Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

Comparative Genome Analysis of *Trichophyton rubrum* and Related Dermatophytes Reveals Candidate Genes Involved in Infection

Diego A. Martinez, Brian G. Oliver, Yvonne Gräser, et al. 2012. Comparative Genome Analysis of *Trichophyton rubrum* and Related Dermatophytes Reveals Candidate Genes Involved in Infection . mBio 3(5): . doi:10.1128/mBio.00259-12.

Updated information and services can be found at:
http://mbio.asm.org/content/3/5/e00259-12.full.html

SUPPLEMENTAL MATERIAL **http://mbio.asm.org/content/3/5/e00259-12.full.html#SUPPLEMENTAL**

REFERENCES This article cites 73 articles, 30 of which can be accessed free at:
http://mbio.asm.org/content/3/5/e00259-12.full.html#ref-list-1

CONTENT ALERTS Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), **more>>**

Information about commercial reprint orders: **http://mbio.asm.org/misc/reprints.xhtml**
Information about Print on Demand and other content delivery options:
http://mbio.asm.org/misc/contentdelivery.xhtml
To subscribe to another ASM Journal go to: **http://journals.asm.org/subscriptions/**

Journals.ASM.org

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

RESEARCH ARTICLE

Comparative Genome Analysis of *Trichophyton rubrum* and Related Dermatophytes Reveals Candidate Genes Involved in Infection

Diego A. Martinez,[a] Brian G. Oliver,[b] Yvonne Gräser,[c] Jonathan M. Goldberg,[a] Wenjun Li,[d] Nilce M. Martinez-Rossi,[e] Michel Monod,[f] Ekaterina Shelest,[g] Richard C. Barton,[h] Elizabeth Birch,[i] Axel A. Brakhage,[g] Zehua Chen,[a] Sarah J. Gurr,[i] David Heiman,[a] Joseph Heitman,[d] Idit Kosti,[j] Antonio Rossi,[e] Sakina Saif,[a] Marketa Samalova,[i] Charles W. Saunders,[k] Terrance Shea,[a] Richard C. Summerbell,[l] Jun Xu,[k] Sarah Young,[a] Qiandong Zeng,[a] Bruce W. Birren,[a] Christina A. Cuomo,[a] and Theodore C. White[m]

Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA[a]; Seattle Biomedical Research Institute, Seattle, Washington, USA[b]; Universitätsmedizin Berlin—Charité, Institute of Microbiology and Hygiene, Berlin, Germany[c]; Duke University, Durham, North Carolina, USA[d]; Universidade de São Paulo, Ribeirão Preto, Sao Paulo, Brazil[e]; Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland[f]; Leibniz Institute for Natural Product Research and Infection Biology (HKI), Jena, Germany[g]; Leeds General Infirmary, Leeds, United Kingdom[h]; University of Oxford, Oxford, England, United Kingdom[i]; Department of Biology, Technion-IIT, Haifa, Israel[j]; Procter and Gamble Co., Cincinnati, Ohio, USA[k]; Sporometrics and Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada[l]; and University of Missouri at Kansas City, Kansas City, Missouri, USA[m]

ABSTRACT **The major cause of athlete's foot is *Trichophyton rubrum*, a dermatophyte or fungal pathogen of human skin. To facilitate molecular analyses of the dermatophytes, we sequenced *T. rubrum* and four related species, *Trichophyton tonsurans*, *Trichophyton equinum*, *Microsporum canis*, and *Microsporum gypseum*. These species differ in host range, mating, and disease progression. The dermatophyte genomes are highly colinear yet contain gene family expansions not found in other human-associated fungi. Dermatophyte genomes are enriched for gene families containing the LysM domain, which binds chitin and potentially related carbohydrates. These LysM domains differ in sequence from those in other species in regions of the peptide that could affect substrate binding. The dermatophytes also encode novel sets of fungus-specific kinases with unknown specificity, including nonfunctional pseudokinases, which may inhibit phosphorylation by competing for kinase sites within substrates, acting as allosteric effectors, or acting as scaffolds for signaling. The dermatophytes are also enriched for a large number of enzymes that synthesize secondary metabolites, including dermatophyte-specific genes that could synthesize novel compounds. Finally, dermatophytes are enriched in several classes of proteases that are necessary for fungal growth and nutrient acquisition on keratinized tissues. Despite differences in mating ability, genes involved in mating and meiosis are conserved across species, suggesting the possibility of cryptic mating in species where it has not been previously detected. These genome analyses identify gene families that are important to our understanding of how dermatophytes cause chronic infections, how they interact with epithelial cells, and how they respond to the host immune response.**

IMPORTANCE **Athlete's foot, jock itch, ringworm, and nail infections are common fungal infections, all caused by fungi known as dermatophytes (fungi that infect skin). This report presents the genome sequences of *Trichophyton rubrum*, the most frequent cause of athlete's foot, as well as four other common dermatophytes. Dermatophyte genomes are enriched for four gene classes that may contribute to the ability of these fungi to cause disease. These include (i) proteases secreted to degrade skin; (ii) kinases, including pseudokinases, that are involved in signaling necessary for adapting to skin; (iii) secondary metabolites, compounds that act as toxins or signals in the interactions between fungus and host; and (iv) a class of proteins (LysM) that appear to bind and mask cell wall components and carbohydrates, thus avoiding the host's immune response to the fungi. These genome sequences provide a strong foundation for future work in understanding how dermatophytes cause disease.**

Received 30 July 2012 **Accepted** 31 July 2012 **Published** 4 September 2012

Citation Martinez DA, et al. 2012. Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. mBio 3(5): e00259-12. doi:10.1128/mBio.00259-12.

Editor Judith Berman, University of Minnesota, GCD

Copyright © 2012 Martinez et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Christina A. Cuomo (issues relating to sequence, annotation, data release, and comparative genomic analysis), cuomo@broadinstitute.org, or Theodore C. White (issues relating to species, strains, biology, and pathogenesis), whitetc@umkc.edu.

Christina A. Cuomo and Theodore C. White are co-senior authors who contributed equally to this work.

Dermatophytes are a monophyletic group of fungi best known for affecting the skin of animals and humans. These fungi cause a variety of skin diseases, including athlete's foot (clinically termed tinea pedis), jock itch (tinea cruris), and ringworm (tinea capitis or tinea corporis, depending on area of the body infected). There are significant geographic patterns to infection. For example, tinea pedis is more common in developed countries, while tinea capitis is more common in developing countries. Dermatophytes cause some of the most common fungal infections in the world and are endemic to all continents, excluding Antarctica. While these infections rarely cause death, they are difficult to treat and contribute to morbidity, pain, and suffering, especially in ag-

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

TABLE 1 Dermatophyte genome statistics

Characteristic	Result for:						
	T. rubrum	*T. tonsurans*	*T. equinum*	*M. canis*	*M. gypseum*	*T. verrucosum*	*A. benhamiae*
CBS no.	118892	112818	127.97	113480	118893	44[a]	112371
Isolation location	Germany	Quebec	Finland	Germany	Germany	Germany	Switzerland
Site of infection	Nail	Cheek	Nail	Head	Skin	Genitals	Face
Yr of isolation	2004	2003	1996	2004	2004		2002
Mating type	*MAT1-1*	*MAT1-1*	*MAT1-2*	*MAT1-1*	*MAT1-1*	*MAT1-1*	*MAT1-1*
Gene ID prefix	TERG	TESG	TEQG	MCYG	MGYG	TRV	ARB
Assembly size (Mb)	22.5	23.0	24.1	23.1	23.2	22.5	22.2
Repeat (%)	1.73	3.70	7.20	2.43	1.54	2.13	1.34
GC (%)	48.31	48.15	47.39	47.5	48.5	48.24	48.75
No. of predicted protein-coding genes	8,707	8,523	8,679	8,915	8,907	8,024	7,980
Mean coding sequence length (nt)	1,393	1,409	1,371	1,459	1,436	1,468	1,483
Mean intron length (nt)	76	89	96	80	85	84	83
Mean exon no. per gene	3.1	3.2	3.2	3.21	3.15	2.6	2.6
Mean intergenic length (nt)	995	1,078	1,174	970	990	1,136	1,160
No. of tRNAs	82	82	85	82	83	81	85
No. of transmembrane proteins	1,601	1,532	1,531	1,605	1,636	1,640	1,679
No. of secreted proteins	545	526	523	546	589	428	469
No. of GPI-anchored proteins	40	35	42	40	43	26	30

[a] *T. verrucosum* strain is designated 44 but is not currently deposited at CBS.

ing populations (such as the United States). The impact to our economy is striking, as over 500 million dollars is spent on the treatment of dermatophytes worldwide every year (1, 2).

Dermatophytes are in the family *Arthrodermataceae*, a group of filamentous fungi closely related to dimorphic fungi in the *Onygenales*, most closely to the genus *Coccidioides*, a human pulmonary pathogen. While three dermatophyte genera (*Trichophyton*, *Microsporum*, and *Epidermophyton*) have been described, the phylogenetic relationships of the dermatophyte species do not support three distinct divisions, suggesting that these genera are not monophyletic (3). For example, some *Trichophyton* species are more distantly related to *T. rubrum*, the major cause of athlete's foot, than some *Microsporum* species. Among the dermatophytes, there are at least 40 species that are known to cause disease in humans (3, 4).

To examine the genomic basis of phenotypic variation, we sequenced five species selected to represent differences in virulence, host range, mating ability, and frequency of infection (4). This included *Trichophyton rubrum*, a species found only on humans (an anthropophile) and the most common cause of tinea pedis. For comparison, we chose species phylogenetically related to *T. rubrum* that infect humans and animals. *Trichophyton tonsurans* is an anthropophile and a common cause of tinea capitis in humans. It is closely related to *Trichophyton equinum*, a zoophile found on horses and other animals. In addition, we sequenced two more distantly related species, *Microsporum canis*, a common zoophile, and *Microsporum gypseum*, a geophile commonly found in soil. We included in our comparison the genomes of two dermatophytes more closely related to *T. rubrum* (5): *Arthroderma benhamiae*, the teleomorph of a group of *Trichophyton* sp. and a zoophile found in rats, causing human infections, and *Trichophyton verrucosum*, a common problem in cattle as well as humans. Only a subset of these seven species has been described as sexually competent, including *A. benhamiae*, *M. canis*, and *M. gypseum* (6). However, it is possible that the other species, including *T. rubrum*, carry out more cryptic sexual cycles that have yet to be detected. Strains selected for sequencing (Table 1) were recent clinical isolates, were confirmed for species by standard and molecular identification, and show typical growth, morphology, and drug susceptibilities (4, 7).

Here, we present a comparative genomic analysis of dermatophytes, which reveals features that differentiate this group from other fungi and from each other. We describe the genome sequences of *T. rubrum* and four additional dermatophytes that include the most prevalent species in human infections and compare these to two previously sequenced species. This analysis identified changes in specific functional categories common to all dermatophyte genomes. In addition, recent gene gain and loss events within the dermatophytes suggest candidates for roles in infection and host specialization, which could help guide the development of new therapies. These species-specific adaptations could be important in host immune system interaction or survival in the environment.

RESULTS

Genome sequence and analysis. The genomes of the dermatophytes are similar in size, ranging from 22.5 Mb for *T. rubrum* to 24.1 Mb for *T. equinum*. These genomes are smaller in size than other sequenced *Onygenales*; the most closely related *Coccidioides* genomes average 27.5 Mb and other *Onygenales* are larger (8). The three genomes sequenced to higher sequence depth (8 to 9 times) (Table 1; see also Table S1 in the supplemental material) contain nearly all of the assembled sequence in a small number of scaffolds; 95% of the assembly is represented by eight scaffolds in *M. canis*, nine scaffolds in *M. gypseum*, and 12 scaffolds in *T. rubrum* (see Table S1). Some of these scaffolds likely represent complete chromosomes, and telomeric sequences can be found at the ends of, or linked to the ends of, six to eight scaffolds in these assemblies, although two to four of these are at the ends of small scaffolds. The dermatophyte genomes show a very low rate of predicted single nucleotide variation that may represent background noise (see Materials and Methods). The lack of clear support for polymorphism within species is consistent with the predicted haploid nature of these species and the use of single

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

germinated conidia as an inoculum to grow the species for DNA extraction.

The dermatophyte genomes have few transposable elements (TEs), ranging from 1.3% to 7.2% of each assembly (see Table S2 in the supplemental material). These TEs have a lower GC content than the genome average; the GC content of the repetitive sequence ranges from 31 to 32% in the three *Trichophyton* spp. to 36% in *M. gypseum* and *M. canis*. The most repetitive genomes with the highest TE content, *T. equinum* and *T. tonsurans*, are closely related, but *T. equinum* contains nearly twice as much repeat content as *T. tonsurans*. The difference in repeat content can be accounted for by expansion of particular transposable element classes, of which some appear specific to a subset of the dermatophyte genomes. Gypsy elements are most frequently found and are expanded particularly in *T. equinum*, which contains nearly twice as many gypsy elements as *T. tonsurans* (see Table S2). The helitron family is most frequently found in *T. verrucosum*, while mariner elements are highly represented in both *T. equinum* and *T. tonsurans* compared to in the other species. The non-LTR/LINE family is also expanded in *T. equinum* and *T. tonsurans* and contributes to the higher repeat content of the *T. equinum* genome. The low levels of transposable elements suggest that genome defense mechanisms are active. We identified homologs of Argonaut and dicer in all dermatophytes, which could constrain transposable element spread in these genomes. The presence of these genes also suggests that RNA interference (RNAi) gene knockdowns could be used for functional analysis. Additionally, we assembled the rRNA repeat unit of *T. rubrum* (see Text S1 in the supplemental material) and analyzed the 5S rRNA repeats in *M. canis* (see Text S1), providing a basis for strain typing and population analysis in dermatophytes.

Genome similarity and synteny. The dermatophyte genomes display a high degree of colinearity interrupted by a small number of inversions (see Fig. S1 in the supplemental material) and a high percent identity at the amino acid level. The average amino acid identity between the proteins of *T. rubrum* compared to that of each of the other dermatophytes (pairs identified by BLASTP, requiring a score of >100) ranged from a low of 79% for *M. canis* and *M. gypseum* (covering 89% of *T. rubrum* proteins) to a high of 97% for *T. verrucosum* (covering 85% of *T. rubrum* proteins).

The amount of each genome represented in syntenic regions ranges from 92% of *T. tonsurans* (compared to *T. equinum*) to 65% of *M. canis* (compared to *T. verrucosum*) (see Fig. S1 in the supplemental material). In comparison to *T. rubrum*, the closely related genome of *A. benhamiae* has undergone only two small inversions. In contrast, *M. canis*, the most distantly related dermatophyte to *T. rubrum* in our analysis, displays 10 instances of inversions (see Fig. S1). As might be expected, the two most closely related species, *T. equinum* and *T. tonsurans*, appear to have no inversions between them and contain the highest percentage of genes that are colinear (90% and 92%, respectively) (data not shown). Comparison of the dermatophytes to the most closely related outgroup species, *Coccidioides immitis*, reveals a higher degree of divergence and rearrangement. Numerous inversions have occurred since the common ancestor of *C. immitis* and the dermatophytes, and colinear regions cover at most 43% of the *C. immitis* genome (see Fig. S1).

Gene conservation and species specificity. The dermatophytes are strikingly similar in gene contents, with a core set of 6,168 orthologous groups common to the seven dermatophyte genomes (Fig. 1A and C). The total number of orthologs found in all dermatophytes ranges between 69% (in the largest genome, *M. canis*) to 77% (in the smallest genome, *A. benhamiae*). Of the dermatophyte core orthologous groups, 308 are completely unique to the dermatophytes, shown by the red bars in Fig. 1A. Despite their conservation, 60% of these 308 genes contain no recognizable Interpro (IPR) domains (9). Of the remaining 40% with IPR domains, many include functional categories that were also identified as significantly enriched in dermatophytes (see the following section).

T. rubrum, *M. gypseum*, and *M. canis* contain the largest numbers of unique genes (801, 937, and 943 genes, respectively), while *T. tonsurans* contains the fewest with 340, likely due to the high sequence similarity with *T. equinum* (Fig. 1A and B). The genes unique to each species contain little predicted functional information; IPR domains are found in as little as 4% of *T. rubrum* unique proteins and up to 17% for *M. canis*. Additionally few of these proteins contain predicted targeting signals; an average of less than 5% of the species-specific genes contain a secretion signal, and therefore predicting the function of these genes will require experimental studies. In Fig. 1C, the number of genes in parentheses next to the species is the number of genes completely unique to that species; these numbers indicate that species-specific adaptations are ongoing and may be more important than what is shared between species with the same host range.

A high proportion of ortholog groups are shared among all dermatophytes, while a few are specific to dermatophytes that inhabit specific ecological niches or host ranges. To identify genes that are involved in the preference of each niche or host (anthropophile, zoophile, or geophile), we examined ortholog conservation based on predominant lifestyle in dermatophytes (Fig. 1C). Few ortholog groups are unique to each host range (animal, human, or soil). Anthropophile-specific ortholog groups had few genes identifiable by Interpro, as only eight of the 48 ortholog groups contained IPR domains; these included three groups of transcription factors, two of transposable elements, and one each of secondary metabolism, protein-protein interactions, and nuclear transport. A larger set of 271 orthogroups were specific to two or more zoophiles (Fig. 1C); 82% of these groups lack an IPR domain, and the only IPR categories found in multiple orthogroups were domains involved in signal transduction, secondary metabolism, and one transporter.

Examining the genes specific to animal-associated dermatophytes identified a total of 988 ortholog groups that contain at least one anthropophilic and one zoophilic dermatophyte and not found in the geophile *M. gypseum* (Fig. 1C). Within this group there were only 209 ortholog groups that contain at least one IPR domain (79% of ortholog groups were thus unknown). We further analyzed the IPR categories that could be found in at least three ortholog groups (the top 15 IPR domain categories). The most common functional classes included kinase domains (33 ortholog groups); secondary-metabolism domains (38 total ortholog groups); major facilitator superfamily MFS-1 (IPR011701, six groups), which may be related to secondary-metabolism production; and five ortholog groups that encoded the zinc finger, C2H2 type (IPR007087). This suggests that signaling and regulation may play the largest role in niche and host preference. Intriguingly, three ortholog groups with glycosyl transferase, family 54 (IPR006759) domain, were enriched in zoophiles and anthropophiles (two to three copies in these species) compared to in

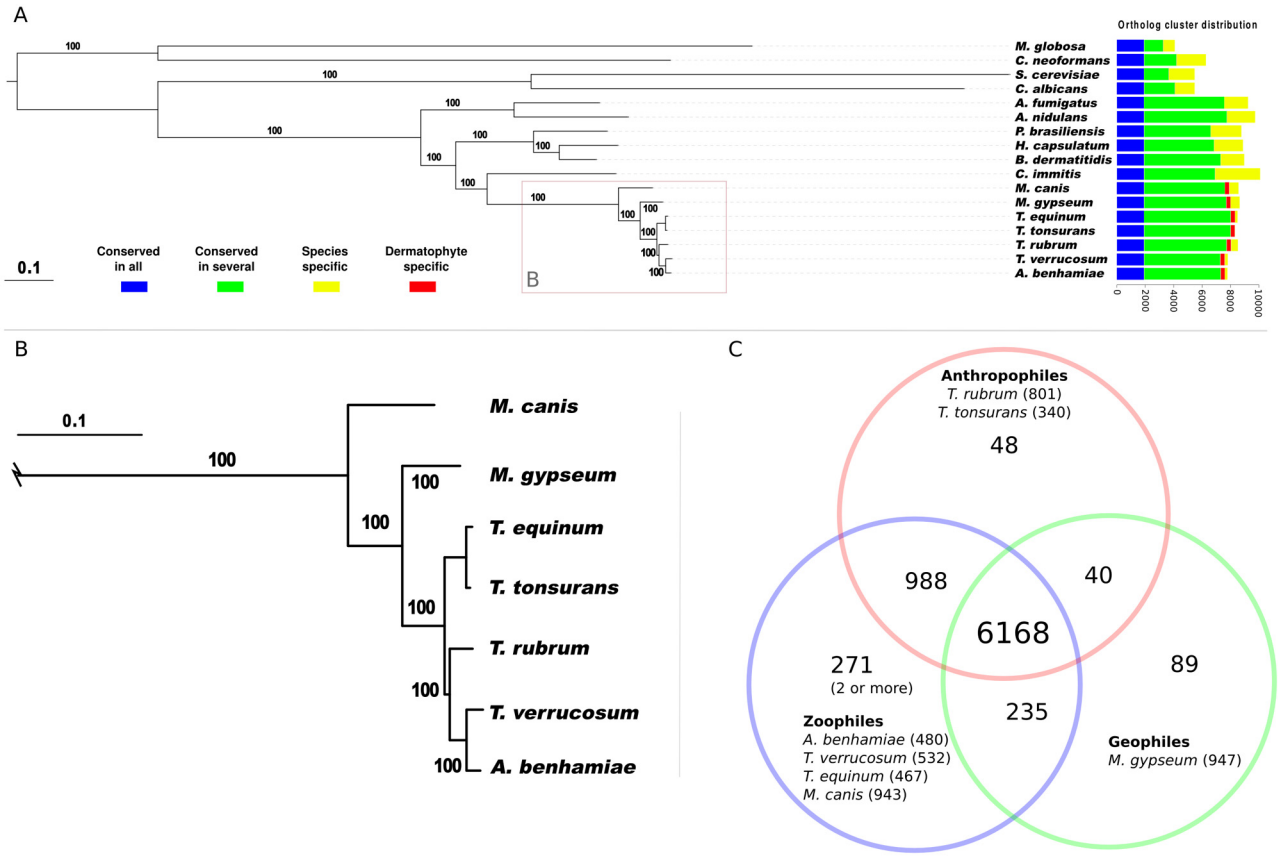

FIG 1 (A) Phylogenetic relationship and gene conservation of the dermatophytes and other species compared in this study; (B) section of the phylogeny in Fig. 1A comprising the dermatophytes; (C) shared and unique ortholog groups (see Materials and Methods) for the seven dermatophyte genomes grouped by host or ecological niche preference. Ortholog groups in this figure include paralogous duplications; therefore, total gene counts including paralogs are higher. The number of genes unique to each species (not contained in any other species in panel A) is shown in parentheses after the species name.

M. gypseum (one copy). GT54 is responsible for "decorating" mannan with *N*-acetyl glucosamine. Some of these genes were predicted to be glycosylphosphatidylinositol (GPI) anchored and several contain a weakly predicted secretion signal (with probabilities ranging from 0.08 to 0.89), indicating that several likely work outside the cell and may be involved in the immunomodulating properties of mannan (reviewed in reference 10). Finally, this analysis also identifies three ortholog groups that contain the peptidoglycan-binding lysin subgroup (IPR002482) and three ortholog groups with a trypsin-like peptidase domain (IPR002482), both of which are described in detail below.

Gene family analysis. To examine gain or loss of functional domains in the dermatophytes, counts of Interpro categories were compared for dermatophytes and several outgroups, including dimorphic fungi (Fig. 1A; see also Materials and Methods). To examine changes across different phylogenetic distances and phenotypic variation, multiple groups were contrasted: dermatophytes compared to all other genomes, dermatophytes compared to dimorphic fungi from the *Onygenales*, dermatophytes compared to two *Aspergillus* genomes, and dermatophytes compared to *C. immitis* (with or without the related, nonpathogenic fungus *Uncinocarpus reesii*).

By comparing the dermatophytes to the 11 other fungal genomes, we found that a total of 103 Interpro domains were significantly enriched (above the false discovery rate [FDR] of >0.05; see Materials and Methods) in the dermatophytes (out of 6,121 total IPR categories). The top 25 domains overrepresented in dermatophytes fall into four large functional categories: secondary metabolism, kinases, proteases, and LysM binding domains (Fig. 2). In contrast, only 23 IPR domains were depleted in the dermatophytes compared to the other fungi (see Data Set S1 in the supplemental material). The depleted domains include categories that are involved in sugar metabolism and plant cell wall breakdown. Several carbohydrate active enzyme (CAZy [11]) classes are depleted in dermatophytes, including glycoside hydrolase (GH) families involved in the breakdown of hemicellulose or pectin, major components of the plant cell wall. An additional set of IPR depleted domains includes several types of sugar transporters. These results suggest that such losses in dermatophytes represent adaptation to the infrequent use of saccharides from plant material and potentially sugar substrates in general. In support of this point, the dermatophyte genomes do not contain any identifiable cellulose-binding domains (IPR000254), typically a hallmark of fungi that utilize plant material for nutrients (12).

Fewer significant IPR categories were identified when the comparison was focused on subgroups most closely related to dermatophytes (Fig. 1A and 2). Compared to the dimorphic fungi in the *Onygenales* as a group, 37 IPR categories are enriched in dermatophytes, while six IPR categories are depleted. Most of the IPR categories enriched in dermatophytes compared to the dimorphic

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

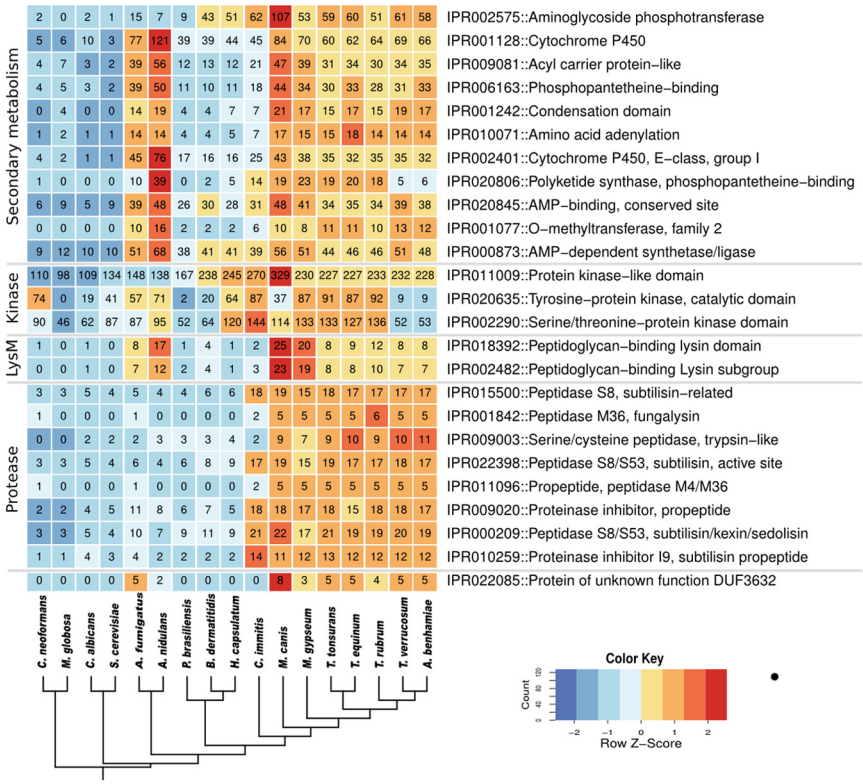

FIG 2 IPR domains most enriched in dermatophytes compared to all nondermatophytes in the study. Values are colored along a blue (low) to red (high) color scale, with color scaling relative to the low and high values of each row. IPR categories are sorted by *P* value within each supercategory. IPR domain enrichments shown have *P* values of at least 3e−8.

fungi (see Data Set S1 in the supplemental material) were also identified in the comparison to all other fungi described above. The top 10 enriched groups include four areas of cellular function; kinases involved in signaling, secondary metabolism, the LysM domain involved in binding of chitin, and transporters. Similarly, when the dermatophytes were compared to the more distantly related *Aspergillus* species, dermatophytes show enrichments for proteases; they also contain fewer domains involved in cellulose and plant cell wall breakdown. A similar excess of proteases over carbohydrate-degrading enzymes was shown to correlate with a preference for protein over saccharide substrates in *Uncinocarpus reesei* (Onygenales) (13).

Comparing the anthropophile dermatophytes (*T. tonsurans* and *T. rubrum*) with the zoophile dermatophytes (*T. verrucosum*, *T. equinum*, *M. canis*, and *A. benhamiae*) revealed a small number of IPR families that could play a role in specifying host preference, consistent with the ortholog analysis described above. The only IPR families enriched in anthropophilic dermatophytes (above the FDR of 0.05) are kinases and transcription factors. These include serine/threonine protein kinase domain (IPR002290), the catalytic domain of tyrosine protein kinase, catalytic domain (IPR020635), and Zn(2)-C(6) fungal-type DNA-binding domain (IPR001138). This strengthens the suggestions that modulation of signaling pathways and transcriptional regulation are rapidly evolving and may play a role in host specificity.

LysM domains. The LysM binding domains (IPR018392 and IPR002482) are highly enriched in the dermatophytes (Fig. 2). The number of genes containing these LysM domains ranges from nine in *T. verrucosum* to 31 in *M. canis*. The LysM domain was originally described in bacteria, where it binds to bacterial cell wall peptidoglycan. The domain has since been identified in other fungal genomes, where it binds chitin and related carbohydrates (reviewed in reference 14). In fungal plant pathogens, LysM proteins are thought to be effectors of host response (15–17). In particular, they are implicated in evasion of the host innate immune response (14) by securing fragments of chitin so that the chitin cannot stimulate the immune response. Similar to what was found in other fungi, dermatophyte LysM-containing genes are predicted to be secreted and found in combination with other functional domains that suggest associated functions.

The LysM domain in dermatophytes (see Fig. S2 in the supplemental material) is similar to that characterized for the mycoparasite *Trichoderma* (18) and differs from the standard Pfam hidden Markov model (hmm) identified in bacteria. The LysM from dermatophytes and other fungi contains three prominent cysteine residues not present in the standard Pfam profile. These substitutions are predicted to alter the structure of the LysM domain by interrupting one of the alpha helices and altering the turn between the helices (see Materials and Methods). These changes were shown to alter binding specificity so that the fungal LysM domain binds chitin (16, 17).

Fungal proteins containing LysM domains form two major related groups (Fig. 3). Clade A (52 genes from dermatophytes) is predominantly composed of genes with LysM, followed by the chitin binding domains GH18 and Chitin_binding_1. Clade B (60 genes from dermatophytes) contains almost entirely genes with

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

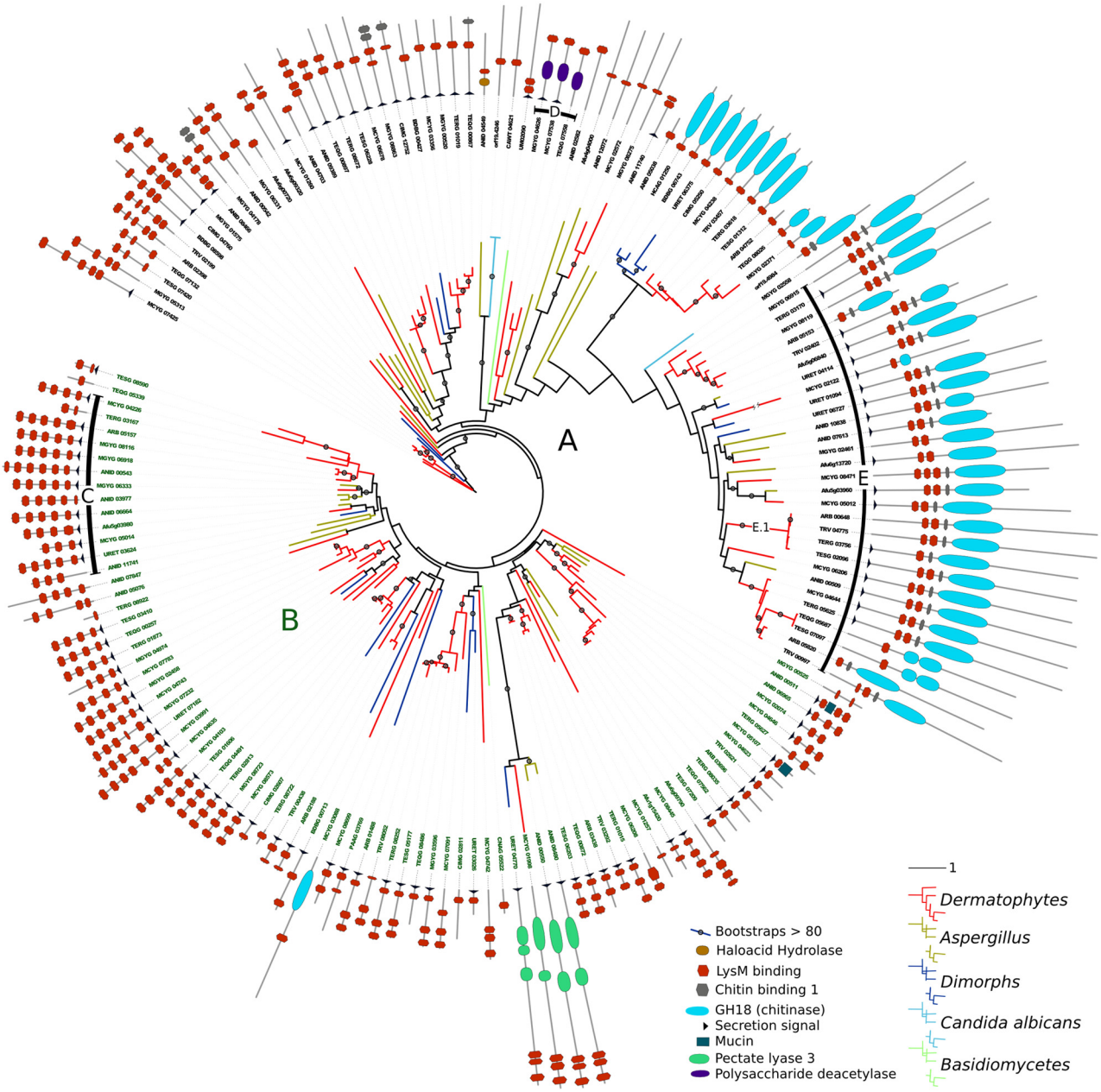

FIG 3 Phylogenetic tree of all fungal proteins containing LysM domains. Phylogeny was estimated using RAxML (see Materials and Methods). Two major clades were identified and labeled clade A (black gene names) and clade B (green gene names). Clade B may have arisen differently, as the majority of proteins in clade B do not have chitinase domains. Clade C contains LysM domains that have evidence of recent intergene duplication of the LysM domain, accounting for the high number of LysM domains in this group. Genes in clade D contain polysaccharide deacetylase domains in combination with the LysM domain. This combination is found only in dermatophytes. Genes in clade E contain the WA/GW signature in the LysM domain, typically the domain closest to the chitinase domain. Clade E.1 (labeled in the phylogeny tree) is a recently expanded group of genes that do not contain the WA/GW phylogenetically embedded within the WA/GW signature-containing group.

multiple LysM copies but no other chitin binding domains. An exception is the *M. canis* LysM gene (MCYG_03088), which contains the only GH18 domain in clade B. This gene is also unique in its domain order, as it is the only gene in which the LysM domain is at the 3′ end of the gene and the GH18 is at the 5′ end. In clade A, LysM domains are always at the 5′ end, followed by the GH18 domain in some proteins (28 dermatophytes genes). Also in clade A, the Chitin_binding_1 domain appears in combination with the LysM domain in genes both with and without the GH18 domain. Genes that do not contain the GH18 domain also appear to be some of the deepest branches in clade A, suggesting that the gain of the GH18 domain gave rise to a combination of all three domains in this clade.

Other domain combinations with LysM highlight the link to chitin binding and amplification of this protein class in dermatophytes. In *M. canis*, *M. gypseum*, and *T. equinum*, LysM domains

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

are found in combination with the polysaccharide deacetylase type 1 (Pfam: Polysacc_deac_1; purple oval in Fig. 3), which is common in chitooligosaccharide deacetylase (19). This combination with LysM can be found only in these three dermatophytes out of all known proteins containing the polysaccharide deacetylase type 1 (http://www.ebi.ac.uk/interpro/). Polysaccharide deacetylases are involved in chitin catabolism, and its presence in these species may allow chitin scavenging in the soil, defense from other fungi, or cell wall modification. Two of these dermatophyte genes (in *M. gypseum* and *M. canis*) contain a secretion signal, supporting a functional role outside the fungus. While the *T. equinum* (TEQG_07558) protein does not have a secretion signal, this may be due to a small gap in the assembly upstream of the gene, so that the gene model is truncated.

LysM domains are often tandemly arrayed in proteins, varying from one to six copies. The variable number and tandem orientation suggests that there could be ongoing duplication of LysM domains within a gene to provide new variants. To test this hypothesis, we examined a phylogeny of each LysM domain from all of the proteins in the dermatophytes and our comparative genome set (see Fig. S3 in the supplemental material; see also Materials and Methods). Only a small subset of the LysM domains in the same protein appear closely related enough to have arisen by recent duplication (see Fig. S3). The few proteins with suggested intraprotein duplication of LysM domains are confined to one section of the LysM protein phylogeny (clade C in Fig. 3). The phylogenetic tree shows that the domain expansion occurred just before speciation (MGYG_06918 and MGYG_08116, both from *M. gypseum*, and ARB_05157, from *A. benhamiae*) (see Fig. S3). However, these proteins appear to be rapidly evolving, as the mean Ka/Ks (see Materials and Methods) is significantly higher in proteins containing LysM domains than in other proteins (*P* value of 0.00018).

In the LysM HMM refined to include dermatophytes, we identified a unique motif, containing the amino acid signature WA/GW (tryptophan-alanine or glutamine-tryptophan). Inspection of the multiple sequence alignments revealed that only 26 of the LysM domains contained the WA/GW signature (clade E in Fig. 3). This WA/GW domain is predicted to extend the loop of the potential binding cleft (see Fig. S4 in the supplemental material). This suggests that the changes may allow a larger molecule to fit in this binding site, such as an alpha sugar that has a different linear structure than beta sugars, such as chitin or bacterial cell wall peptidoglycan. The proteins that contain the WA/GW signature also contain GH18 chitinase catalytic domains (Fig. 3, clade E). Phylogenetic analysis suggests multiple lineage-specific expansions of WA/GW LysMs in *M. canis* (four copies) and in *M. gypseum* (three copies) as well as in *Aspergillus* spp. and *U. reesei*. Interestingly, four closely related genes in clade E (Fig. 3, clade E.1) do not contain the WA/GW signature, suggesting that the signature was lost fairly recently.

The expansion of a gene family can be driven by tandem duplication, and in at least one other fungal genus, *Trichoderma*, some LysM-containing proteins are organized in clusters (18). In *Trichoderma*, a gene containing both GH18 and LysM will often be adjacent to a gene containing LysM only. In *T. rubrum*, two loci have a similar organization, where a protein containing chitinase and LysM domains is adjacent to a protein containing only LysM domains; no other closely linked copies of LysM genes were found in the dermatophyte genomes. This clustered genomic location could be important for coregulation, as frequently observed for secondary-metabolism gene clusters (20). One pair of genes located next to each other in *T. rubrum* (TERG_05625 and TERG_05627) appears differentially expressed. When *T. rubrum* is grown on soy protein, TERG_05625 is expressed, and when grown in keratin, TERG_05627 is expressed (21). Another two of the 15 LysM-containing genes in *T. rubrum* (TERG_01019 and TERG_03618) are transcriptionally active when grown in standard yeast extract-peptone-dextrose (YPD) medium at several time points (22). However, the available expressed sequence tags (ESTs) likely underrepresent gene expression, and thus a whole genome transcription analysis is needed for a thorough study.

In addition to LysM domains, several IPR domains that encode chitinase are enriched in dermatophytes. These include glycoside hydrolase family 18 catalytic domain (GH18; IPR001223), glycoside hydrolase, chitinase active site domain (IPR001579), and the chitinase II domain (IPR011583) (see Data Set S1 in the supplemental material). A domain for binding chitin, chitin-binding, type 1 domain (IPR001002), is significantly enriched as well (see Data Set S1). Another enriched domain, glycoside hydrolase, family 24 domain (GH24, IPR002196), is common to lysozyme genes that typically hydrolyze the bond between *N*-acetylmuramic acid and *N*-acetyl-d-glucosamine bonds in bacterial cell walls. In the CAZy database (http://cazy.org), only seven nondermatophyte fungi contain a GH24 domain. In our study, we found that all dermatophytes except for *Arthroderma benhamiae* contain two genes that encode the GH24 domain. Additionally, we identified one GH24 copy each in the dimorphic fungi *B. dermatitidis* and *H. capsulatum* and in *A. nidulans*. The duplication of GH24 in dermatophytes along with other chitin and related carbohydrate-specific adaptations suggests that these proteins could aid growth in a wide variety of niches, including the soil and human skin.

Protease family expansions in dermatophytes. Secreted subtilisin proteases (Merops family S8 [23]) are among the most important gene families for fungi that live on skin. The subtilisins are expanded in the dermatophytes (5) and could play a role in keratin degradation, but these proteases are also found in other fungi associated with animals that are not skin pathogens (24). While most dermatophytes contain 12 subtilisins, *T. tonsurans* has gained an additional copy with an I9 inhibitory domain (see Fig. S5 in the supplemental material), and one subtilisin appears to have been lost in *M. canis*.

Other protease families involved in degrading keratinized tissue also show dynamic copy number variation. Previous studies found that the M35 (deuterolysins) (5) and M36 (fungalysins) (25) metalloendopeptidase families were expanded in several dermatophytes. Both the M35 and M36 families are present in at least 5 copies in each dermatophyte genome. In the M35 family, both *M. canis* and *T. verrucosum* appear to have gained an extra copy of this gene compared to other dermatophytes (6 copies); for comparison, there are 7 M35 genes in *C. immitis* and 3 to 4 in the *Aspergilli*. In contrast, in the M36 family, most species have 5 copies, while *T. rubrum* has gained an extra copy (6 copies); M36 is present in only 2 copies in *C. immitis* and zero or one in all other analyzed nondermatophyte fungi (see Fig. S5 in the supplemental material). Strikingly, an M36 fungalysin conserved in *T. equinum* and *T. tonsurans* contains a predicted GPI anchor that is not found in other fungi (see Materials and Methods). This suggests that these proteases are likely active at the cell wall in these species, similar to the novel GPI anchoring of the S10 protease family in

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

TABLE 2 Common and unique nonribosomal peptide synthase (NRPS) and polyketide synthase (PKS) gene clusters in dermatophytes

Type of cluster	Total no. of clusters	No. of clusters common to four or more dermatophytes[a]	No. of clusters unique to less than three dermatophytes						
			T. rubrum	*T. tonsurans*	*T. equinum*	*A. benhamiae*	*T. verrucosum*	*M. canis*	*M. gypseum*
Nonreducing PKS	8	3	0	1	1	0	0	4	0
Reducing PKS	13	5	0	1	1	1	1	4	6
NRPS	14	12	0	1	1	1	0	0	0
Reducing PKS/NRPS hybrids	8	3	0	0	0	0	0	3	2
Species totals, including common[b]	43	23	23	22	23	25	24	32	27

[a] In these clusters, some of the genes in a given species may be missing because of gaps in the sequence.
[b] Total numbers within a species include those clusters in common as well as those that are unique. See Table S3 in the supplemental material for all details. For a comparison, *C. immitis* has 16 total clusters (5).

dermatophytes (26). This illustrates that dynamic adaptation to protein substrates is still ongoing after speciation in these dermatophytes.

Additionally, the dermatophyte genomes are significantly enriched for other protease families (see Fig. S5 in the supplemental material), including both endopeptidases (S41, M4, M43) and exopeptidases (S10, M28, M14, S41, S12). This expansion of both classes of proteases, which are able to cleave from the middle or ends of protein chains, respectively, could enable highly efficient degradation of protein substrates, as suggested based on analyses of the *A. benhamiae* and *M. canis* secreted proteomes (27).

Secondary metabolism. Dermatophytes are enriched for genes connected to secondary-metabolite production. To understand what types of enzymes are present, we characterized the nonribosomal peptide synthase (NRPS) and both reducing and nonreducing polyketide synthase (PKS) and PKS/NRPS hybrid genes (Table 2; see also Table S3 in the supplemental material). The dermatophytes have a large number of NRPS/PKS genes, ranging from 22 in *T. tonsurans* to 32 in *M. canis*. For comparison, *C. immitis* (5) has only 16 NRPS/PKS genes (Table 2). The most striking differences are in the NRPS and in the PKS/NRPS hybrids. The largest group in the dermatophytes is PKS (including the PKS portion of PKS/NRPS hybrid genes), which is present in 11 to 21 copies in the seven dermatophyte genomes (Table 2; see also Table S3). Unlike *Aspergillus* (5, 28), dermatophytes demonstrate prevalence of reducing over nonreducing PKSs, particularly in *M. gypseum*. In addition, *M. gypseum* contains fewer NRPS and nonreducing PKS genes than the other dermatophytes. As *M. gypseum* is the lone soil preferring sequenced dermatophyte (see Table S3), this suggests that these genes may be more important for association with hosts.

Twenty-eight of the secondary-metabolite genes are present in less than three dermatophytes species, with no close orthologs in other species (Table 2). These include eleven unique genes in *M. canis* and eight unique genes in *M. gypseum*. These 28 unique genes in the dermatophytes are strong candidates for virulence factors contributing to the unique host range and pathogenesis of these species.

While most NRPS and PKS genes are orthologous across the seven dermatophytes, the accessory genes in the secondary-metabolite gene clusters are more dynamic. We used the SMURF cluster detection tool (29) and integrated the results from SMURF with a phylogenetic analysis of NRPS/PKS genes (grouping them into families of orthologs with recent paralogs) to identify clusters most likely to produce novel metabolites (see Materials and Methods).

While half of NRPS clusters appear conserved, the rest have differences in gene content that suggest different products could be generated. In 7 of 14 NRPS families, the same decorating genes flanked the NRPS gene, with the exception of several (one to four genes) hypothetical genes. In the six NRPS clusters with the most variable accessory gene content, the difference was typically due to alterations of the cluster in *M. canis* or *M. gypseum*. In one case, an *M. canis* NRPS (MCYG_04549) gene had moved to a region near a PKS (MCYG_04545) gene forming a hybrid cluster. In two families for *M. gypseum* and in one family for *M. canis*, the NRPS gene clusters contained large insertions not found in any other dermatophyte (MGYG_02587 had a 13-gene insertion, MGYG_08842 had an 8-gene insertion, and MCYG_08441 had a 15-gene insertion); these dramatic alterations are likely to impact the secondary metabolite produced. A cluster in *T. equinum* (NRPS gene TEQG_0583) is interrupted by a 31-kb insertion containing gypsy and LINE elements, which may disrupt the coexpression of the genes in *T. equinum*. In one family, the *T. rubrum* (TERG_02711) SMURF cluster was altered due to a 22-gene deletion with respect to *T. equinum* and *T. tonsurans* but was otherwise syntenic. One NRPS gene cluster (TEQG_01720) conserved across the *Trichophyton* spp. was recently noted to be substituted in *M. gypseum* by a distinct NRPS (MGYG_05790) cluster, which may produce porphyrin compounds important for iron metabolism (30).

We found the PKS families to be similarly variable across the dermatophytes. In 7 out of 18 families, accessory genes were conserved, and 1 family contained similar accessory genes, with the exception of a duplication in the PKS gene for *M. gypseum* (MGYG_04674 and MGYG_01493). Seven families contain predominantly duplications in *M. canis* and *M. gypseum* (or losses in the *Trichophyton* spp.), with only three of these families including orthologs in the *Trichophyton* spp. In one of the three cases with a large difference in predicted SMURF cluster size, an apparent translocation of genes in *M. gypseum* (MGYG_00039) doubled the number of genes in the PKS cluster, from eight genes (as predicted in *T. rubrum*, TERG_01487) to 16 genes. Similarly, the *M. canis* cluster for PKS gene MCYG_03598 contains an 11-gene insertion not found in any other dermatophyte. In another PKS family, transposable elements in *T. equinum* (TEQG_05640) and *T. tonsurans* (TESG_01634) have created a 25-kb inserted region of gypsy and LINE elements, decreasing the number of accessory genes from 12 to 6.

Protein kinases. Dermatophytes contain an unusual number of novel eukaryotic protein kinases (ePKs) not seen before in fungi. The sets of ePKs (kinomes) of the dermatophytes were de-

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

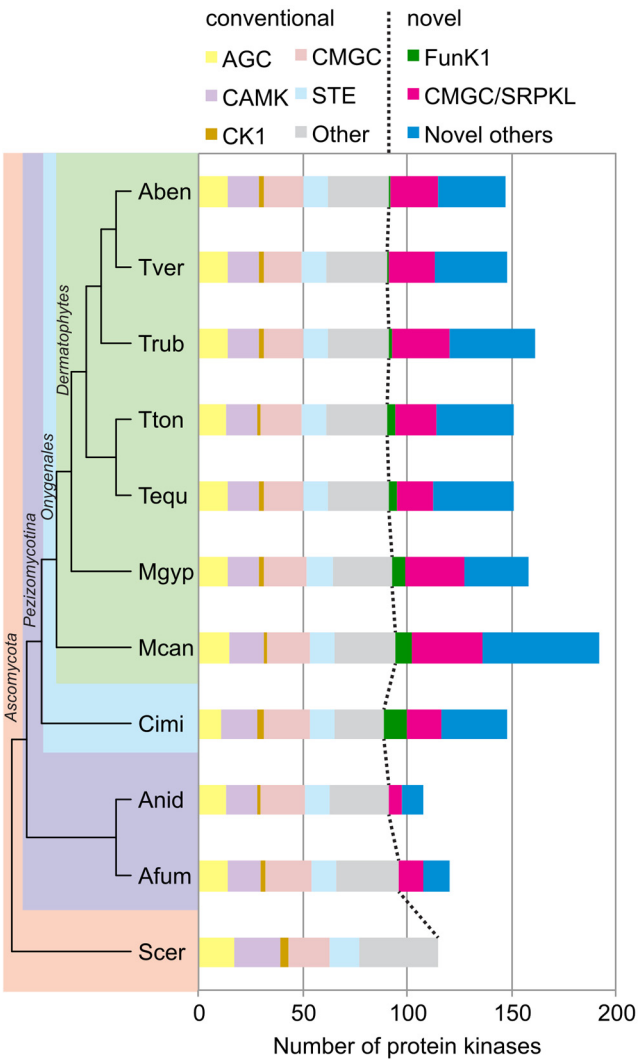

FIG 4 Eukaryotic protein kinase superfamily members (kinomes) of the dermatophytes compared with the kinomes of other fungi. Kinases from established families are shown on the left of the black line, while novel or recently discovered kinases are shown on the right. Kinases are classified into major groups shown as colored blocks (76). Abbreviations: Aben, *A. benhamie*; Tver, *T. verrucosum*; Trub, *T. rubrum*; Tton, *T. tonsurans*; Tequ, *T. equinum*; Mgyp, *M. gypseum*; Mcan, *M. canis*; Cimi, *C. immitis*; Anid, *A. nidulans*; Afum, *A. fumigatus*; Scer, *S. cerevisiae*; AGC, protein kinases A, G, and C; CAMK, calcium/calmodulin-dependent kinases; CK1, casein kinase 1; CMGC, cyclin-dependent kinases (CDK), mitogen-activated, glycogen-synthase, and CDK-like kinases; STE, sterile phenotype kinases; FunK1, fungal-specific kinase 1; SRPKL, serine/arginine-rich protein-specific kinase-like; other, kinases not falling into major groups.

termined (see Data Set S2 in the supplemental material) and compared to those of other fungi (Fig. 4; see Data Set S2). The dermatophyte kinomes, summarized in Fig. 4, contain from 147 to 194 ePKs and are significantly larger than the kinome of the best characterized fungus, *Saccharomyces cerevisiae*, which contains only 118 ePKs (31). Dermatophytes contain from 55 (*A. benhamiae*) to 94 (*M. canis*) novel kinases that do not fall into known families (Fig. 4, right side of line), as well as 90 to 94 widely conserved conventional kinases (Fig. 4, left side of line).

Phylogenetic analysis of novel kinase sequences revealed large superclusters above the orthology level. Maximum likelihood trees (see Fig. S6 in the supplemental material) show that these superclusters can be classified as serine arginine protein kinase-like 1 (SRPKL1), SRPKL2, SRPKL3, PezK1, and PezK2, where PezK indicates the restriction of these kinases to *Pezizomycotina* (see Data Set S2 in the supplemental material). In contrast, several dermatophyte SRPKLs have orthologs in the basidiomycete *C. cinerea* (MCYG_00930:CC1G_15433, MCYG_01080:CC1G_02719, MCYG_08457:CC1G_05630), suggesting an early fungal origin for this family. While a large number of kinases from widely conserved families are present in many species, including *S. cerevisiae* (Fig. 4), the dermatophyte kinomes contain several families that have been lost from the budding yeast (see Data Set S2), underlining the reduced kinome of *S. cerevisiae* compared to other fungi, as also noted compared to *S. pombe* and *C. cinerea* (32, 33). The dermatophyte kinome includes from one to eight members of the recently described FunK1 kinase family (Fig. 4), which is restricted to *Pezizomycotina* and *Basidiomycota* (13, 33). These novel kinases are more likely than conventional ones to contribute to the unique biology of the dermatophytes and could provide pathogen-specific drug or diagnostic targets.

Microsporum canis, the earliest branching and most divergent of the sequenced dermatophytes, contains 30% more novel kinases than the average of the others (Fig. 4). While *M. canis* has a higher number of proteins overall, this does not account for the kinase expansions, as on average there are only 5.2% more proteins compared to the other genomes. The increased number of kinases is explained by a combination of reduced ortholog loss and increased duplication to create paralogs in many ortholog groups (see Data Set S2 in the supplemental material). Kinase ortholog groups are most highly conserved in *M. canis*, which contains 61 ortholog groups versus an average of 53 in other dermatophytes. Also, the average number of genes in an ortholog group in *M. canis* is 1.48, compared with 1.2 for the others, indicating a higher rate of paralog creation. Additionally, *M. canis* has 22 kinases that are not in ortholog groups (compared to a dermatophyte average of 4.7 kinases), indicating a higher rate of paralog creation. The relatively diverse kinome of *M. canis* may provide additional control points for signaling pathways.

Many of the novel kinases have sequence motifs similar to SRPK kinases, including a distinctive KLG(WFYH) GXXXSTVWL motif near the N terminus of the kinase domain. Other novel kinases, in contrast, do not share motifs with known families. The novel kinases have significantly higher average Ka/Ks values than the conventional ones (0.228 versus 0.082, respectively, Mann-Whitney U P value of $<2.2e-16$), indicating that they are changing more quickly and may be evolving in response to variable external factors. As observed in other kinase expansions within pathogens (34), a significant proportion of the novel kinases, in this case 33%, of SRPKL and PezK kinases (see Data Set S2 in the supplemental material) are missing one or more essential active site residues and appear to be inactive or "pseudokinases." Several of the *T. rubrum* pseudokinases are expressed, including an SRPKL1 (TERG_06939), three PezK2s (TERG_05772, TERG_00915, TERG_01919), and two unclassified kinases (TERG_03779, TERG_06932). Of these, orthologs of three (TERG_05772, TERG_03779, and TERG_06932) are predicted to be inactive in other dermatophytes, suggesting a conserved function.

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

Analyses of other genes and gene families with important roles in virulence. (i) Mating and meiosis genes. Conservation of the mating type locus (*MAT*) and meiosis-specific genes in all dermatophytes suggests that all species have the capacity to complete sexual reproduction, which is correlated with virulence in some fungal species (35). Based on mating assays and indirect evidence from population genetic studies, it has been widely accepted that geophilic and most zoophilic dermatophytes have extant sexual cycles, whereas most anthropophilic dermatophytes have lost the ability to sexually reproduce (4). A recent study characterized both *MAT* locus idiomorphs in *M. gypseum* and one idiomorph in the sequenced genomes of the other species (Table 1) (36). Outside of the *MAT* locus, many meiotic genes that serve important roles in sexual reproduction are conserved in all dermatophytes. The meiosis-specific genes *SPO11*, *HOP1*, *HOP2*, *MND1*, *REC8*, *DMC1*, *MSH4*, and *MSH5* are found in all seven dermatophytes. Of a larger set of 49 genes important for mating and meiosis (5), 41 are conserved in all dermatophyte species, three were absent from all dermatophytes, and the remaining five did not show a clear correlation with the mating ability of the species. As anthropophilic dermatophytes have an extant *MAT* locus and meiotic genes, the apparent inability to detect mating under laboratory conditions in some species may be explained by as-yet-unknown genetic, epigenetic, or environmental factors (such as loss of mating partner) that either limit or are necessary for sexual reproduction.

(ii) Comparison of dermatophytes to other skin-adapted fungi. Divergent fungi whose ecological niche is animal skin have specialized differently than dermatophytes to this environment. For example, the basidiomycete *Malassezia* sp. is distantly related to dermatophytes, has a much smaller genome (9 Mb) (37), and is found in sebum-rich areas of skin. *Malassezia* sp. is enriched for proteases similar to the dermatophytes (37). However, whole-genome analysis showed that *Malassezia globosa* has deficiencies in fatty acid synthesis genes (fatty acid synthase, delta-9 desaturase) and enrichments in lipid-hydrolyzing enzymes (37, 38). The missing fatty acid synthase gene is present, with some modifications, in the dermatophytes (see below). In addition, the acid sphingomyelinase gene family is expanded in *M. globosa* and absent in the dermatophytes. It is likely that these genome differences reflect the phylogenetic distances and the differences in niche occupied by these fungi, with *M. globosa* dependent on lipid provided by the host.

Fatty-acid synthase is made up of an alpha subunit and a beta subunit. These two fatty acid synthase subunits are absent in *Malassezia globos*, duplicated and altered in dermatophytes, and duplicated in *C. immitis* (as recovered by orthoMCL). As the gene duplications are not present in the other dimorphs, the duplication must be in the common ancestor of *C. immitis* and the dermatophytes. In *M. gypseum*, *T. equinum*, and *T. tonsurans*, the duplicated beta subunit diverges from the FAS beta Interpro domain (IPR016452) but is detected by homology in the MaoC dehydratase-like domain, one of the domains in the FAS beta subunit (see Table S4 in the supplemental material). The alpha subunit is also duplicated in dermatophytes (except for *M. gypseum*) and in *C. immitis*. In *T. tonsurans* and *T. equinum*, one copy of the alpha subunit has undergone a dramatic fusion to the upstream gene annotated as encoding a drug resistance efflux protein (membrane embedded). At another point in fatty acid synthesis, the delta-9 desaturase, which places a double bond at the 9th carbon of fatty acids, is present in most dermatophyte genomes in one copy, with *M. gypseum* and *M. canis* having 2 copies.

(iii) Ergosterol metabolism. Although the number of antifungals currently available has increased, most antifungal drugs target ergosterol, the cholesterol analog found predominantly in fungal cell membranes. Fungal growth is challenged in two ways: by interference caused by direct interaction of polyenes, such as amphotericin B, with ergosterol, leading to disruption of the membrane, and by inhibition of specific steps of sterol biosynthesis, with drugs such as allylamines, thiocarbamates, azoles, and morpholines, that limit the availability of ergosterol.

Most of the genes involved in ergosterol biosynthesis (ERG) in dermatophytes are conserved, while a subset (four families in particular) varies in copy number (see Table S4 in the supplemental material). Compared to *C. immitis*, the *ERG3* family appears to have undergone expansion in the dermatophytes to two copies (*T. verrucosum* and *A. benhamiae*) or three copies (other dermatophytes). Similarly, the *ERG4* gene was duplicated in *T. equinum*, *T. tonsurans*, and *M. gypseum*; *ERG12* was duplicated in *M. canis*; and *ERG25* was duplicated in *M. canis* and *M. gypseum*. In contrast, *ERG7* appears to have been duplicated in *C. immitis*, but it is the only *ERG* gene to have done so in this fungus. The presence of duplicated genes in the ergosterol biosynthesis pathway could be a fungal strategy to modulate the composition and fluidity of the cell membrane, which could also confer adaptive advantages to reduce susceptibility to antifungal drugs (39).

DISCUSSION

Our study identifies families of genes that are likely involved in the interactions between dermatophytes and their hosts. These gene families may help to provide explanations for long-standing questions in dermatophyte pathogenesis. In particular, it has been observed that *T. rubrum* and *M. canis* (when in cat, the preferred host) can become almost invisible to the host immune system, possibly resulting in chronic infections (40). Previously, this was thought to be due to masking recognizable components of the fungal cell wall and the use of proteases such as DppIV to degrade the inflammatory protein substance P (40). Our analysis suggests additional mechanisms by which dermatophytes can counteract the host immune response.

To evade the immune system, dermatophytes must prevent host identification of several cell wall components. The human immune system has pattern recognition receptors that identify pathogen-associated molecular patterns (PAMPs). Two recently discovered mammalian genes, dectin-1 and dectin-2, recognize beta-1,3-glucan and beta-1,6-glucan, major components of fungal cell walls. While fungal mannans may inhibit dectin-2 (41), less is known about chitin as a PAMP. The mammalian immune system has two chitinases and several chi-lectins which identify chitin (42). While studies link chitin identification to an innate or adaptive immune reaction, little is known about how chitin recognition is linked to the immune system (42). LysM genes in fungal plant pathogens mask chitin, which is on the external surface of the hyphae, from detection by the plant immune response (16, 43, 44). It is likely that LysM domains in dermatophytes perform a similar chitin-disguising function, thereby preventing chitin immunostimulation.

The paucity of LysM-containing genes in the closely related dimorphic fungi provides support for an immunosuppressive function. *Histoplasma capsulatum*, which contains only one LysM

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

gene, is able to disguise beta-glucans by using a protective layer of alpha-1,3-glucans (45). The *Blastomyces dermatitidis* gene BAD-1 downregulates immune response directly by suppressing tumor necrosis factor alpha (TNF-α) from macrophages; this species has only four LysM genes. The small number of LysM genes in these fungi may indicate that chitin protection is not necessary for pathogenesis. Neither are there reports that dermatophytes use strategies like *H. capsulatum* or *B. dermatitidis*. The related pathogen *A. fumigatus* may utilize multiple mechanisms, as it both has a large set of seventeen LysM genes and has also been shown to use a galactosaminogalactan coat, which contains polysaccharides similar to chitin (46). It is notable that the clade of LysM domains with the WA/GW signature, with a modified LysM structure, contains genes from both dermatophytes and *Aspergillus* spp. This raises the possibility that the dermatophytes may be protecting themselves with a similar polysaccharide coat and that the modification in LysM protein folding in both fungi mediates attachment of the fungal cell wall to the polysaccharide coat or to the host.

This is the first report of an animal fungal pathogen containing significant expansions of genes with the LysM domain. Given the above-described evidence, we propose that the expansion and variation in the LysM gene family in dermatophytes provides protection in eluding the animal immune system, similar to the protection found in plant pathogens to the invading fungus. The variation in domain content suggests that this could be involved in chitin (or related carbohydrate) metabolism as well as in host attachment or other defense mechanisms. However, a recent analysis has suggested a broader role in fungal biology for LysM-containing proteins (47).

The large number of novel kinases may provide flexibility for the different dermatophyte species to occupy different ecological niches. *M. canis* can survive on a variety of animals, which may be enabled by having a larger kinome to allow more flexibility in its response to environmental challenges and a broader range of conditions. An alternative, although not exclusive, hypothesis is that the common dermatophyte ancestor utilized this diverse repertoire of kinases, more of which have been retained in *M. canis*. Inactive kinases or "pseudokinases" within the dermatophytes may function in one or more of four ways (reviewed in reference 34). They may (i) compete for substrate with active kinases, (ii) act as inhibitory pseudosubstrates, (iii) allosterically modulate the activity of other biomolecules, or (iv) act as scaffolds. The intracellular apicomplexan parasite *Toxoplasma gondii* secretes a variety of pseudokinases into host cells at the onset of infection (34). In contrast, the absence of secretion signals or transmembrane helices in the dermatophyte pseudokinases indicates that they function within the dermatophyte cells.

Gene content points to evolution of pathogens. Both the dermatophytes and dimorphic fungi are animal pathogens, yet the patterns of gene expansion and loss are strikingly different, indicating a different approach to pathogenesis. The exception is *C. immitis*, which contains similarities to both groups, as it is the most closely related dimorphic fungus to the dermatophytes. The protease content is lower in the clade containing *P. brasiliensis*, *B. dermatitidis*, and *H. capsulatum* than in the clade containing *C. immitis* and dermatophytes, suggesting a shift in the ancestor of both *C. immitis* and the dermatophytes to allow more protein degradation. More recent expansion of secondary-metabolite enzymes in the dermatophytes suggests that they require an extensive arsenal of antimicrobial or toxic compounds. This arsenal may be necessary to survive in the various niches where dermatophytes deal with different environmental stresses (soil, skin, hair), with different hosts (plants and animals), and with different cohabitating microbes. Secondary metabolites impacting the host might be identified by determining which gene clusters are actively transcribed during infection. Similarly, the expansion of the LysM domains suggests this domain's important role in pathogenesis of the dermatophytes.

The dermatophytes: subtle changes lead to a large impact. The dermatophyte species are distinguished only by small differences in gene content and genome organization. Very similar gene contents (Fig. 2), the high degree of colinearity (see Fig. S1 in the supplemental material), and the sparse number of unique genes in each species (Fig. 1C) highlight the short time since a common ancestor (Fig. 1B). Despite this short divergence time, each fungus has adapted to seemingly specific niches, host range, and preferences for different proteinaceous substrates (hair versus skin, horse or cattle versus human). These adaptations could be at the level of differential regulation of enzymes involved in these processes, and study of this would require a comparison of gene expression between species.

Another difference in the dermatophytes is that the anthropophiles appear to reproduce asexually under normal conditions. However, our analysis suggests a genomic competency for mating in the anthropophiles that is similar to that in the other species. In *Aspergilli* presumed to be asexual, the initial genomic analysis noted conservation of genes involved in mating (48, 49), suggesting that mating should be possible, and this was later shown to occur experimentally (50). If present, sexual reproduction might occur only under harsher environments where they might switch from asexual reproduction to sexual reproduction to generate recombinant progeny. In addition, the high homology of the *MAT* locus and meiotic genes among the dermatophytes indicates that crosses between different species may be possible, but the hybrids may not undergo meiosis/sporulation because of genetic divergence, which was recently observed in a cross between *T. rubrum* and *Arthroderma simii* (51). Thus, interspecies crosses (hybrids) may have a limited contribution to genetic exchange and may reflect species boundaries.

Future work addressing important questions in dermatophyte biology will benefit from the foundation provided by these genome sequences. For example, comparing genomes of sequential clinical isolates will determine if tinea reoccurrence is the result of reactivation of a latent commensal colonization or reinfection from external sources. Comparing genomes and gene expression of clinical isolates before and after antifungal treatment to these reference genomes will help address why dermatophytes appear to respond to antifungal treatment but not develop resistance after repeated treatments. Extending previous studies of gene expression (21, 52,) to compare different species in parallel under similar conditions may help explain the phenotypes of different species or strains. Finally, analysis of the expression of kinase genes and other signaling systems can help characterize how dermatophytes respond to host signals in both active infection and latent colonization.

MATERIALS AND METHODS

Selection of isolates for sequencing and genomic DNA preparation. The strains selected for sequencing are not the type strains, which have been in

culture for many decades. Instead, we chose to use recently isolated strain isolates (within the last 10 years, with the exception of *T. equinum*), all of which were cultured from humans (4). The species of selected strains were all determined by standard and molecular analyses. They all grow at standard rates, form conidia normally, and make protoplasts. They are all susceptible to antifungal drugs and drugs used in transformation (7) (data not shown). Most strains are available from the Centraalbureau voor Schimmelcultures (CBS) (Table 1). Genomic DNA was prepared from strains grown in liquid culture for 5 to 7 days (see Text S1 in the supplemental material).

Sequencing and assembly. Genomes were sequenced by a whole-genome shotgun approach using Sanger technology. Three genomes, *T. rubrum*, *M. canis*, and *M. gypseum*, were sequenced to 8 to 10× coverage; two plasmid libraries and one fosmid library were used for the two *Microsporum*, and one plasmid and one fosmid library were used for *T. rubrum* (see Table S1 in the supplemental material). *T. tonsurans* and *T. equinum* were sequenced to 5 to 6× coverage using two plasmid libraries for each genome (see Table S1). All genomes were assembled with Arachne and show high continuity (see Table S1). Evidence for polymorphism was examined with the Arachne HQDAnnotator (http://www.broadinstitute.org/crd/wiki/index.php/HQDAnnotator); this identified a low rate of high-quality single-nucleotide disagreements from the reads in four of the assemblies (1 in 32,000 bases in *M. canis* and *T. rubrum*, 1 in 46,000 bases in *T. tonsurans*, and 1 in 70,000 bases in *T. equinum*) and a higher rate in *M. gypseum* (1 in every 8,000 bases).

Transposable elements. TEs were identified using the programs LTRFinder (53), LTRHarvest 1.3.5 (54), transposonPSI (http://transposonpsi.sourceforge.net), and RepeatModeler open-1.0.4 (http://www.repeatmasker.org/RepeatModeler.html). The identified TEs were clustered with CD-HIT version 4.3 (55), requiring at least 80% sequence identity to find the consensus TEs in each genome. The repeats were classified by aligning the TEs to RepBase 15.04 library. Spurious TEs were removed by translating the TE sequences into six frames with the EMBOSS 5.0 (56) tool transeq and identifying domains with Pfam 24.0. TEs with no recognizable domain related to TEs and a high BLAST bit score (above 500) to a predicted gene model were filtered out.

Gene prediction and syntenic analysis. Protein-coding genes were initially predicted using a combination of gene models from the gene prediction programs FGENESH (57), GENEID (58), and GeneMark-ES (59) as well as EST-based automated and manual gene models (see Text S1 in the supplemental material). We improved the consistency of the gene models among this group of genomes by examining the alignments of protein orthology groups identified using OrthoMCL (60). The gene sets were then filtered by removing spurious gene models based on matches to repeat and low-complexity sequences (see Text S1). Regions of colinear genes between the genomes were determined using DAGChainer (61) anchored with related genes in each genome (defined as a BLASTp E value of less than 1e−10) and requiring at least 6 colinear genes, while not allowing a gap of more than 5 genes.

Gene family and protein domain analysis. Gene families were constructed using OrthoMCL (60) for the seven dermatophytes and 10 additional genomes (*Coccidioides immitis*, *Histoplasma capsulatum*, *Paracoccidioides brasiliensis*, *Blastomyces dermatitidis*, *Aspergillus nidulans*, *Aspergillus fumigatus*, *Candida albicans*, *Saccharomyces cerevisiae*, *Cryptococcus neoformans*, and *Malassezia globosa*). These data are available as Data Set S3 in the supplemental material and on the Broad Institute Dermatophyte website (http://www.broadinstitute.org/annotation/genome/dermatophyte_comparative/MultiDownloads.html). Secretion signals were predicted with SignalP version 3.0. Proteins with a secretion signal were analyzed with the fungus-specific GPI prediction program BIG-PI (62) to predict GPI anchoring sites using the website http://mendel.imp.ac.at/gpi/fungi_server.html. Genes were annotated with Interpro domains using runInterproScan (version 4.7.1, database version 29.0) (9). Significant differences in domain content between the compared genomes were identified in R using the hypergeometric distribution to calculate a *P* value of the difference in the number of Interpro (IPR) domains in dermatophytes and compared to other genomes. The results were filtered using a q value (set to 5% allowable FDR) (63) to account for multiple testing.

LysM domain search. Genes from each genome containing the LysM domain (Fig. 3) were identified by Interpro. To ensure that all possible LysM domains were identified where evidence indicated that the structure was likely to be altered, we compiled all identifiable LysM domains in the genomes to be compared and created a fungus-specific LysM HMM using HMMER 3.0 (64), repeating the process through three iterations until convergence was reached. This model was used to identify additional LysM domains in the compared fungi.

Unique LysM region. To understand what may be different about the WA/GW domains, we used the second LysM domain from the 5′ end in the *M. canis* MCYG_04644 gene (labeled MCYG_04644_2_2) as a representative for the WA/GW LysM domain. The three-dimensional (3-D) structure of the *M. canis* LysM domain was predicted using the Phyre2 server (65). This structure aligned to the *Magnaporthe oryzae* LysM domain from the c2l9yA gene (mgg_03307) (16) in the package VMD (66) with the STAMP three-dimensional alignment tool (67), as shown in Fig. S4 in the supplemental material.

LysM domain expansion. To analyze the relationship of LysM domains, each domain was assigned a unique identifier referring to the number of domains in the protein and the order for this particular domain (as in the example of MCYG_04644_2_2). All domains were aligned with MUSCLE (68) using default settings. The best amino acid model of evolution was estimate with ProtTest (69), as the WAG + G model. This model was used in RAxML (70) to estimate the phylogeny of all LysM domain reconstructions using 2,000 bootstrap replicates. The R (http://cran.org) package APE (71) was used to calculate the cophenetic distance (tree space) between LysM domains from the same gene (using the cophenetic.phylo command). Distances on an ultrametric version of the LysM domain tree that were less than 0.01 indicated a very close relationship, while multiples of 0.01 indicated further distances on the tree. The relationships were then mapped onto the RAxML tree (a section is shown in Fig. S3 in the supplemental material) for visual inspection and classification of how close the relationship is compared to the whole gene tree in Fig. 3.

Ka/Ks analysis. To determine the evolutionary rates of gene families, we applied the codeml tool (using the branch invariant model) in the PAML package (72) using the *C. immitis* genome as the outgroup in orthoMCL groups. To test the statistical significance between any groups of particular gene families, we used the Mann-Whitney *U* test in R (http://cran.org).

Mating and meiosis genes. To determine the potential for mating and meiosis, the conservation of eight genes (see Results) specific to meiosis and of a wider set of 49 genes important for mating and meiosis (5) was examined. Of the set of 49 genes, 41 were conserved in all dermatophytes. The remaining eight include three not conserved in any of the dermatophyte species (MutA, Asd-1, and GpgA/Ste18), two short pheromone precursors that are not well conserved (PpgA, PpgB), and three genes present in only some but not all species (Rce1, SfaD/Ste4m AreB). The lack of complete conservation of these later three could reflect differences in assembly or annotation of the seven genomes.

Secondary metabolism. Prediction of the PKS and NRPS genes was made by the SMURF tool (http://www.jcvi.org/smurf/index.php) (29). An additional Blast search was run using KS and C domains in order to find sequences skipped by SMURF. The PKS and NRPS domain architecture was predicted using the InterProScan (73). Sequences were aligned with MUSCLE (68). The phylogenetic analysis was performed using PHYML (74) for the construction of the maximal likelihood tree with the Jones-Taylor-Thornton (JTT) model of the amino acid substitution. We manually inspected the gene order of clusters with the results from DAGChainer (61). We considered a difference of 40% in the number of accessory genes to be significantly different.

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

Protein kinases. The sets of eukaryotic protein kinases (kinomes) of the seven dermatophyte species analyzed in detail in this paper were identified by searching their proteins against a protein kinase HMM derived from an alignment of *Dictyostelium* protein kinases (75) using a cutoff score of −66. Low-scoring sequences were additionally screened for conservation of known protein kinase sequence motifs. The kinases thus identified were classified using the system of Hanks and Hunter (76) and Manning and coworkers (77). Classifications of dermatophyte kinases with *S. cerevisiae* orthologs were mapped from the curated data set maintained at http://www.kinase.com using OrthoMCL (60); kinases without orthologs in *S. cerevisiae* were classified by BLAST against the curated set and classified if their expected values were less than 10^{-30} and the top three hits agreed. Updated kinase classifications were provided by G. Manning (personal communication). Novel kinases were classified using maximum-likelihood trees by RAxML (70) with 1,000 bootstrap replicates for verification. To accommodate the large number of novel sequences, HMM consensus sequences of the ortholog groups were generated. Orthologous sequences were aligned using muscle (68), and HMMs were built using the hmmbuild functionality of the HMMER3 package (64). HMM consensus sequences consisting of maximum probability residues at each match state were obtained using the hmmemit functionality. Consensus sequences for outgroups were built from the six major kinase groups (AGC, CAMK, CK1, CMGC, TKL, and TK) (76) downloaded from http://www.kinase.com, and a consensus sequence was obtained from an alignment of true dermatophyte SRPK kinases. The HMM consensus sequences were combined with unclustered sequences and aligned using hmmalign, and columns with gap contents exceeding 20% were masked. For SRPK-like kinases, families were created at bootstrap support levels exceeding 50%. For other protein kinases, families containing four or more ortholog groups were created at bootstrap support levels exceeding 80%.

Nucleotide sequence accession numbers. Assemblies and annotations were submitted to GenBank under the following accession numbers: *T. rubrum* CBS 118892 (ACPH01000000), *T. tonsurans* CBS 112818 (ACPI01000000), *T. equinum* CBS127.97 (ABWI01000000), *M. canis* CBS 113480 (ABVF01000000), *M. gypseum* CBS 118893 (ABQE01000000). The *T. rubrum* ribosomal DNA was submitted to GenBank under the accession number JX431933.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00259-12/-/DCSupplemental.

Text S1, DOCX file, 0.1 MB.
Data Set S1, XLS file, 0.1 MB.
Data Set S2, XLS file, 0.1 MB.
Data Set S3, XLS file, 6.6 MB.
Figure S1, TIF file, 15.2 MB.
Figure S2, TIF file, 9.6 MB.
Figure S3, TIF file, 15.2 MB.
Figure S4, TIFF file, 2.6 MB.
Figure S5, TIF file, 20.1 MB.
Figure S6, TIF file, 14.8 MB.
Table S1, DOCX file, 0 MB.
Table S2, DOCX file, 0 MB.
Table S3, DOCX file, 0 MB.
Table S4, DOCX file, 0 MB.

ACKNOWLEDGMENTS

We acknowledge the Broad Institute Sequencing Platform for generating all DNA sequences described here, and Matthew Henn for coordinating the sequencing. We thank Mark Feuermann for discussion concerning protease annotation, Gerard Manning for updated protein kinase classifications and discussions regarding fungal kinase annotation, Peter Sudbery for discussions concerning pseudokinases and pseudophosphatases, Jian Yang for sharing TrED information regarding growth conditions, Neil Gow for discussions concerning LysM binding, Colin Jackson for discussions about rRNA, Divya Sain and Jason Stajich for sharing results of fungal cell wall protein conservation, Gerwald A. Köhler for sharing results of efflux pump protein conservation, and Antonis Rokas for comments on the manuscript.

This work was supported by the National Human Genome Research Institute and by NIH R21-AI081235 to T.C.W.

REFERENCES

1. **Achterman RR, White TC.** 2012. Dermatophyte virulence factors: identifying and analyzing genes that may contribute to chronic or acute skin infections. Int. J. Microbiol. **2012**:358305.
2. **Achterman RR, White TC.** 2012. A foot in the door for dermatophyte research. PLoS Pathog. **8**:e1002564.
3. **Gräser Y, Scott J, Summerbell R.** 2008. The new species concept in dermatophytes—a polyphasic approach. Mycopathologia **166**:239–256.
4. **White TC, Oliver BG, Gräser Y, Henn MR.** 2008. Generating and testing molecular hypotheses in the dermatophytes. Eukaryot. Cell **7**:1238–1245.
5. **Burmester A, et al.** 2011. Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biol. **12**: R7.
6. **Kwon-Chung KJ, Bennett JE.** 1992. Dermatophytoses, p 136–139. *In* Medical mycology. Lea & Febiger, Philadelphia, PA.
7. **Achterman RR, Smith AR, Oliver BG, White TC.** 2011. Sequenced dermatophyte strains: growth rate, conidiation, drug susceptibilities, and virulence in an invertebrate model. Fungal Genet. Biol. **48**:335–341.
8. **Cuomo CA, Birren BW.** 2010. The fungal genome initiative and lessons learned from genome sequencing. Methods Enzymol. **470**:833–855.
9. **Hunter S, et al.** 2009. Interpro: the integrative protein signature database. Nucleic Acids Res. **37**:D211–D215.
10. **Netea MG, Brown GD, Kullberg BJ, Gow NA.** 2008. An integrated model of the recognition of *Candida albicans* by the innate immune system. Nat. Rev. Microbiol. **6**:67–78.
11. **Cantarel BL, et al.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. **37**: D233–D238.
12. **Martinez D, et al.** 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat. Biotechnol. **26**:553–560.
13. **Desjardins CA, et al.** 2011. Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. PLoS Genet. **7**:e1002345.
14. **de Jonge R, Thomma BP.** 2009. Fungal LysM effectors: extinguishers of host immunity? Trends Microbiol. **17**:151–157.
15. **de Jonge R, et al.** 2010. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science **329**:953–955.
16. **Koharudin LM, et al.** 2011. Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus *Magnaporthe oryzae*. Structure **19**:662–674.
17. **Marshall R, et al.** 2011. Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. Plant Physiol. **156**:756–769.
18. **Gruber S, et al.** 2011. Analysis of subgroup C of fungal chitinases containing chitin-binding and LysM modules in the mycoparasite *trichoderma atroviride*. Glycobiology **21**:122–133.
19. **Freiberg C, et al.** 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. Nature **387**:394–401.
20. **Yu JH, Keller N.** 2005. Regulation of secondary metabolism in filamentous fungi. Annu. Rev. Phytopathol. **43**:437–458.
21. **Zaugg C, et al.** 2009. Gene expression profiling in the human pathogenic dermatophyte *Trichophyton rubrum* during growth on proteins. Eukaryot. Cell **8**:241–250.
22. **Wang L, et al.** 2006. Analysis of the dermatophyte *Trichophyton rubrum* expressed sequence tags. BMC Genomics **7**:255.
23. **Rawlings ND, Barrett AJ, Bateman A.** 2012. Merops: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. **40**:D343–D350.
24. **Muszewska A, Taylor JW, Szczesny P, Grynberg M.** 2011. Independent subtilase expansions in fungi associated with animals. Mol. Biol. Evol. **28**:3395–3404.
25. **Jousson O, et al.** 2004. Multiplication of an ancestral gene encoding

Downloaded from mbio.asm.org on September 4, 2012 - Published by mbio.asm.org

secreted fungalysin preceded species differentiation in the dermatophytes *Trichophyton* and *Microsporum*. Microbiology **150**:301–310.
26. **Zaugg C, Jousson O, Léchenne B, Staib P, Monod M.** 2008. *Trichophyton rubrum* secreted and membrane-associated carboxypeptidases. Int. J. Med. Microbiol. **298**:669–682.
27. **Sriranganadane D, et al.** 2011. Identification of novel secreted proteases during extracellular proteolysis by dermatophytes at acidic pH. Proteomics **11**:4422–4433.
28. **Chiang YM, Oakley BR, Keller NP, Wang CC.** 2010. Unraveling polyketide synthesis in members of the genus *Aspergillus*. Appl. Microbiol. Biotechnol. **86**:1719–1736.
29. **Khaldi N, et al.** 2010. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet. Biol. **47**:736–741.
30. **Zhang H, Rokas A, Slot JC.** 2012. Two different secondary metabolism gene clusters occupied the same ancestral locus in fungal dermatophytes of the *Arthrodermataceae*. PLoS One **7**:e41903.
31. **Manning G, Plowman GD, Hunter T, Sudarsanam S.** 2002. Evolution of protein kinase signaling from yeast to man. Trends Biochem. Sci. **27**: 514–520.
32. **Rhind N, et al.** 2011. Comparative functional genomics of the fission yeasts. Science **332**:930–936.
33. **Stajich JE, et al.** 2010. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereous*). Proc. Natl. Acad. Sci. U. S. A. **107**:11889–11894.
34. **Reese ML, Boyle JP.** 2012. Virulence without catalysis: how can a pseudokinase affect host cell signaling? Trends Parasitol. **28**:53–57.
35. **Hsueh YP, Heitman J.** 2008. Orchestration of sexual reproduction and virulence by the fungal mating-type locus. Curr. Opin. Microbiol. **11**: 517–524.
36. **Li W, Metin B, White TC, Heitman J.** 2010. Organization and evolutionary trajectory of the mating type (MAT) locus in dermatophyte and dimorphic fungal pathogens. Eukaryot. Cell **9**:46–58.
37. **Xu J, et al.** 2007. Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. Proc. Natl. Acad. Sci. U. S. A. **104**:18730–18735.
38. **Ashbee HR.** 2007. Update on the genus Malassezia. Med. Mycol. **45**: 287–303.
39. **Ferreira ME, et al.** 2005. The ergosterol biosynthesis pathway, transporter genes, and azole resistance in *Aspergillus fumigatus*. Med. Mycol. **43**(Suppl 1):S313–S319.
40. **Vermout S, et al.** 2008. Pathogenesis of dermatophytosis. Mycopathologia **166**:267–275.
41. **Almeida SR.** 2008. Immunology of dermatophytosis. Mycopathologia **166**:277–283.
42. **Vega K, Kalkum M.** 2012. Chitin, chitinase responses, and invasive fungal infections. Inter. J. Microbiol. 2012:920459.
43. **Bolton MD, et al.** 2008. The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. Mol. Microbiol. **69**:119–136.
44. **Mentlak TA, et al.** 2012. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. Plant Cell **24**:322–335.
45. **Rappleye CA, Goldman WE.** 2008. Fungal stealth technology. Trends Immunol. **29**:18–24.
46. **Fontaine T, et al.** 2011. Galactosaminogalactan, a new immunosuppressive polysaccharide of *Aspergillus fumigatus*. PLoS Pathog. **7**:e1002372.
47. **Stergiopoulos I, et al.** 2012. In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. Mol. Biol. Evol.
48. **Galagan JE, et al.** 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature **438**:1105–1115.
49. **Nierman WC, et al.** 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature **438**: 1151–1156.
50. **O'Gorman CM, Fuller HT, Dyer PS.** 2009. Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. Nature **457**: 471–474.
51. **Anzawa K, Kawasaki M, Mochizuki T, Ishizaki H.** 2010. Successful mating of *Trichophyton rubrum* with *Arthroderma simii*. Med. Mycol. **48**: 629–634.
52. **Staib P, et al.** 2010. Differential gene expression in the pathogenic dermatophyte *Arthroderma benhamiae* in vitro versus during infection. Microbiology **156**:884–895.
53. **Xu Z, Wang H.** 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**:W265–W268.
54. **Ellinghaus D, Kurtz S, Willhoeft U.** 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics **9**:18.
55. **Li W, Godzik A.** 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22**: 1658–1659.
56. **Rice P, Longden I, Bleasby A.** 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. **16**:276–277.
57. **Salamov AA, Solovyev VV.** 2000. Ab initio gene finding in *Drosophila* genomic DNA. Genome Res. **10**:516–522.
58. **Parra G, Blanco E, Guigó R.** 2000. GeneID in *Drosophila*. Genome Res. **10**:511–515.
59. **Borodovsky M, Lomsadze A, Ivanov N, Mills R.** 2003. Eukaryotic gene prediction using GeneMark.hmm. Curr. Protoc. Bioinformatics Chapter 4:Unit4.6. PubMed.
60. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.
61. **Haas BJ, Delcher AL, Wortman JR, Salzberg SL.** 2004. DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics **20**:3643–3646.
62. **Eisenhaber B, Schneider G, Wildpaner M, Eisenhaber F.** 2004. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. J. Mol. Biol. **337**:243–253.
63. **Storey JD, Tibshirani R.** 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. **100**:9440–9445.
64. **Eddy SR.** 2011. Accelerated profile HMM searches. PLoS Comput. Biol. **7**:e1002195.
65. **Kelley LA, Sternberg MJ.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4**:363–371.
66. **Humphrey W, Dalke A, Schulten K.** 1996. VMD: visual molecular dynamics. J. Mol. Graph. **14**:33–38.
67. **Russell RB, Barton GJ.** 1992. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. Proteins **14**:309–323.
68. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.
69. **Abascal F, Zardoya R, Posada D.** 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**:2104–2105.
70. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688–2690.
71. **Paradis E, Claude J, Strimmer K.** 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics **20**:289–290.
72. **Yang Z, Bielawski JP.** 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**:496–503.
73. **Mulder N, Apweiler R.** 2007. Interpro and InterProScan: tools for protein sequence classification and comparison. Methods Mol. Biol. **396**:59–70.
74. **Guindon S, Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**: 696–704.
75. **Goldberg JM, et al.** 2006. The dictyostelium kinome—analysis of the protein kinases from a simple model organism. PLoS Genet. **2**:e38.
76. **Hanks SK, Hunter T.** 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J **9**:576–596.
77. **Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S.** 2002. The protein kinase complement of the human genome. Science **298**: 1912–1934.