fungal evolution suggests that interspecific antagonistic interactions are important determinants of fungal evolution, community development and functioning. This result is in line with the idea that genes involved in stress-related functions can tolerate, or are even under selection for, increases in copy number (Wapinski et al. 2007).

Species with yeast-like or monocentric growth styles such as *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *S. pombe* and *B. dendrobatidis* all have low numbers of GH18 genes (2, 3, 1 and 1 genes) as compared with other fungi, even with closely related species exhibiting filamentous growth e.g. *Candida albicans* (5 genes). This suggests that filamentous growth requires the combined action of several GH18s as compared with a yeast-like growth style, even though the reduction in the non-filamentous species are only significant for *S. pombe* ($p = 0.023$).

The analysis of the GH18 gene family size in basidiomycetes shows no conflict with a random process. However, 5 of the 8 GH18 genes in *Coprinopsis cinerea* and 5 of the 10 GH18 genes in *Phanerochaete chrysosporium* are found in subgroup A-III, compared to only one representative from *Ustilago maydis* and *Filobasidiella neoformans*. Further studies of more closely related species with shorter coalescence times will be needed to determine if the apparent

expansion of GH18 subgroup A-III in saprotrophic basidiomycetes and the apparent contraction in pathogenic basidiomycetes is a consequence of the different life-styles of the species. Based on the observed expansion of GH18 subgroups involved in interspecific interactions in ascomycetes, we can hypothesize about the involvement of A-III GH18s in fungal-fungal interactions in basidiomycetes.

Our analysis shows the usefulness of the combination of a stochastic birth and death model and phylogenetic information in a probabilistic framework for identification of lineages with unusually evolving gene families. However, the birth and death model assumes independence among individual genes. This means that any large-scale chromosome duplication, deletion or polyploidization that acts on several gene family members at once violates the assumption of the model (Hahn et al. 2005). Interpretation of gene family size differences between taxa that are separated by genome duplications should be made with caution. There are indications of a recent polyploidization in *R. oryzae* (Taylor and Berbee, 2006) which suggest the possibility that the observed non-random GH18 family size in this species may not be entirely related to adaptive selection. On the other hand, the two GH18 genes in the duplicated *S. cerevisiae* genome (Kellis et al. 2004) belong to the fundamentally different A and B clusters. This indicate that the loss of one of the duplicated *CTS1* and *CTS2* paralogues have been selected for during evolution after whole-genome duplication. *Cts1* and *cts2* are involved in cell separation during budding and in sporulation (Kuranda and Robbins, 1991; Giaever et al. 2002), which is in line with the idea that genes involved in growth-related functions are under selection against changes in copy numbers (Wapinski et al. 2007).

Another violation against the assumption of independence among individual genes may be seen in *Neurospora crassa*. This species has the lowest number of GH18 genes (12) among the Sordariomycetes, which may be attributed to the presence of a wide array of genome defence mechanisms, including repeat-induced point mutations, greatly slowing down the creation of new genes (Galagan et al. 2003). The low number of GH18 genes in this species is not significantly violating a random process ($p = 0.256$), but interpretation of data for other gene families in *N. crassa* should be done with caution.

In the current study we have used fungal genome data in comparative work to infer phylogenetic relationships in the fungal GH18 gene family and to detect non-random expansions and contractions. This approach can be used to establish links between phylogeny of the GH18 gene family with the ecological role of the species, and identify specific GH18 subgroups as important evolutionary targets in specific fungal lineages. Within the fungal GH18 gene family we observe selection against changes in copy number in GH18s involved in growth and development as well as selection for increased copy number in GH18s involved in stress-related functions, supporting the idea of a bipolar principle that governs tolerance to duplications and losses (Wapinski et al. 2007). The results also indicate that antagonistic fungal-fungal interactions constitute an important evolutionary force in soil borne ascomycetes, but not for fungal species that are pathogenic in humans.

Acknowledgements

This work was supported by a grant from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Uppsala Microbiomics Center). Genome sequence data were produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>), The Fungal Genome Initiative at the BROAD Institute (<http://www.broad.mit.edu/annotation/fungi/fgi/index.html>) and Génolevures (<http://cblabri.u-bordeaux.fr/Genolevures/blast/index.php>). We thank Timothy James, Åke Olson and Nils Högberg for critical reading of the manuscript.

References

- Adams, D.J. 2004. Fungal cell wall chitinases and glucanases. *Microbiol.*, 150:2029–35.
- Altschul, S.F., Madden, T.L., Schäffer, A.A. et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389–402.
- Anisimova, M. and Gascuel, O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.*, 55:539–52.
- Benitez, T., Rincon, A.M., Limon, M.C. et al. 2004. Biocontrol mechanisms of *Trichoderma* strains. *Int. Microbiol.*, 7:249–60.
- Bishop, J.G., Dean, A.M. and Mitchell-Olds, T. 2000. Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. U.S.A.*, 97:5322–7.
- Bowman, B.H., White, T.J. and Taylor, J.W. 1996. Human pathogenic fungi and their close nonpathogenic relatives. *Mol. Phyl. Evol.*, 6:89–96.
- Clark, N.L., Aagaard, J.E. and Swanson, W.J. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction*, 131:11–22.
- De Bie, T., Cristianini, N., Demuth, J.P. et al. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, 22:1269–71.
- Dong, L.Q., Yang, J.K. and Zhang, K.Q. 2007. Cloning and phylogenetic analysis of the chitinase gene from the facultative pathogen *Paecilomyces lilacinus*. *J. Appl. Microbiol.*, 103:2476–88.
- Duo-Chuan, L. 2006. Review of fungal chitinases. *Mycopath.*, 161:345–60.
- Felse, P.A. and Panda, T. 1999. Regulation and cloning of microbial chitinase genes. *App. Microbiol. Biotech.*, 51:141–51.
- Galagan, J.E., Calvo, S.E., Borkovich, K.A. et al. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*, 422:859–68.
- Gan, Z., Yang, J., Tao, N. et al. 2007a. Cloning of the gene *Lecanicillium psalliotae* chitinase *Lpch1* and identification of its potential role in the biocontrol of root-knot nematode *Meloidogyne incognita*. *Appl. Microbiol. Biotechnol.*, 76:1309–17.
- Gan, Z., Yang, J., Tao, N. et al. 2007b. Cloning and expression analysis of a chitinase gene *Crchi1* from the mycoparasitic fungus *Clonostachys rosea* (syn. *Gliocladium roseum*). *J. Microbiol.*, 45:422–30.
- Giaever, G., Chu, A.M., Ni, L. et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, 418:387–91.
- Gilkes, N.R., Henrissat, B., Kilburn, D.G. et al. 1991. Domains in microbial beta-1,4-glycanases—Sequence conservation, function, and enzyme families. *Microbiol. Rev.*, 55:303–15.
- Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, 52:696–704.
- Hahn, M.W., De Bie, T., Stajich, J.E. et al. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.*, 15:1153–60.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.*, 41:95–8.
- Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, 280:309–16.
- Henrissat, B. and Bairoch, A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, 781–8.
- Henrissat, B. and Davies, G.J. 2000. Glycoside hydrolases and glycosyltransferases Families, modules, and implications for genomics. *Plant Physiol.*, 124:1515–9.
- Hoell, I.A., Klemsdal, S.S., Vaaje-Kolstad, G. et al. 2005. Overexpression and characterization of a novel chitinase from *Trichoderma atroviride* strain P1. *Biochim. Biophys. Acta. Prot. Proteom*, 1748:180–90.
- Howell, C.R. 2003. Mechanisms employed by *Trichoderma* species in the biological control of plant diseases: The history and evolution of current concepts. *Plant Dis.*, 87:4–10.
- Iseli, B., Armand, S., Boller, T. et al. 1996. Plant chitinases use two different hydrolytic mechanisms. *FEBS Lett.*, 382:186–8.
- Jones, D.T., Taylor, W.R. and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comp. Appl. Biosci.*, 8:275–82.
- Karlsson, M., Nygren, K. and Johannesson, H. 2008. The evolution of the pheromonal signal system and its potential role for reproductive isolation in heterothallic *Neurospora*. *Mol. Biol. Evol.*, 25:168–78.
- Kasuga, T., White, T.J. and Taylor, J.W. 2002. Estimation of nucleotide substitution rates in eukaryotic fungi. *Mol. Biol. Evol.*, 19:2318–24.
- Kellis, M., Birren, B.W. and Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, 428:617–24.
- Keyhani, N.O. and Roseman, S. 1999. Physiological aspects of chitin catabolism in marine bacteria. *Biochim. Biophys. Acta. Gen. Sub.*, 1473:108–22.
- Klemsdal, S.S., Clarke, J.H.L., Hoell, I.A. et al. 2006. Molecular cloning, characterization, and expression studies of a novel chitinase gene (*ech30*) from the mycoparasite *Trichoderma atroviride* strain P1. *FEMS Microbiol. Lett.*, 256:282–9.

- Kuranda, M.J. and Robbins, P.W. 1991. Chitinase is required for cell-separation during growth of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, 266:19758–67.
- Limon, M.C., Margolles-Clark, E., Benitez, T. et al. 2001. Addition of substrate-binding domains increases substrate-binding capacity and specific activity of a chitinase from *Trichoderma harzianum*. *FEMS Microbiol. Lett.*, 198:57–63.
- Magliani, W., Conti, S., Gerloni, M. et al. 1997. Yeast killer systems. *Clin. Microbiol. Rev.*, 10:369–400.
- Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F. et al. 2005. CDD: a conserved domain database for protein classification. *Nucleic Acids Res.*, 33:D192–6.
- Padovan, A.C.B., Sanson, G.F.O., Brunstein, A. et al. 2005. Fungi evolution revisited: Application of the penalized likelihood method to a bayesian fungal phylogeny provides a new perspective on phylogenetic relationships and divergence dates of ascomycota groups. *J. Mol. Evol.*, 60:726–35.
- Sanderson, M.J. 2003. Molecular data from 27 proteins do not support a Precambrian origin of land plants. *Am. J. Bot.*, 90:954–956.
- Seidl, V., Huemer, B., Seiboth, B. et al. 2005. A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS J.*, 272:5923–39.
- Stark, M.J.R., Boyd, A., Mileham, A.J. et al. 1990. The plasmid-encoded killer system of *Kluyveromyces lactis*—a review. *Yeast*, 6:1–29.
- Suzuki, K., Taiyoji, M., Sugawara, N. et al. 1999. The third chitinase gene (chiC) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases. *Biochem. J.*, 343:587–96.
- Taylor, J.W. and Berbee, M.L. 2006. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia*, 98:838–49.
- Taylor, T.N., Hass, H., Kerp, H. et al. 2005. Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism (vol 96, pg 1403, 2004). *Mycologia*, 97:269–85.
- Terwisscha van Scheltinga, A.C.T., Kalk, K.H., Beintema, J.J. et al. 1994. Crystal-structures of hevamine, a plant defense protein with chitinase and lysozyme activity, and its complex with an inhibitor. *Structure*, 2:1181–9.
- Thompson, J.D., Gibson, T.J., Plewniak, F. et al. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25:4876–82.
- Tiffin, P. 2004. Comparative evolutionary histories of chitinase genes in the genus *Zea* and family Poaceae. *Genetics*, 167:1331–40.
- van Aalten, D.M.F., Komander, D., Synstad, B. et al. 2001. Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proc. Natl. Acad. Sci. U.S.A.*, 98:8979–84.
- van Aalten, D.M.F., Synstad, B., Brurberg, M.B. et al. 2000. Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9-angstrom resolution. *Proc. Natl. Acad. Sci. U.S.A.*, 97:5842–7.
- Wapinski, I., Pfeffer, A., Friedman, N. et al. 2007. Natural history and evolutionary principles of gene duplication in fungi. *Nature*, 449:54–61.
- Warren, R.A.J. 1996. Microbial hydrolysis of polysaccharides. *Ann. Rev. Microbiol.*, 50:183–212.
- Wattanalai, R., Boucias, D.G. and Tartar, A. 2004. Chitinase gene of the dimorphic mycopathogen, *Nomuraea rileyi*. *J. Invertebr. Pathol.*, 85:54–57.
- Zdobnov, E.M. and Apweiler, R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 17:847–8.

Comparative Evolutionary Histories of the Fungal Chitinase Gene Family Reveal Non-Random Size Expansions and Contractions due to Adaptive Natural Selection

Magnus Karlsson and Jan Stenlid

Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, P.O. 7026, SE-75007, Uppsala, Sweden.

Supplemental Information

Tree File S1. Phylogenetic trees with estimates of divergence times used in the current study.

Tree 1. Phylogenetic tree and estimates of divergence time in millions of years for all included fungi based on the Pezizomycotina calibration.

(((((N_crassa:200,M_grisea:200):10,(H_jecorina:200,G_zeae:200):10):190,(((C_immitis:38,U_reesii:38):119,A_capsulatum:157):63,E_nidulans:220):180):200,((S_cerevisiae:145,C_albicans:145):115,Y_lipolytica:260):340):50,S_pombe:650):70,(((P_chrysosporium:240,C_cinerea:240):245,F_neoformans:485):60,U_maydis:545):175):72,R_oryzae:792):101,B_dendrobatidis:893);

Tree 2. Phylogenetic tree and estimates of divergence time in millions of years for a subset of the included fungi based on the Ascomycota calibration.

(((((N_crassa:120,M_grisea:120):10,(H_jecorina:120,G_zeae:120):10):90,(C_immitis:130,E_nidulans:130):180):90,((S_cerevisiae:90,C_albicans:90):180,Y_lipolytica:270):90):140,S_pombe:390):60,(((P_chrysosporium:150,C_cinerea:150):150,F_neoformans:300):40,U_maydis:340):110):30,R_oryzae:480);

Tree 3. Phylogenetic tree and estimates of divergence time in millions of years for a subset of the included fungi based on the Pezizomycotina calibration.

(((((N_crassa:200,M_grisea:200):10,(H_jecorina:200,G_zeae:200):10):190,(C_immitis:220,E_nidulans:220):180):200,((S_cerevisiae:145,C_albicans:145):115,Y_lipolytica:260):340):50,S_pombe:650):70,(((P_chrysosporium:240,C_cinerea:240):245,F_neoformans:485):60,U_maydis:545):175):72,R_oryzae:792);

Tree 4. Phylogenetic tree and estimates of divergence time in millions of years for a subset of the included fungi based on the Sordariomycete calibration.

(((((N_crassa:400,M_grisea:400):10,(H_jecorina:400,G_zeae:400):10):290,(C_immitis:420,E_nidulans:420):280):480,((S_cerevisiae:280,C_albicans:280):590,Y_lipolytica:870):310):130,S_pombe:1310):160,(((P_chrysosporium:500,C_cinerea:500):480,F_neoformans:980):120,U_maydis:1100):370):130,R_oryzae:1600);

Table S2. Non-randomly evolving branches in the fungal GH18 gene family using alternative calibration points for dating fungal divergences.

Data set	Branch ID	Sordariomycetes ¹ p-value ⁴	Likelihood ratio ⁴	Pezizomycotina ² p-value ⁴	Likelihood ratio ⁴	Ascomycota ³ p-value ⁴	Likelihood ratio ⁴
All GH18 genes		1		<0.001		<0.001	
	Pezizomycotina	-	-	0.007	22	0.007	22
	Emericella	-	-	0.006	14	0.006	14
	Rhizopus	-	-	0.024	1	0.024	1
Cluster A GH18 genes	Schizosaccharomyces	-	-	0.023	3	0.023	3
		1		0.002		0.002	
	Pezizomycotina	-	-	0.007	16	0.007	16
	Gibberella	-	-	0.001	5	0.006	5
Cluster B GH18 genes	Emericella	-	-	0.004	39	<0.001	39
	Rhizopus	-	-	0.043	1	0.043	1
		1		0.407		0.033	
	Hypocrea	-	-	0.002	14	<0.001	98
	Rhizopus	-	-	0.018	1	0.005	2

¹Estimates of divergence dates calibrated with *P. devonicus* as representing Sordariomycetes, giving an estimate for the fungal phylum at 1630 millions of years (Taylor and Berbee, 2006).
²Estimates of divergence dates calibrated with *P. devonicus* as representing Pezizomycotina, giving an estimate for the fungal phylum at 923 millions of years (Taylor and Berbee, 2006).
³Estimates of divergence dates calibrated with *P. devonicus* as representing Ascomycota, giving an estimate for the fungal phylum at 495 millions of years (Taylor and Berbee, 2006).
⁴De Bie, T., Cristianini, N., Demuth, J.P. et al. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, 22:1269–71.
Abbreviation: -:no significant GH18 gene family expansion or contraction.