RESEARCH ARTICLE

Species-Specific Chitin-Binding Module 18 Expansion in the Amphibian Pathogen *Batrachochytrium dendrobatidis*

John Abramyan and Jason E. Stajich

Department of Plant Pathology and Microbiology, University of California, Riverside, California, USA

ABSTRACT *Batrachochytrium dendrobatidis* is the causative agent of chytridiomycosis, which is considered one of the driving forces behind the worldwide decline in populations of amphibians. As a member of the phylum *Chytridiomycota*, *B. dendrobatidis* has diverged significantly to emerge as the only pathogen of adult vertebrates. Such shifts in lifestyle are generally accompanied by various degrees of genomic modifications, yet neither its mode of pathogenicity nor any factors associated with it have ever been identified. Presented here is the identification and characterization of a unique expansion of the carbohydrate-binding module family 18 (CBM18), specific to *B. dendrobatidis*. CBM (chitin-binding module) expansions have been likened to the evolution of pathogenicity in a variety of fungus species, making this expanded group a prime candidate for the identification of potential pathogenicity factors. Furthermore, the CBM18 expansions are confined to three categories of genes, each having been previously implicated in host-pathogen interactions. These correlations highlight this specific domain expansion as a potential key player in the mode of pathogenicity in this unique fungus. The expansion of CBM18 in *B. dendrobatidis* is exceptional in its size and diversity compared to other pathogenic species of fungi, making this genomic feature unique in an evolutionary context as well as in pathogenicity.

IMPORTANCE Amphibian populations are declining worldwide at an unprecedented rate. Although various factors are thought to contribute to this phenomenon, chytridiomycosis has been identified as one of the leading causes. This deadly fungal disease is cause by *Batrachochytrium dendrobatidis*, a chytrid fungus species unique in its pathogenicity and, furthermore, its specificity to amphibians. Despite more than two decades of research, the biology of this fungus species and its deadly interaction with amphibians had been notoriously difficult to unravel. Due to the alarming rate of worldwide spread and associated decline in amphibian populations, it is imperative to incorporate novel genomic and genetic techniques into the study of this species. In this study, we present the first reported potential pathogenicity factors in *B. dendrobatidis*. *In silico* studies such as this allow us to identify putative targets for more specific molecular analyses, furthering our hope for the control of this pathogen.

Address correspondence to Jason E. Stajich, jason.stajich@ucr.edu.

Global decline in biodiversity is edging towards prehistoric proportions (1). This grim fact is epitomized by the recent catastrophic decline in amphibians, of which ~40% are imperiled (2). Various causes for this phenomenon are suspected, including climate change and habitat loss (3), but *Batrachochytrium dendrobatidis*, the causative agent of chytridiomycosis, is often cited as the most robust explanation (2). *B. dendrobatidis* is a pathogenic fungus of the phylum *Chytridiomycota* (4), a group comprised of generalist saprobes and parasites of various organisms ranging from algae to microscopic animals (5). Of the two pathogenic *Chytridiomycete* species identified to be vertebrate specific, *B. dendrobatidis* is the only one specific to amphibians (6, 7). This specificity has evolved despite the complex antimicrobial skin secretions which amphibians are known for (8). In addition to causing individual mortality, chytridiomycosis devastates entire populations, causing declines leading to eventual extinction (2).

From its discovery in 1998 until 2009, the physiological effect of *B. dendrobatidis* on the amphibian host was unknown (9). Re-

cently, in a series of studies, Voyles and colleagues found that *B. dendrobatidis* causes electrolyte imbalance in the amphibian system (10-12). Additionally, *B. dendrobatidis*-infected frogs show a weak immune response to the pathogen. When transcripts of amphibian immunity-related genes were measured, the few genes that did show up-regulation did so at later time points of infection, contrary to common belief, indicating possible suppression of the immune system by *B. dendrobatidis* (13).

To date, the majority of studies involving *B. dendrobatidis* have focused on the ecology and the population-specific virulence of the fungus, providing little information for understanding its pathogenicity on a molecular level. However, the sequencing of genomes from two *B. dendrobatidis* isolates (JEL423 and JAM81) has opened the gates to genetic analyses on various levels, ranging from population-level genomics to functional genetics (14). Since *B. dendrobatidis* is aquatic and likely specific to its amphibian host (a vertebrate with a biphasic lifestyle), substantial restructuring of its genome would likely have been required in order to adapt to

this new environmental niche, overcome host defense factors, and acquire nutrients from this new source.

Pathogens are in a constant evolutionary arms race with the hosts' coevolved defenses in order to effectively attack the host as well as obtain nutrients (15). This close interaction is the likely reason why virtually all pathogenic fungi are host specific (none attack all plants or animals) and, furthermore, tissue specific (16). Some of the key innovations associated with the evolution of a pathogenic lifestyle in fungi include production of a host-specific toxin (*Alternaria alternata* [17]), morphological restructuring (conversion from yeast to hyphal state) (*Candida albicans* [18]), expansion of protein-coding genes (*C. albicans* [19]), and specific gene loss associated with reduction of diet breadth and host range (*Metarhizium anisopliae* [20]). Recent studies of the phylum *Ascomycota* revealed a striking number of gene gains and losses, numbering in the thousands for some species, in lineages where pathogenic lifestyles have evolved (21, 22).

One of the more commonly observed genomic modifications is lineage-specific expansion (LSE) of gene families and domains within a gene. These expansions often occur through tandem duplication events (23). In eukaryotes, nearly half of all paralogous protein clusters have been generated through LSEs (24). A recent study of the *Coprinus cinereus* genome revealed an overrepresentation of paralogous, multicopy genes in genomic regions with high rates of meiotic recombination, describing a possible mechanism for LSE generation in fungi (25). Due to the redundancy of the duplicates, gene and domain expansions provide raw material for the evolution of novel functions (26). New functional roles can arise through domain accretion or reshuffling (27, 28). Additionally, the availability of genomic material through LSEs also allows for a more effective response to new and diverse environmental pressures (24).

Domain expansions, which result in multimodularity, are classified as "homogenous" and "heterogeneous" (29). Homogenous expansions occur when domains are more closely related to each other than to a domain on another gene, likely being formed by tandem duplication events. Homogenous multimodularity is likely to lead to stronger binding affinity, since the overall affinity of the CBMs (chitin-binding modules) operating in tandem is greater than that of the individual domains working independently (29). Heterogeneous domain expansions, on the other hand, are observed when one of the domains is more closely related to a domain on another locus than to any domain in the same gene. Heterogeneous multimodularity allows for distinct ligand-binding specificities, potentially giving an enzyme multiple roles (29). Each type of domain expansion enhances protein function in different ways that may play specific roles in recognition of or adhesion to external chitin molecules.

While LSEs allow for the attainment of the genetic profile required for pathogenicity, they do not necessarily lead directly to pathogenesis but provide a foundation which increases the likelihood of development towards a pathogenic lifestyle. This pattern has been observed for keratinase and peptidase gene families in the *Onygenales* fungi and implicated in the evolution of *Coccidioides* as a vertebrate pathogen (30). Many LSEs found in fungi have pathogenicity-related functions, such as secondary metabolite production, melanin biosynthesis, carbohydrate binding, and cell wall degradation (24, 31). We observed that carbohydrate-binding module family 18 (CBM18 [chitin-binding module 18]; Pfam ID PF00187) is an LSE in *B. dendrobatidis* and sought to test

whether its recent expansion could be predicted as a pathogenesis factor.

CBM18 is a subclass of chitin recognition domains that have evolved convergently in fungi, plants, and arthropods (32). The stereotypical chitin-binding domain is the hevein domain, first discovered in the latex of rubber trees (*Hevea brasiliensis*) (33). Most CBMs have 30 to 43 residues organized around a conserved four-disulfide core, known as the hevein domain of a chitin-binding motif (32, 34). Although proteins such as chitinases, chitin-binding lectins, and other similar molecules contain hevein-like domains, their roles are radically different due to their subcellular localization, molecular sizes, and overall structure (35). Additionally, binding may not be relegated to chitin itself but may also extend to any complex glycoconjugates containing *N*-acetyl-d-glucosamine (GlcNAc, the monomeric unit of chitin) or *N*-acetylneuraminic acid (Neu5Ac) (34).

When the genome of *B. dendrobatidis* was initially analyzed (C. A. Cuomo, J. E. Stajich, and I. V. Grigoriev, unpublished data), the expansion of CBM18 was particularly apparent due to its extraordinary size and breadth. Although multiple CBM loci are common (29, 36), we focused on this expansion due to its potential as a unique cell wall and/or cell surface protein component which may contribute significantly to the pathogenic character of *B. dendrobatidis*. In the rice blast fungus, *Magnaportha oryzae*, there are 36 copies of a similar domain distributed among 21 predicted proteins (37). The expanded domain in *Magnaportha oryzae* has a cysteine pattern similar to those of the *B. dendrobatidis* CBMs and the AVR4 protein, which was previously shown to protect fungal cell walls from chitinases (38). To better characterize the history of the expansion of CBM18 in the *B. dendrobatidis* genome, we performed a phylogenetic analysis of the domains and their encoding loci. We detected both homogenous and heterogeneous expansions within a relatively large number of *B. dendrobatidis* genes. We also tested for positive selection both within loci and within the domains themselves, uncovering significant evidence of recent positive selection at individual sites, which is suggestive of an adaptive role for many CBM18 domains.

RESULTS

Pfam search results. When the CBM18 Pfam domain HMM was used to search against the predicted protein coding genes from the *B. dendrobatidis* genome, with an *E* value cutoff of 0.01, we identified 67 domain copies located on 18 genes (Fig. 1). The number of domains per locus ranged from a single domain to 11. The significance of this finding was furthered by our analysis of the closest sequenced genome of another chytridiomycete, *Homolaphlyctis polyrhiza*. *H. polyrhiza* has only 10 CBM18 copies. The *B. dendrobatidis* chitin-binding domain had a homologous (cysteine) pattern of $X_3CGX_7CX_4CCSX_4CX_6CX_3C$. This motif is also conserved between the *B. dendrobatidis* CBM18 and the hevein domain of plants (see Fig. S1 in the supplemental material). In addition to the CBMs, some of these genes also contained additional binding domains such as a tyrosinase domain (tyrosinase like [TL]) and a deacetylase domain (deacetylase-like [DL]) as identified using PHOG Universal Proteome Explorer, v.2.0. The third group consisted of genes with no secondary domains (lectin-like [LL]) (Fig. 1). Although other organisms have larger numbers of genes with CBMs, the 11 copies of CBM18 on BDEG_01757 likely represent one of the largest expansion within a gene (http://www.cazy.org/). Corresponding BDEG numbers also reflect the
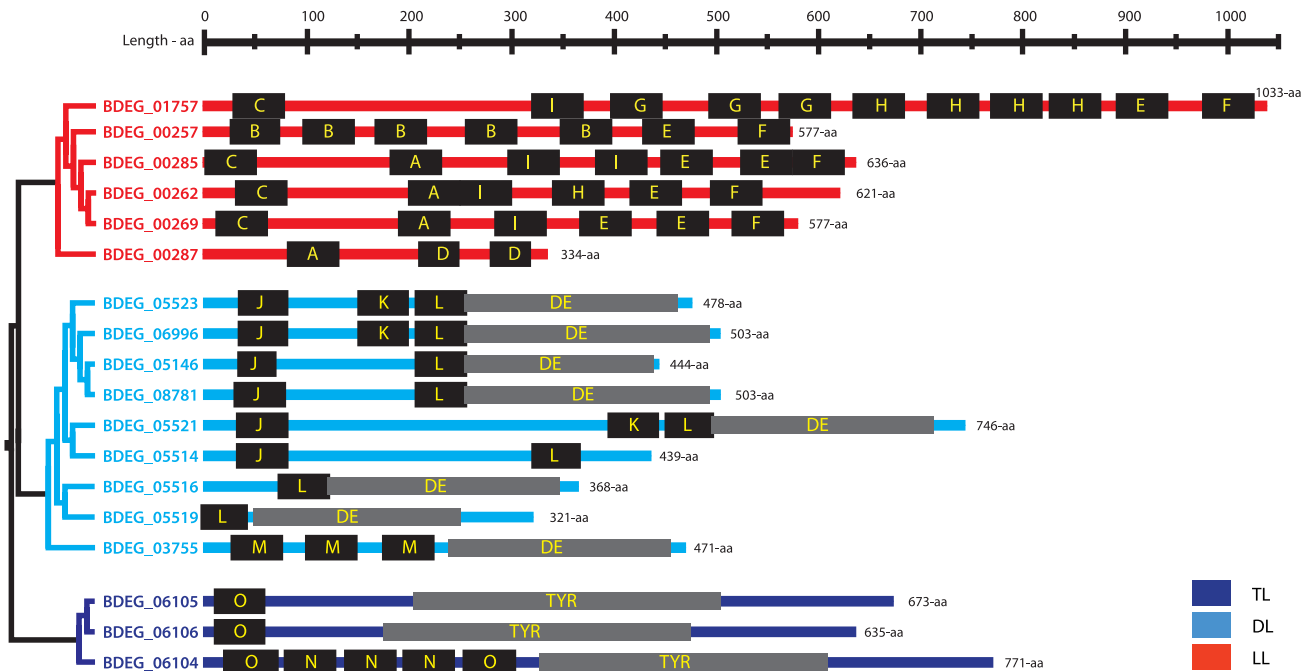
FIG 1 Domain positions on corresponding genes. CBM18 (chitin-binding module 18) domains are represented by black boxes with the phylogenetic domain groups labeled in yellow. Gray boxes represent non-CBM (chitin-binding module) domains identified on some genes. All domain lengths and positions are calculated on an amino acid level, beginning from the first residue of the gene. Widths of boxes and interdomain spaces are representative of relative domain lengths and spaces, respectively. Genes are color coded according to type. An unrooted cladogram generated from a Bayesian tree shows the phylogenetic relationships between genes. TL, tyrosinase-like; DL, deacetylase-like; LL, lectin-like; DE, deacetylase domain; TYR, tyrosinase domain; aa, amino acids.

clustering of certain groups of genes with respect to chromosomal position (e.g., BDEG_06104 to BDEG_06106, BDEG_05514, BDEG_05516, BDEG_05519, BDEG_05521, and BDEG_05523), indicative of a tandem duplication pattern.

Gene tree. In order to understand the evolutionary relationship of the CBM18 genes, a Bayesian analysis was performed with the *H. polyrhiza* genes as a representative outgroup (see Fig. S2 in the supplemental material). Most major terminal clades in the tree show strong support with Bayesian posterior probabilities (BPP) ranging from 0.9 to 1.0. The tree is divided into three monophyletic and strongly supported clades consisting of TL (tyrosinase-like), DL, and LL genes. The clade representing the TL group is paraphyletic due to the presence of *H. polyrhiza* genes within the clade. The DL group is sister to the LL group, with a BPP of 1. Furthermore, intron-exon patterns and sizes display a distinct similarity within members of each group (see Fig. S3 in the supplemental material).

Domain tree. After the characterization of the CBM containing loci, the 67 individual domain sequences were isolated and aligned with the 10 from the outgroup *H. polyrhiza* (Fig. 2). This alignment was then used to build a phylogram using a Bayesian analysis in order to understand the relationship of the domains within and between loci. The Bayesian analysis revealed a relationship between domains that varies significantly from that of their parent loci. To simplify our analysis, we decided to partition the large number of domains further by assigning a letter to each monophyletic clade, referred to here as groups (Fig. 2).

Additionally, the individual domain groups were all monophyletic with respect to the type of gene they were located on (e.g., group N and O members are all found in various TL genes, while

group J and L members are all found in DL genes). The domains found on the LL genes form a monophyletic clade, although with weak support, while the clade containing groups J, K, L, M, N, and O is comprised of members from both DL and TL genes (Fig. 2). Furthermore, group J, which is found in multiple DL genes, is monophyletic and sister to LL genes. Two *H. polyrhiza* domains were within the DL and LL clades, while the rest were paraphyletic with the TL group, in a similar manner to the gene analysis. Overall, the Bayesian analysis yielded a well-resolved phylogenetic tree which, although reflective of the evolutionary history of the loci, still shows marked individual evolutionary trajectories for the domains, independent of the type of gene they are located on.

Domain order. When the domain positions were mapped on the genes, we observed some general patterns which were reflective of the phylogenetic relationships of the domains. First, groups J, K, L, M, N, and O were restricted to the DL and TL genes, while the LL genes contained members of groups A, B, C, D, E, F, G, H, and I (Fig. 1). Furthermore, the members of each group occupied the same position in homologous genes (e.g., BDEG_06996, BDEG_05523, and BDEG_05521 contain domains from groups J, K, and L in the same order). There were some anomalous genes, however, such as locus BDEG_03755, which consisted entirely of domains from group M, which seems to be a sister group to the N group found in a TL gene (Fig. 2). Additionally, the DL genes mostly contained J, K, and L domains, with various losses of K and J observed. TL genes all contained group O domains, with one, BDEG_06104, containing three copies of N group domains, a unique domain found only in this locus. Interestingly, when the Bayesian analysis was performed on the domains, members of group O and N do not form a monophyletic clade (Fig. 2).
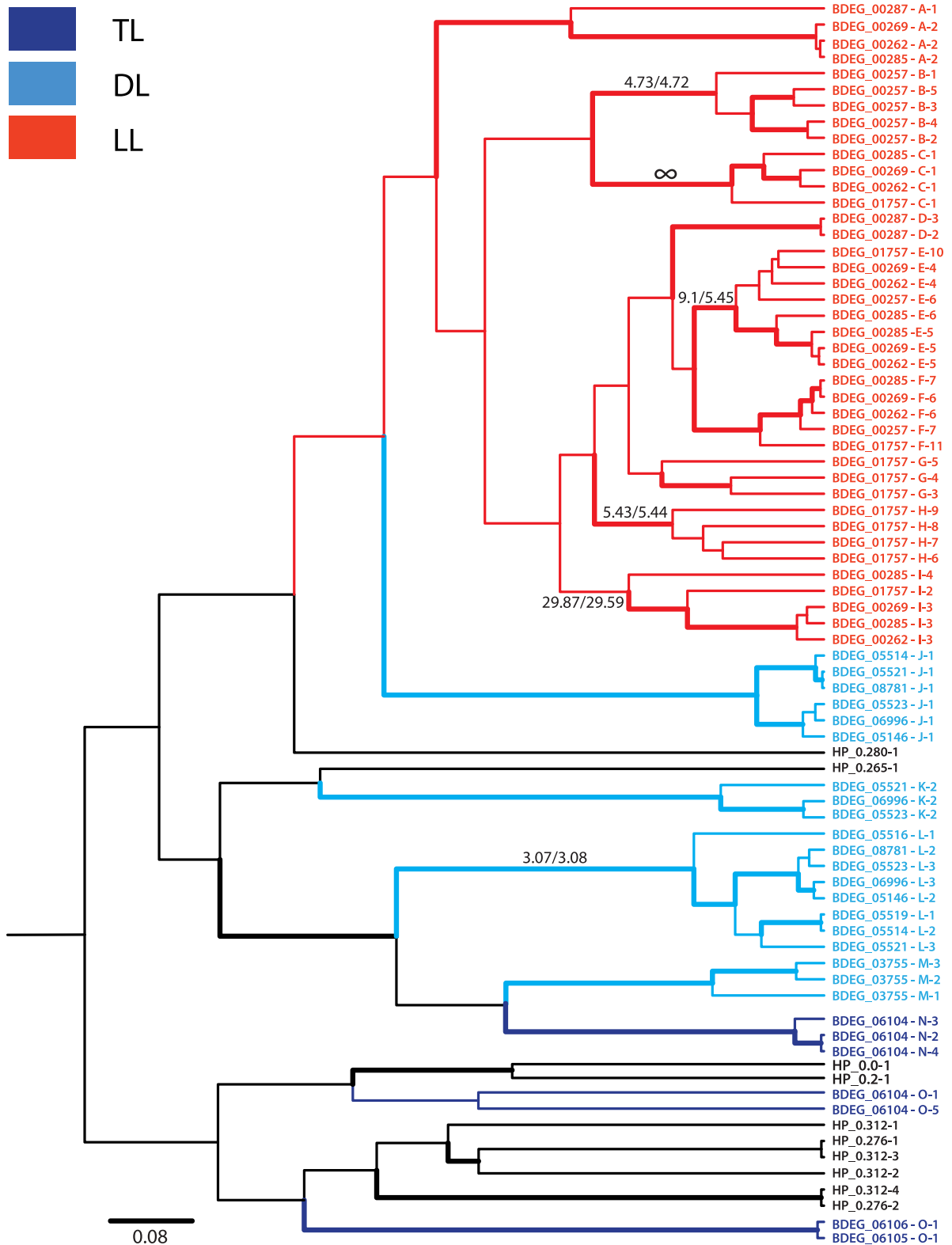
FIG 2 CBM18 domain phylogeny. Unrooted phylogram constructed using only domain sequences and implemented in BEAST v. 1.6.1 (65). *H. polyrhiza* domains were used as an outgroup and are represented in black. Domain names consist of the corresponding gene identifier, the phylogenetic group to which the domain belongs to, and the position in the gene which the specific domain holds. Thicker branches indicate Bayesian posterior probabilities of ≥0.80. ω values are given on branches which show positive selection. Branches are color coded according to gene type. TL, tyrosinase-like; DL, deacetylase-like; LL, lectin-like; HP, *H. polyrhiza*.

TABLE 1 Domain groups with adaptive selection signal

Comparison	Proportion (%)[a]	ω	$2\Delta l$	df	χ^2
Group B					
M1 versus M2	45	4.73	5.27	2	0.036
M7 versus M8	22	4.72	5.39	2	0.034
Group C					
M1 versus M2	4.5	928.63	20.54	2	1.73×10^{-5}
M7 versus M8	4.8	999.0	25.11	2	1.77×10^{-6}
Group E					
M1 versus M2	12	9.10	15.46	2	0.00022
M7 versus M8	29	5.45	15.43	2	0.00022
Group H					
M1 versus M2	37	5.43	13.38	2	0.00062
M7 versus M8	37	5.44	13.46	2	0.00060
Group I					
M1 versus M2	5.2	29.87	10.03	1	0.00083
M7 versus M8	5.2	29.59	10.03	2	0.00331
Group L					
M1 versus M2	25	3.07	6.15	2	0.023
M7 versus M8	24	3.08	7.74	2	0.010

[a] Percentage of sites with positive selections.

The LL genes also have unique patterns of gains and losses between them. In comparison to the DL and TL genes, the LL genes have a larger number of domains on average, ranging from 6 to 11. BDEG_00287 and BDEG_0285 in the LL clade, show a more distinct domain pattern than the rest (Fig. 1; also, see Fig. S2 in the supplemental material). The other four members of the LL clade (BDEG_00262, BDEG_00269, BDEG_00285, and BDEG_01757) show a similar pattern of the domain groups A, C, D, E, F, G, H, and I, with various duplications and losses throughout. BDEG_01757 stands out in all of the genes due to the 11 copies of the CBM18 domains it contains, including the unique G and H domains which, although related to the other LL domains, are unique (Fig. 1).

Test for selection in domains. Due to the disparities we saw between the domain phylogeny and the gene phylogeny, we decided to test for adaptive selection within domain groups. Convergent functional evolution may explain the paraphyletic groups which the DL and TL domains form (Fig. 2). Due to the short sequence lengths (30 to 43 residues), we decided to use the site models to identify specific amino acid sites within the domains potentially under positive selection. Although all of the domains contained characteristic CBMs, they showed a high degree of variability in their coding sequence. Therefore, analyzing the domains according to their monophyletic groups (A to O) gave more precise results than comparing all domains at once. We observed evidence for positive selection in 7 of the 14 groups analyzed.

The domains were analyzed under models M1 and M7 (null neutral models) and M2 and M8 (selection site models) (Table 1) as implemented in PAML 4.4. Of the 14 groups tested (excluding group D, which had only two members), we identified statistically significant adaptive selection in groups B, C, E, H, J, K, and L (Table 1; also, see Table S1 in the supplemental material). Group K was subsequently removed from the analysis due to an insufficient number of group members, which adversely affected the parameters of the analysis. The rejection of the neutral model in all cases was statistically significant, with P being <0.05 for group D and <0.005 for all other domain groups (Table 1; Table S1). In these groups, positively selected sites ranged from only 4.5% of the domains in group F to 37% of those in group J. Likewise, ω values

ranged from a low of 3.07 in group L to ∞ (999.0) in group C, which is indicative of a dS value of ~0 (39). The selection models, M2 and M8, were in consensus with regard to the sites identified as showing strong adaptive selection, with M8 generally having a few additional sights coupled with stronger BEB support (see Table S1).

Test for selection in whole genes. From the previous analysis, we found that the majority of domain groups showing adaptive selection were from the LL clade. This finding, coupled with the fact that this clade contains the largest number of domain copies, led us to test for adaptive selection within the all LL genes. For this analysis, we decided to utilize the branch model as well as the branch site model in addition to the previously described site models, with the DL and TL clades as indicators of background evolutionary rate.

When the branch model was applied to the data set, the one-ratio model (null, neutral) was compared to the two-ratio model (selection). Using a chi-square test, we found that the two-ratio model fits significantly better than the one ratio model (see Table S2 in the supplemental material). The two-ratio model revealed that $\omega 1f$ (foreground) was 0.81273, while $\omega 0b$ (background) was 0.46233. Although the foreground LL clade had almost twice the ω value of the background, there was no indication of strong positive selection (see Table S2).

Given the difference in foreground and background ω scores, we decided to further our analysis by applying the branch-site model to isolate the specific sites which may be under positive selection. When model A was applied, and compared to the null neutral model M1 from the site models, we found a likelihood ratio statistic of 154.132 with a P value of <0.001, allowing us to reject the null neutral model M1. Of the sites, 22.7% showed significant positive selection, with a ω value of 4.612 for the foreground branches. Of the positively selected sites, >22% showed significant statistical support using BEB analysis (see Table S2 in the supplemental material). When model B was compared to M3, the neutral model (M3) was once again rejected with highly significant statistical support. Model B suggests moderate positive selection across all of the lineages ($\omega 1b = 1.232$) with strong adaptive selection in the LL clade, with a ω value of 5.0 at 20.6% of sites.

Lastly, the site models were used to analyze the entire genes, and models M1 and M2 were compared (see Table S2 in the supplemental material). The neutral model (M1) was rejected, although only 12.6% of sites were found to be under selection, with a ω value of 5.624 as identified by M2. Of this group, >15% showed high statistical significance using BEB. When models M7 and M8 were run, model M7 was once again rejected in favor of the selection model M8. Model M8 identified 16% of sites as showing positive selection with a ω value of 4.245; again >15% of these sites were statistically significant using BEB. Between models M2 and M8, model M8 had two additional sites, while the other six sites were shared between the two, with differing posterior probabilities.

DISCUSSION

Of the more than 100,000 species of fungi described to date, only a very small proportion are pathogenic (40). Furthermore, many pathogens have nonpathogenic sister species, a pattern observed frequently in this kingdom. This pattern suggests that pathogenicity arose multiple times in the various fungal groups (15, 41). Due to the diverse genetic backgrounds that pathogens arise from,

genomic modifications to attain pathogenicity are varied. However, one of the commonly identified elements that many fungi evolve en route to pathogenesis is an expansion of CBMs. Across the phylum *Ascomycota*, Soanes et al. (40) identified CBM18s (Pfam ID PF00187) in almost three times as many proteins (on average) in pathogenic lineages as in their nonpathogenic counterparts. Despite the association of CBMs and fungal pathogenesis, the exact roles of these domains are not known in fungi and are likely to be as diverse as the domains are among themselves.

The involvement of CBMs in pathogenesis is particularly intriguing in this study due to the pathogenicity-associated characteristics of the CBM18-containing genes themselves. The phylogenetic analysis of these genes revealed three well-supported clades: DL, TL, and LL. The TL group is also monophyletic with some of the *H. polyrhiza* genes, which is likely indicative of an ancestral association before the TL group gave rise to the LL and DL clades. The former two have secondary domains which have both been implicated as pathogenicity factors (genes essential for successful completion of the pathogenic life cycle but dispensable for saprophytic growth) (31, 40). The LL group contains only genes with CBM domains, which are themselves thought to be involved in pathogenesis. CBM domains do not necessarily function in catalytic activity, but they may be essential in chitin binding and substrate affinity for an enzyme (42). Furthermore, the LL group contains the largest fraction of the identified CBMs, both as a group and as individual genes. Through our analyses, we predicted that the members of the LL group function as lectins, which are commonly known to have multiple CBMs, as opposed to genes with a second type of functional domain, which are generally accompanied by only one CBM (34).

Although a role for the LL molecules in *B. dendrobatidis* has not been described, work with the tomato pathogen *Cladosporium fulvum* has shown that CBMs can bind to the chitin on the cell wall of a fungus, essentially functioning as a competitor of and limiting access for chitinases (38). A CBM-containing gene has been shown to protect fungi from basic plant chitinases through this mechanism (38). It is not unlikely that *B. dendrobatidis* has evolved this CBM expansion to function in defense of its cell wall in a manner similar to that of *Cladosporium*. Alternatively, *B. dendrobatidis* may utilize a chitin-binding ability to bind to nonhost chitinous structures (e.g., discarded insect integuments) in an as-yet-undescribed aspect of its ecology, an association chytrids are known for (5). Multiple CBMs can function synergistically to enhance binding affinity of a lectin for chitin (29, 36). The potential ability of *B. dendrobatidis* to survive within chitinous structures in the environment, without amphibian hosts being present, could have major implications for conservation and eradication efforts.

Such long repeats as we have observed in LL loci are common in proteins that are involved in protein-protein interactions and likely occur through tandem duplications (23, 43). One such example is the cytochrome P450 gene, which has been observed to have 40 to 140 copies in filamentous ascomycetes and basidiomycetes, compared to yeast, which has three (40, 25). However, the 11 copies of CBM18 in BDEG_01757 constitute the largest CBM18 expansion by far of any organism that we are aware of and may have functional properties that will be important to characterize in follow-up molecular genetic explorations. Alternative search approaches, as well as reanalysis with higher *E* values, are likely to discover more CBM18 domains as well as domain-carrying genes. However, these analyses would simply further

support our conclusions of an expansion of this group of domains in this species.

In addition to their domain abundance and diversity, the LL clade also appears to show the strongest signal of adaptive evolution of the three groups. When we tested the domain groups for selection, 5 of the 6 groups identified as being under positive selection were from LL loci. Furthermore, when a *dN/dS* analysis was performed on the whole gene sequence, using the selection levels of DL and TL genes as background rates, LL genes showed a significantly stronger signal for adaptive evolution.

Anisimova et al. (44) discuss the effects of sequence length as well as sample numbers with respect to the power and effectiveness of the likelihood method to detect positive selection in the CBM18 domains. In their analyses, short sequences were more likely to produce a $2\Delta l$ value of 0, similar to sequences with low divergence. Sequence length may also have an effect on the χ^2 approximations; however, Anisimova and colleagues (44) also showed that sequences as short as 50 codons show a normal χ^2 distribution pattern in computer simulations. Our sequences are somewhat shorter than these averaging ~41 codons. However, none of the clades we report in this study as having positive selection display any of the aforementioned effects of short sequences.

Furthermore, Anisimova et al. (44) found that the number of taxa did not affect the shape of the $2\Delta l$ distribution in their simulation studies, while they used as few as five taxa. However, the number of taxa did negatively affect the power of the LRT, which may have had an effect on our analysis of clade K, which has only three members and thus was not reported as a significant result. Furthermore, Yang (45) suggests a minimum of four to five sequences. While some of our clades have less than this recommended number of sequences, most do not. All other clades reported in our analyses meet this criterion.

Additionally, de Jonge and Thomma (46) suggest that CBM-rich genes may be involved in repression of immune response by sequestering any chitin oligosaccharide breakdown products which may be released during invasion, thereby concealing the presence of the pathogen from the host. This may explain the lack of immune response observed by Rosenblum et al. (13).

While the LL loci are CBM rich and appear to be evolving at a higher rate than other loci, the DL and TL loci also reveal an interesting link to pathogenicity. The TL loci encode tyrosinase/catechol oxidase-like domains in addition to CBM18. Genes with putative tyrosinase activity are typically expanded in copy number among pathogenic fungal lineages and are associated with melanin biosynthesis (31). In fungi, melanins are involved in virulence by defending the cell against UV radiation, free radicals, and hydrolytic enzymes which can attack the cell wall and cellular components (47, 48). Melanized cell walls also show an electron-dense outer layer compared to albino mutants in *Wangiella dermatitidis* (49). Melanins also bind metals and function as redox-buffers, acting as a sink for unpaired electrons (50, 51). If such an electron-dense outer layer is formed in *B. dendrobatidis*, it may participate in electron transport disruption, which is thought to be the main cause of host amphibian death from *B. dendrobatidis* infections (12).

Interestingly, CBM18 domains in the TL group showed no statistically significant positive selection, indicating that these loci are most likely under purifying selection, possibly due to a functional link with the tyrosinase domain. However, there does seem to be some indication of convergent evolution, since the domains

from group N, which likely arose from a tandem duplication of group O domains on a TL gene, form a well supported monophyletic clade with group M from the DL clade. The aforementioned scenario results in a TL gene with multiple, distinct domains. A recent study identified three chitin-binding sites in the locus of the archaeon *Thermococcus kodakaraensis* KOD1, where one domain had 1/3 the binding force of the other (52). The diversification in CBM function in this gene may fine-tune this TL gene to a multivalent substrate (29).

The DL clade is comprised of eight genes containing a deacetylase domain, which can be found to play a role in pathogenic lineages. Chitin deacetylases convert chitin into chitosan by a de-N-acetylation process, either changing the chitin polymers of the cell wall directly or providing a chitosan mask on top of it (53, 54). This process is significant because chitosan is a poor substrate for standard antifungal hydrolases such as chitinases and β-1,3-glucanases, which primarily target the outermost layers of the fungal cell wall (53). There is sparse literature on skin-secreted chitinases in amphibians, but amphibians do produce chitinases in the digestive system and pancreas and may form a symbiotic relationship with bacteria that produce these enzymes on their skin in order to fight off fungal infection (55, 56). Additionally, amphibian skin secretions have been shown to affect *B. dendrobatidis* survival directly (57). Of the 880 antimicrobial peptides identified in anurans to date, >20% are produced within special glands in their skin (58). Therefore, the expansion of a chitin deacetylase gene may represent evolved defensive measures that protect *B. dendrobatidis* from host-derived antimicrobial peptides which attack chitin (31).

In conclusion, we have identified a unique expansion of the CBM18 domain in *B. dendrobatidis* and evidence that positive selection has shaped the recent molecular evolution of these loci. Although these genes are putative with regard to their functional roles, their structures strongly indicate roles in pathogenesis and defense, two critical factors shaping genetic modifications in the *de novo* evolution of pathogenicity. Further molecular studies are required to show that the expression and function of these genes in the life cycle of *B. dendrobatidis* are relevant to defense, but bioinformatic analyses indicate a strong connection of the expansion of this gene family and the evolution of the pathogenic lifestyle of this unique fungus.

MATERIALS AND METHODS

Predicted coding sequences (CDS) and their protein translations were obtained for *B. dendrobatidis* isolate JEL423 (accession no. AATT00000000) and *H. polyrhiza* isolate JEL142 (accession no. AFSM00000000) (http://fungalgenomes.org/public/Hp_JEL142/) (59). The CBM18 protein domain family (Chitin_bind_1) hidden Markov model (HMM) representation was downloaded from the Sanger Pfam24.0 database (60). The HMMER3 (61) application hmmsearch was used to perform a search of the CBM18 Pfam HMM against the *H. polyrhiza* and *B. dendrobatidis* protein databases with an expectation value (*E* value) of 0.01. The domains were identified and then extracted from the database using a custom Perl script utilizing BioPerl (62). Domain positions in the genes were identified manually.

To align the individual domains, the 67 *B. dendrobatidis* and 10 *H. polyrhiza* domain sequences identified by HMMER were extracted and aligned with T-Coffee (63). The resulting multiple protein sequence alignment was back translated based on the

known CDS to create a codon alignment guided by the amino acid alignment using the BioPerl script (https://github.com/bioperl/bioperl-live/blob/master/scripts/utilities/bp_mrtrans.pl). This codon alignment was then used for molecular evolution and phylogenetic analyses of the domains. However, T-Coffee proved to be insufficient for the alignment of the entire locus due to variation in domain numbers, domain positions, and overall gene lengths. Instead, we used an anchored multiple alignment method implemented in the DIALIGN alignment tool (64) on the input CDS gene sequences (the coding regions are similar enough at the DNA level). This tool allows for user-specified anchor points to constrain the alignment to prioritize the alignment of the user-selected sequences (e.g., conserved domains) while the remaining parts of the sequence are aligned automatically. This method allowed us to create biologically meaningful alignments based on the *a priori* knowledge of homologous regions the alignment algorithms may not automatically recognize.

Phylogenetic analyses of the domain and gene CDS alignments were performed using a Bayesian Markov chain Monte Carlo (MCMC) analysis implemented in BEAST v. 1.6.1 (65). The program uses an MCMC method to sample tree and parameter space and can be summarized to generate a consensus tree with posterior probabilities for each clade representing confidence in the presence of the clade from the sampled phylogenies. The analysis was run with a minimum of 10 million generations and sampling every 100th tree. Subsequently, a consensus tree was generated using TreeAnnotator v.1.6.1 (65) and viewed using FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). Additionally, the PHOG (PhyloFacts Orthology Group) Universal Proteome Explorer v.2.0 was used to annotate functionally and structurally important regions in the protein sequences in order to predict functional roles for proteins (66).

Test for selection. To test for selection both within lineages as well as within specific amino acid sites, we compared rates of synonymous (silent; *dS*) and nonsynonymous (amino acid replacement; *dN*) substitutions in the CBM18 domains. The *dN/dS* ratio (ω) measures selective pressure on amino acids. Thus, if *dN* significantly exceeds *dS*, divergence is likely to have occurred through positive selection. In other words, under selectively neutral conditions or even purifying selection, ω is less than or equal to 1, whereas ω will exceed 1 if an amino acid change is selectively advantageous. Finding that ω is significantly greater than 1 shows convincing evidence of adaptive selection (67).

To test for positive selection, we use a tree-based likelihood approach as described by Yang (68). Bayesian phylogenetic trees were constructed, as previously described, for each data set. Codon alignment files as well as their corresponding Bayesian tree files were submitted to the CODEML program in PAML v. 4.4 (45). CODEML was used to analyze the data under three different models and parameters.

The branch model is used to estimate positive selection along a specific lineage (54). Specifically, it allows the *dN/dS* ratio on a specified branch of the tree to differ from the average *dN/dS* ratio across the rest of the tree (two-ratio model). The null model fixes *dN/dS* at 1 on the specified branch. The one-ratio model is neutral and assumes the same ω ratio for all branches. The two-ratio model allows for two ω rates (69). Likelihood estimates assumed the codon substitution model of Goldman and Yang (70).

The site models allows the ω ratio to vary between codons (71). The null, neutral models do not allow for positive selection ($\omega \leq 1$) and are compared to an alternative hypothesis which does allow $\omega > 1$. The first set of models consists of M1 and M2. M1 is a nearly neutral while M2 allows for selection, indicative of adaptive selection. Model M3 also allows for positive selection. Models M7 and M8 are two other models. M7 does

not allow any positively selected sites ($\omega > 1$), whereas M8 allows for an $\omega > 1$ category of sites. M8 is less conservative and may identify positive selection in sites that the M2 analysis did not detect (72).

The branch-site model allows for the detection of positively selected sites on preselected lineages (45, 69). This model requires us to label specific clades or branches as a "foreground" group, which is being tested for selection in comparison to the "background" (all other branches in the tree). There are two models in the branch-site analysis: model A, which assumes that ω_0 is 0 and ω_1 is 1, and model B, which allows ω_0 and ω_1 to vary. Models A and B are identical to models M1 and M3, respectively, from the site model analysis (which is discussed above), except that the site models M1 and M3 assume ω_0 and ω_1 are the same across all branches of the phylogeny, since the site model does not use a labeled tree. If ω_2 is >1, and if model A or B fits significantly better than model M1 or M3, then this is indicative of positive selection in the foreground branches. Confidence of specific evolutionary rate classes of sites is computed with a Bayes empirical Bayes (BEB) approach (73). Yang (45) describes these three models in greater detail. The likelihood estimates for each were compared using a hierarchical likelihood ratio test (hLRT) of twice the difference in the log likelihood value of the models being compared ($2\Delta l$), with the result approximating a chi square distribution.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00150-12/-/DCSupplemental.

FIG S1, DOCX file, 0.1 MB.
FIG S2, PDF file, 0.2 MB.
FIG S3, PDF file, 0.2 MB.
TABLE S1, DOCX file, 0.1 MB.
TABLE S2, DOCX file, 0.1 MB.
TABLE S3, DOCX file, 0.1 MB.
TABLE S4, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. **Barnosky AD, et al.** 2011. Has the Earth's sixth mass extinction already arrived? Nature **471**:51–57.
2. **Cheng TL, Rovito SM, Wake DB, Vredenburg VT.** Coincident mass extirpation of neotropical amphibians with the emergence of the infectious fungal pathogen *Batrachochytrium dendrobatidis*. Proc. Natl. Acad. Sci. U. S. A., in press.
3. **Stuart SN, et al.** 2004. Status and trends of amphibian declines and extinctions worldwide. Science **306**:1783–1786.
4. **James TY, et al.** 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. Nature **443**:818–822.
5. **Sparrow FK.** 1960. Aquatic phycomycetes, 2nd ed, vol **XV**. University of Michigan Press, Ann Arbor.
6. **Fisher MC, Garner TW, Walker SF.** 2009. Global emergence of *Batrachochytrium dendrobatidis* and amphibian chytridiomycosis in space, time, and host. Annu. Rev. Microbiol. **63**:291–310.
7. **Gleason FH, Kagami M, Lefevre E, Sime-Ngando T.** 2007. The ecology of chytrids in aquatic ecosystems: roles in food web dynamics. Fungal Biol. Rev. **22**:17–25.
8. **Lazarus LH, Attila M.** 1993. The toad, ugly and venomous, wears yet a precious jewel in his skin. Prog. Neurobiol. **41**:473–507.
9. **Longcore JE, Pessier AP, Nichols DK.** 1999. *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. Mycologia **91**:219–227.
10. **Voyles J, et al.** 2007. Electrolyte depletion and osmotic imbalance in amphibians with chytridiomycosis. Dis. Aquat. Organ. **77**:113–118.
11. **Voyles J, Rosenblum EB, Berger L.** 2011. Interactions between *Batrachochytrium dendrobatidis* and its amphibian hosts: a review of pathogenesis and immunity. Microbes Infect. **13**:25–32.
12. **Voyles J, et al.** 2009. Pathogenesis of chytridiomycosis, a cause of catastrophic amphibian declines. Science **326**:582–585.
13. **Rosenblum EB, et al.** 2009. Genome-wide transcriptional response of *Silurana* (*Xenopus*) *tropicalis* to infection with the deadly chytrid fungus. PLoS One **4**:e6494.
14. **Rosenblum EB, et al.** 2010. A molecular perspective: biology of the emerging pathogen *Batrachochytrium dendrobatidis*. Dis. Aquat. Organ. **92**:131–147.
15. **Soanes DM, Richards TA, Talbot NJ.** 2007. Insights from sequencing fungal and oomycete genomes: what can we learn about plant disease and the evolution of pathogenicity? Plant Cell **19**:3318–3326.
16. **Yoder OC, Turgeon BG.** 2001. Fungal genomics and pathogenicity. Curr. Opin. Plant Biol. **4**:315–321.
17. **Akagi Y, Akamatsu H, Otani H, Kodama M.** 2009. Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. Eukaryot. Cell **8**:1732–1738.
18. **Bastidas RJ, Heitman J.** 2009. Trimorphic stepping stones pave the way to fungal virulence. Proc. Natl. Acad. Sci. U. S. A. **106**:351–352.
19. **Butler G, et al.** 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature **459**:657–662.
20. **Wang S, Leclerque A, Pava-Ripoll M, Fang W, St. Leger RJ.** 2009. Comparative genomics using microarrays reveals divergence and loss of virulence-associated genes in host-specific strains of the insect pathogen *Metarhizium anisopliae*. Eukaryot. Cell **8**:888–898.
21. **Moran GP, Coleman DC, Sullivan DJ.** 2011. Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. Eukaryot. Cell **10**:34–42.
22. **Wapinski I, Pfeffer A, Friedman N, Regev A.** 2007. Natural history and evolutionary principles of gene duplication in fungi. Nature **449**:54–61.
23. **Björklund AK, Ekman D, Elofsson A.** 2006. Expansion of protein domain repeats. PLoS Comput. Biol. **2**:e114.
24. **Lespinet O, Wolf YI, Koonin EV, Aravind L.** 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. **12**:1048–1059.
25. **Stajich JE, et al.** 2010. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). Proc. Natl. Acad. Sci. U. S. A. **107**:11889–11894.
26. **Ohno S.** 1970. Evolution by gene duplication. Springer-Verlag, Berlin.
27. **Prince VE, Pickett FB.** 2002. Splitting pairs: the diverging fates of duplicated genes. Nat. Rev. Genet. **3**:827–837.
28. **Zhang J.** 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. **19**:292–298.
29. **Abbott DW, Eirín-López JM, Boraston AB.** 2008. Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. Mol. Biol. Evol. **25**:155–167.
30. **Sharpton TJ, et al.** 2009. Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. Genome Res. **19**:1722–1731.
31. **Powell AJ, Conant GC, Brown DE, Carbone I, Dean RA.** 2008. Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. BMC Genomics **9**:147.
32. **Suetake T, et al.** 2000. Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. J. Biol. Chem. **275**:17929–17932.
33. **Archer BL.** 1960. The proteins of *Hevea brasiliensis* latex. 4. Isolation and characterization of crystalline hevein. Biochem. J. **75**:236–240.
34. **Raikhel NV, Lee HI, Broekaert WF.** 1993. Structure and function of chitin-binding proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. **44**:591–615.
35. **Li SS, Claeson P.** 2003. Cys/Gly-rich proteins with a putative single chitin-binding domain from oat (*Avena sativa*) seeds. Phytochemistry **63**:249–255.
36. **Beintema JJ.** 1994. Structural features of plant chitinases and chitin-binding proteins. FEBS Lett. **350**:159–163.
37. **Dean RA, et al.** 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature **434**:980–986.
38. **van den Burg HA, Harrison SJ, Joosten MH, Vervoort J, de Wit PJ.** 2006. *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. Mol. Plant Microbe Interact. **19**:1420–1430.

39. **Yang Z, dos Reis M.** 2011. Statistical properties of the branch-site test of positive selection. Mol. Biol. Evol. **28:**1217–1228.

40. **Soanes DM, et al.** 2008. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One **3:**e2300.

41. **Bowman BH, Taylor JW, White TJ.** 1992. Molecular evolution of the fungi: human pathogens. Mol. Biol. Evol. **9:**893–904.

42. **Iseli B, Boller T, Neuhaus JM.** 1993. The N-terminal cysteine-rich domain of tobacco class I chitinase is essential for chitin binding but not for catalytic or antifungal activity. Plant Physiol. **103:**221–226.

43. **Ekman D, Light S, Björklund AK, Elofsson A.** 2006. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? Genome Biol. **7:**R45.

44. **Anisimova M, Bielawski JP, Yang Z.** 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. **18:**1585–1592.

45. **Yang Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24:**1586–1591.

46. **de Jonge R, Thomma BP.** 2009. Fungal LysM effectors: extinguishers of host immunity? Trends Microbiol. **17:**151–157.

47. **Rappleye CA, Goldman WE.** 2006. Defining virulence genes in the dimorphic fungi. Annu. Rev. Microbiol. **60:**281–303.

48. **Sexton AC, Howlett BJ.** 2006. Parallels in fungal pathogenesis on plant and animal hosts. Eukaryot. Cell **5:**1941–1949.

49. **Jacobson ES.** 2000. Pathogenic roles for fungal melanins. Clin. Microbiol. Rev. **13:**708–717.

50. **Butler MJ, Day AW, Henson JM, Money NP.** 2001. Pathogenic properties of fungal melanins. Mycologia **93:**1–8.

51. **Langfelder K, Streibel M, Jahn B, Haase G, Brakhage AA.** 2003. Biosynthesis of fungal melanins and their importance for human pathogenic fungi. Fungal Genet. Biol. **38:**143–158.

52. **Kikkawa Y, et al.** 2008. Interaction force of chitin-binding domains onto chitin surface. Biomacromolecules **9:**2126–2131.

53. **El Gueddari NE, Rauchhaus U, Moerschbacher BM, Deising HB.** 2002. Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. New Phytol. **156:**103–112.

54. **Ride JP, Barber MS.** 1990. Purification and characterization of multiple forms of endochitinase from wheat leaves. Plant Sci. **71:**185–197.

55. **Brucker RM, et al.** 2008. Amphibian chemical defense: antifungal metabolites of the microsymbiont *Janthinobacterium lividum* on the salamander *Plethodon cinereus*. J. Chem. Ecol. **34:**1422–1429.

56. **Harris RN, James TY, Lauer A, Simon MA, Patel A.** 2006. Amphibian pathogen *Batrachochytrium dendrobatidis* is inhibited by the cutaneous bacteria of amphibian species. Ecohealth **3:**53–56.

57. **Woodhams DC, et al.** 2007. Resistance to chytridiomycosis varies among amphibian species and is correlated with skin peptide defenses. Anim. Conserv. **10:**409–417.

58. **VanCompernolle SE, et al.** 2005. Antimicrobial peptides from amphibian skin potently inhibit human immunodeficiency virus infection and transfer of virus from dendritic cells to T cells. J. Virol. **79:**11598–11606.

59. **Joneson S, Stajich JE, Shiu SH, Rosenblum EB.** 2011. Genomic transition to pathogenicity in chytrid fungi. PLoS Pathog. **7:**e1002338.

60. **Bateman A, et al.** 2002. The Pfam protein families database. Nucleic Acids Res. **30:**276–280.

61. **Eddy SR.** 2009. A new generation of homology search tools based on probabilistic inference. Genome Inform. **23:**205–211.

62. **Stajich JE, et al.** 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. **12:**1611–1618.

63. **Notredame C, Higgins DG, Heringa J.** 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302:**205–217.

64. **Morgenstern B, Prohaska SJ, Pöhler D, Stadler PF.** 2006. Multiple sequence alignment with user-defined anchor points. Algorithms Mol. Biol. **1:**6.

65. **Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:**214.

66. **Krishnamurthy N, Brown DP, Kirshner D, Sjölander K.** 2006. PhyloFacts: an online structural phylogenomic encyclopedia for protein functional and structural classification. Genome Biol. **7:**R83.

67. **Hill RE, Hastie ND.** 1987. Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature **326:**96–99.

68. **Yang Z.** 1998. On the best evolutionary rate for phylogenetic analysis. Syst. Biol. **47:**125–133.

69. **Mathews S, Burleigh JG, Donoghue MJ.** 2003. Adaptive evolution in the photosensory domain of phytochrome A in early angiosperms. Mol. Biol. Evol. **20:**1087–1097.

70. **Goldman N, Yang Z.** 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11:**725–736.

71. **Nielsen R, Yang Z.** 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:**929–936.

72. **Metzger KJ, Thomas MA.** 2010. Evidence of positive selection at codon sites localized in extracellular domains of mammalian CC motif chemokine receptor proteins. BMC Evol. Biol. **10:**139.

73. **Yang Z, Rannala B.** 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. **54:**455–470.