Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens

Jun Xu*, Charles W. Saunders*, Ping Hu*, Raymond A. Grant*, Teun Boekhout[†‡], Eiko E. Kuramae[†§], James W. Kronstad[¶], Yvonne M. DeAngelis*, Nancy L. Reeder*, Kevin R. Johnstone*, Meredith Leland*, Angela M. Fieno*, William M. Begley*, Yiping Sun*, Martin P. Lacey*, Tanuja Chaudhary*, Thomas Keough*, Lien Chu[‖], Russell Sears**[††], Bo Yuan**[‡‡], and Thomas L. Dawson, Jr.*[§§]

*Procter & Gamble Co., Miami Valley Innovation Center, Cincinnati, OH 45253-8707; [†]Centraalbureau voor Schimmelcultures, Fungal Biodiversity Centre, 3584 CT, Utrecht, The Netherlands; [‡]Division of Acute Medicine and Infectious Diseases and Eijkman–Winkler Centre for Microbiology, Infectious Diseases and Inflammation, University Medical Centre, 3508 GA, Utrecht, The Netherlands; [§]Netherlands Institute of Ecology, Centre for Terrestrial Ecology, 6666 ZG, Heteren, The Netherlands; [¶]The Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC, Canada V6T 1Z4; [‖]Integrated Genomics, 2201 West Campbell Park Drive, Suite 15, Chicago, IL 60612-3547; and **Department of Biomedical Informatics and Pharmacology, Ohio State University, Columbus, OH 43210

Edited by Joan Wennstrom Bennett, Rutgers, The State University of New Jersey, New Brunswick, NJ, and approved October 1, 2007 (received for review July 18, 2007)

Fungi in the genus *Malassezia* are ubiquitous skin residents of humans and other warm-blooded animals. *Malassezia* are involved in disorders including dandruff and seborrheic dermatitis, which together affect >50% of humans. Despite the importance of *Malassezia* in common skin diseases, remarkably little is known at the molecular level. We describe the genome, secretory proteome, and expression of selected genes of *Malassezia globosa*. Further, we report a comparative survey of the genome and secretory proteome of *Malassezia restricta*, a close relative implicated in similar skin disorders. Adaptation to the skin environment and associated pathogenicity may be due to unique metabolic limitations and capabilities. For example, the lipid dependence of *M. globosa* can be explained by the apparent absence of a fatty acid synthase gene. The inability to synthesize fatty acids may be complemented by the presence of multiple secreted lipases to aid in harvesting host lipids. In addition, an abundance of genes encoding secreted hydrolases (e.g., lipases, phospholipases, aspartyl proteases, and acid sphingomyelinases) was found in the *M. globosa* genome. In contrast, the phylogenetically closely related plant pathogen *Ustilago maydis* encodes a different arsenal of extracellular hydrolases with more copies of glycosyl hydrolase genes. *M. globosa* shares a similar arsenal of extracellular hydrolases with the phylogenetically distant human pathogen, *Candida albicans*, which occupies a similar niche, indicating the importance of host-specific adaptation. The *M. globosa* genome sequence also revealed the presence of mating-type genes, providing an indication that *Malassezia* may be capable of sex.

fungal genomics | fungal proteomics | seborrheic dermatitis | skin | fungal mating

For >100 years, *Malassezia* have been associated with dandruff (1), an easily recognizable skin flaking condition occurring in 30–95% of humans (2). The presence of *Malassezia* is not sufficient to cause either dandruff or the more extreme skin condition, seborrheic dermatitis; in fact, many people harbor *Malassezia* without showing symptoms. However, *Malassezia* must have an essential role in these conditions, because scalp-flaking symptoms are improved by treatment with a variety of antifungal materials that remove *Malassezia* (1, 3–5). These fungi also cause systemic infections in neonates (6–8) and the skin disease pityriasis versicolor (1, 3, 5), as well as many common veterinary disorders, such as canine dermatitis and otitis externa (9). *Malassezia* are also thought to contribute to the common skin disease atopic eczema by host sensitization to fungal protein allergens (10).

Despite their widespread occurrence and association with multiple common skin disorders, remarkably little is known about these fungi other than their unusual dependence on external lipids for *in vitro* growth (1). To understand the role of *Malassezia* in dandruff, we have focused on *Malassezia globosa* and *Malassezia restricta*, because they are the most commonly isolated species from human scalp (3, 11) and are associated with prevalent skin disorders, including dandruff (4).

We describe the genome sequence, secretory proteome, and *in vivo* expression of selected genes of *M. globosa* and a partial genome sequence and secretory proteome of *M. restricta.* This report highlights a conflict between phylogeny and host-specific adaptation, because genome- and proteome-based evidence indicates that the basidiomycete *Malassezia* species share similar sets of extracellular hydrolases with the phylogenetically distant *Candida albicans* that occupies an overlapping skin-related niche. In contrast, *Malassezia* is closely related to the plant pathogen *Ustilago maydis*, but these fungi are easily distinguished by differences in their sets of extracellular hydrolases and their host preferences.

Results and Discussion

Whole-genome shotgun sequencing (7× coverage) was performed on *M. globosa* type strain CBS 7966 [supporting information (SI) Table 2]. The assembled genome size of 9 Mb closely matches the value reported by Boekhout *et al.* (12). The genome size is among the smallest of the sequenced genomes of free-

Author contributions: J.X. and C.W.S. contributed equally to this work; C.W.S. and T.L.D. designed research; J.X., C.W.S., R.A.G., T.B., Y.M.D., N.L.R., K.R.J., M.L., A.M.F., W.M.B., Y.S., M.P.L., T.C., T.K., L.C., and T.L.D. performed research; J.X., C.W.S., P.H., E.E.K., R.S., and B.Y. analyzed data; and J.X., C.W.S., P.H., R.A.G., T.B., J.W.K., and T.L.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank database [accession nos. AAYY00000000 (*M. globosa*) and AAXK00000000 (*M. restricta*)]. The versions described in this paper are the first versions, AAYY01000000 and AAXK01000000.

[††]Present address: Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720.

[‡‡]Present address: Department of Computer Science, Institute of Biocomputing and Systems Biology, Shanghai Jiaotong University, Shanghai 200290, China.

[§§]To whom correspondence should be addressed. E-mail: dawson.tl@pg.com.

This article contains supporting information online at www.pnas.org/cgi/content/full/0706756104/DC1.

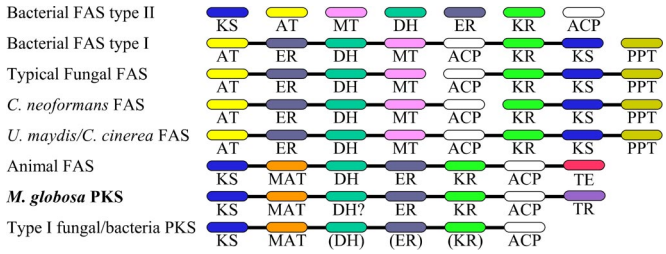

Fig. 1. *M. globosa* does not have a traditional fungal fatty acid synthase. Ovals represent fatty acid synthase domains, with black connecting lines representing covalent connections, as predicted from the DNA sequences. FAS, fatty acid synthase; KS, β-ketoacyl synthase; AT, acetyl-CoA-ACP transacetylase; MT, malonyl-CoA-ACP-transacylase; DH, dehydratase; ER, β-enoyl reductase; KR, β-ketoacyl reductase; ACP, acyl carrier protein; PPT, phosphopantetheinyl transferase; MAT, malonyl-CoA-acetyl-CoA-ACP-transacylase; TE, thioesterase; PKS, polyketide synthase; TR, thioester reductase. The ''?'' next to DH activity of *M. globosa* PKS indicates the uncertainty of the activity based on homology alone. A domain name in parentheses represents a domain included in some, but not all, of this set of enzymes. Shown here is one example of type I PKS; many variations of this architecture exist. The two FAS subunits of *U. maydis* and several other fungi appear to be fused into a single polypeptide. See SI Tables 6–8 for more information.

living fungi (SI Table 3), similar in size to the *Eremothecium* (*Ashbya*) *gossypii* genome (13, 14). *M. globosa* CBS 7966 appears to be haploid, because the genome shows essentially no polymorphisms (<0.0004%) within regions of high-quality sequence. Gene models for *M. globosa* were constructed by using *in silico* predictions and sequence data from an EST library constructed from cultured *M. globosa* cells. We predicted 4,285 protein-coding genes, again among the lowest value previously reported for any genome of a free-living fungus (SI Table 4). Introns were found in only 27% of the genes (SI Table 2), a value similar to that seen in *U. maydis* and lower than that observed for other basidiomycete fungi (15).

Whole-genome shotgun sequencing (1× coverage) was also carried out on *M. restricta* type strain CBS 7877 (SI Table 2). Although this was insufficient coverage to assemble large sections of the *M. restricta* genome, this sequence did allow a comparison of gene content. There were many similarities in the genomes, and only three genes were found in the *M. restricta* genome that were missing in the *M. globosa* genome sequence (*SI Text*). *M. globosa* and *M. restricta* showed an average 77% amino acid sequence identity among reciprocal best-matched proteins.

The *M. globosa* genome encodes the basic metabolic components for proficiency in glycolysis, the tricarboxylic acid (TCA) cycle, the glyoxylate cycle, the pentose phosphate shunt, the synthesis of all 20 canonical amino acids, and the synthesis of the five bases typically found in nucleic acids. The notable predicted metabolic deficiency is the apparent absence of a gene encoding fatty acid synthase, a deficiency apparently shared by *M. restricta* (SI Table 5) and a likely reason why these *Malassezia* species depend on fatty acids for growth. *M. globosa* and *M. restricta* do contain one gene that resembles an animal fatty acid synthase, but careful examination revealed a more likely function as a polyketide synthase (*SI Text, Section IV*). The domain organization is similar to fungal polyketide synthases and animal fatty acid synthases and different from that of fungal fatty acid synthases (Fig. 1). A similarity search revealed that 68 of the 100 most-similar proteins are annotated as polyketide synthases (SI Table 6). In contrast, a similarity search with a *U. maydis* putative fatty acid synthase revealed many fatty acid synthases and no polyketide synthases among the 100 most similar proteins. Furthermore, the thioesterase domain in previously reported animal fatty acid synthases is replaced in the *M. globosa* polypeptide with a thioester reductase domain found in many nonribosomal peptide synthetases and polyketide synthases (16) (SI Table 7). In contrast to the situation for the *Malassezia* species, the available genomes of all other free-living fungi contain fatty acid synthase genes (SI Table 8).

The need for *Malassezia* species to assimilate fatty acids from external sources is also reflected in the presence of multiple genes for secreted lipases and phospholipases (Table 1). The *M. globosa* genome encodes 14 lipases, 13 of which are predicted to be secreted. Twelve of the secretory lipases can be divided into two families represented by the PFAM categories LIP (PF03583) and Lipase 3 (PF01764); the latter contains the previously described *M. globosa* enzyme LIP1 (17). There are also genes for nine phospholipases, including six phospholipase C enzymes that are predicted to be secreted. RT-PCR and proteomics experiments were performed to confirm the expression of the lipase and phospholipase genes because of their relevance to the lipophilic lifestyle. The majority of the genes for the predicted lipases and phospholipases were expressed on human scalp, as demonstrated by RT-PCR (Fig. 2*A*; data not shown). Thus, these genes are expressed when the fungus is at the location where it would be expected to use host lipids.

Of course, the enzymes would need to be extracellular to mediate the release of fatty acids from host sources. We therefore performed proteomics experiments to identify the proteins secreted by *M. globosa* into the culture medium and/or bound to the cell surface. In a representative experiment (Fig. 2*B*), we detected a lipase, two aspartyl proteases, two glucose methanol choline oxidoreductases, two cell wall-modifying enzymes, four proteins related to known *Malassezia* allergens (1), and several unknown proteins. From this and 11 other experiments in which growth and extraction conditions were varied, we identified >50 secreted proteins (SI Tables 9 and 10). As hypothesized, among these secreted proteins were eight lipases and three phospholipases, thus confirming the extracellular location of these lipid-degrading enzymes. The *M. restricta* genome also contains multiple genes for each of these six sets of proteins, but the limited sequence coverage prevented quantification of the numbers of genes. Because these proteins are secreted, they would be the most likely to interact with skin and would therefore likely mediate *Malassezia* pathogenic processes and be relevant therapeutic targets.

We used the available whole-genome sequences for several fungi to perform a phylogenetic analysis with the goal of identifying evolutionary relationships for *M. globosa* and *M. restricta*. This analysis revealed that, as reported using ribosomal RNA-encoding DNA region sequences (18), *M. globosa* is most closely related to the maize pathogen *U. maydis* among the fungi with complete genome sequences (15) (Fig. 3*A*; SI Fig. 4 and SI Table 11). This close relationship is striking, given the remarkably different host preferences. A comparison of the reciprocal best-matched proteins encoded by the two genomes revealed they show an average of 52% amino acid sequence identity. Despite this high level of similarity, there is limited conservation of synteny. For example, only 15% of adjacent genes in *M. globosa* have adjacent orthologs in *U. maydis*, and <3% of the genes in *M. globosa* appear in *U. maydis* in synteny groups longer than two genes.

The similarities with *U. maydis* extend to the genes responsible for mating. *Ustilago* species contain an *a* locus encoding pheromones and pheromone receptors and a *b* locus encoding homeodomain transcription factors (19). *M. globosa* contains genes resembling those at the *a* and *b* loci, and the genes share organizational similarity with the linked mating-type genes of the bipolar heterothallic species *Ustilago hordei* (20). That is, the putative pheromone and pheromone receptor genes of *M. globosa* occur in the same supercontig as the homeodomain-containing transcription factor genes with similarity to *bW* and *bE* of *U. hordei* (Fig. 3*B*). We hypothesize that these genes define

MICROBIOLOGY

Table 1. Comparison of enzyme classes among several fungal species

Enzyme	Gene copies (predicted secreted proteins)				
	M. globosa	*U. maydis*	*C. albicans*	*C. neoformans*	*S. cerevisiae*
Lipase	14(13)	7(2)	14(12)	3(1)	3(0)
Phospholipase C	6(6)	0(0)	0(0)	0(0)	0(0)
Phospholipase D	2(0)	3(0)	2(0)	4(0)	2(0)
Phospholipase B	1(0)	1(1)	7(5)	1(1)	4(4)
Acid sphingomylinase	4(4)	1(1)	4(3)	1(1)	0(0)
Aspartyl protease	18(15)	10(7)	15(15)	9(6)	8(7)
Cysteine protease	7(1)	9(0)	9(1)	9(1)	8(1)
Metalloprotease	31(3)	35(7)	42(3)	44(5)	38(5)
Serine protease	10(2)	17(9)	15(10)	18(10)	13(8)
Glycosyl hydrolase	15(3)	65(36)	45(27)	52(27)	38(24)
Cutinase	0(0)	4(4)	0(0)	0(0)	0(0)
Pectin lyase	0(0)	1(0)	0(0)	0(0)	0(0)
Pectin esterase	0(0)	1(1)	0(0)	0(0)	0(0)
Polyketide synthase	1(0)	3(0)	0(0)	0(0)	0(0)
Total gene number	4,285	6,902	7,677	6,810	6,604

The number of genes within the indicated haploid genomes that encode a selected set of enzymes are shown to illustrate differences in coding capacity. The entry is highlighted in pink for the species with the largest number of genes for each category. The number of proteins that are predicted to be secreted, on the basis of a predicted signal sequence, using SignalP3.0 (37), are listed in parentheses for each category. The phospholipase C entry refers to the secretory enzyme similar to *Pseudomonas aeruginosa* phospholipase C and not the more commonly discussed cytoplasmic phosphoinositol-specific phospholipase C. Detailed methods for the construction of the table are described in *SI Text*. A comparison of *M. globosa* and *C. albicans* aspartyl proteases is depicted in SI Table 13.

a *MAT* locus of ≈170 kb in *M. globosa*. There was no indication of alternate mating-type loci in the genome of the sequenced strain CBS 7966 of *M. globosa*, again suggesting that this strain is haploid, and that mating-type switching does not occur. Several fragments of putative mating-related genes were also found in *M. restricta: pra1* (MRE_contig_0425), *bE1* (MRE_contig_2721), and *bW1* (MRE_contig_3923 and _1863).

The 170-kb *MAT* region of *M. globosa* does not contain the high density of repetitive elements or possess orthologs of the other genes found at the 527-kb *MAT* locus of *U. hordei* (20). Furthermore, a comparison of the 82 genes predicted for the *MAT* locus of *M. globosa* with the genes found at the *MATα* locus of another basidiomycete fungus, *Cryptococcus neoformans*, revealed that these loci share orthologs of the pheromone receptor gene, as expected, and both contain an ortholog of the *CID1* gene thought to play a role in the cell cycle (21). The significance of the shared possession of this gene at the *MAT* locus is not clear but may have implications for the evolution of basidiomycete mating type loci. The *MAT* locus of *C. neoformans* is remarkable, because it encodes components of the pheromone response pathway (21); this is not the case for the *MAT* locus of *M. globosa*.

The presence of candidate genes for mating in the *Malassezia* genomes is interesting, because mating has never been observed in *Malassezia* species. The presence of these genes suggests that the fungus may be capable of sexual development and that the characterization of the *MAT* region in other strains may identify one or more additional mating specificities for use in functional mating tests. We used the genome sequence to investigate the possibility of sexual competency in more detail by searching for orthologs of genes known to function in pheromone signaling,

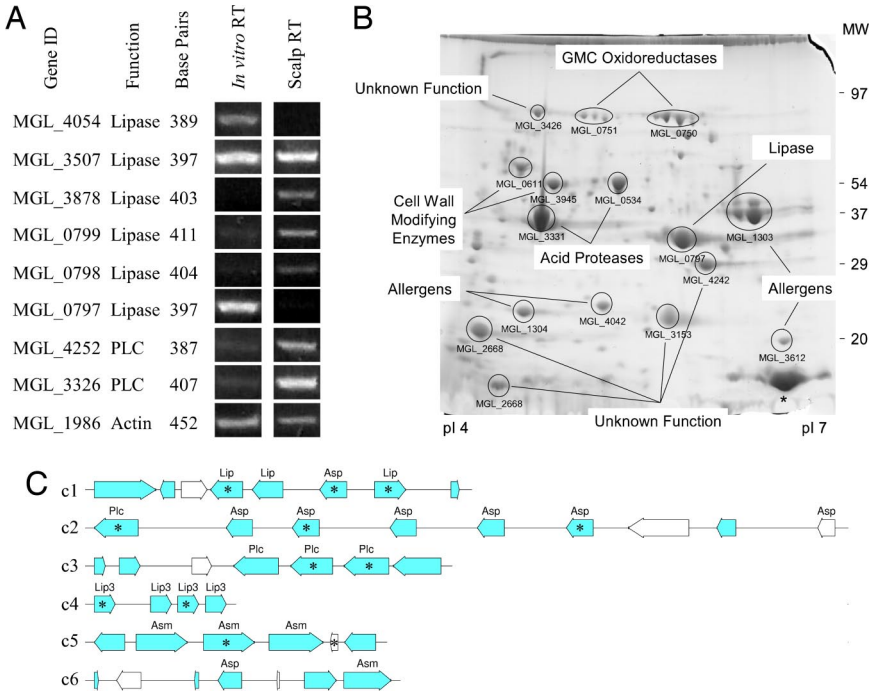

Fig. 2. Secreted proteins: gene expression and gene clusters. (*A*) Transcripts for lipase and phospholipase genes were detected from *M. globosa* cells isolated from scalp and from cells cultured *in vitro*. Isolated RNA was used for first-strand cDNA synthesis with reverse transcriptase and random hexamers and then PCR-amplified by using gene-specific primers, as described (17). (*B*) The 2D gel was stained with Coomassie blue. Multiple spots within a circle are different isoforms of the same protein. The dark spot in the bottom right corner of the gel (*) was resistant to multiple attempts at digestion and could not be identified. See also SI Tables 9 and 10. (*C*) Cyan symbols, proteins predicted to be secreted. Open symbols, proteins predicted not to be secreted. * in the middle of the gene indicates the protein was found in the extracellular fraction in the proteomics experiments. Lip, lipase family LIP; Lip3, lipase family LIP3; Plc, phospholipase C; Asp, aspartyl protease; Asm, acid sphingomyelinase. C1–6 refer to gene cluster names, with the contig numbers, gene numbers, and predicted gene functions listed in SI Table 12.

meiosis and sporulation in other fungi. This analysis revealed that the *M. globosa* genome encodes candidate orthologs for many of these functions. For example, orthologs were found for components of the *Saccharomyces cerevisiae* pheromone responsive MAPK module encoded by *STE11*, *STE7*, and *FUS3,* and for the *GPA1*- and *STE4*-encoded components of the heterotrimeric G protein complex that couples pheromone receptor signaling with the MAPK module (22). Overall, this analysis supports the idea that *M. globosa* is capable of sexual reproduction. The possibility of a sexual cycle in *Malassezia* yeasts is intriguing, because it may have a profound effect on the distribution of virulence-related genes among populations (23). Furthermore, the presence of a sexual state may open possibilities for classical genetic experimentation. This situation is reminiscent of the discovery of mating in *C. albicans* as a consequence of genomic evidence for mating type genes (24), and it is interesting that *C. albicans* mating is enhanced on skin (25).

There are important differences in gene content that may be linked to the adaptation of *M. globosa* and *U. maydis* to their specific host environments (Fig. 3*C*). Overall, *U. maydis* contains 2,536 genes with no apparent homologs in *M. globosa* (at $e < 10^{-5}$). The most striking difference is the presence of more carbohydrate-hydrolyzing proteins (glycosyl hydrolases) encoded by *U. maydis* than by *M. globosa* (Table 1). In particular, xylanase, cellulase, pectin esterase, polygalacturonase, pectin lyase, α-galactosidase, and α-amylase genes are present in *U. maydis* but not in *M. globosa*. *M. globosa* has only 325 genes that are missing in *U. maydis*, and most of these genes have unknown functions. The most prominent difference in genes with assignable function is that *M. globosa* encodes more lipases, aspartyl proteases, phospholipase C enzymes, and acid sphingomyelinases, as described above (Table 1). Many of these *M. globosa* genes are clustered in the genome (Fig. 2*C* and SI Table 12). It is remarkable that clusters of genes for secreted proteins are also found in the *U. maydis* genome, although these proteins appear to function in plant-specific interactions, such as defense suppression and tumor induction (15). In general, the different sets of secreted proteins likely reflect the specificity in the host molecules targeted by the fungi.

The abundance of genes for phospholipases, lipases, aspartyl proteases, and acid sphingomyelinases in *M. globosa* is also seen in *C. albicans* (26) (Table 1), a distantly related (SI Fig. 4) human pathogenic and commensal fungus that also colonizes skin and mucosal surfaces (27). In contrast, these specific genes are not prevalent in the genome of the fermentative yeast *S. cerevisiae* (Table 1). The *C. albicans* genes encoding aspartyl proteases and phospholipase B are implicated in virulence in animal models (27, 28). The prevalence of similar gene families in two fungal species, *M. globosa* and *C. albicans*, that are opportunists on human skin may point to similar roles in colonization and virulence, although *C. albicans* does not provoke the specific skin disorders associated with *Malassezia* species.

Lipases, phospholipases, and allergens have been proposed to be important in *Malassezia* virulence, although definitive genetic tests have not been made, because gene disruption has not yet been demonstrated in *Malassezia* species. The presence of phospholipase activity correlates with the virulence of *Malassezia pachydermatis* in dogs (29). There is a model linking *Malassezia* lipase activity to the pathogenesis of dandruff and seborrheic dermatitis (17). *Malassezia* antigens may play a role in atopic eczema (1), and the *M. globosa* genome encodes proteins similar to 12 of 13 *Malassezia sympodialis* and *Malassezia furfur* proteins identified as reacting with IgE from atopic eczema patients (10), extending previous observations based on PCR analysis for a

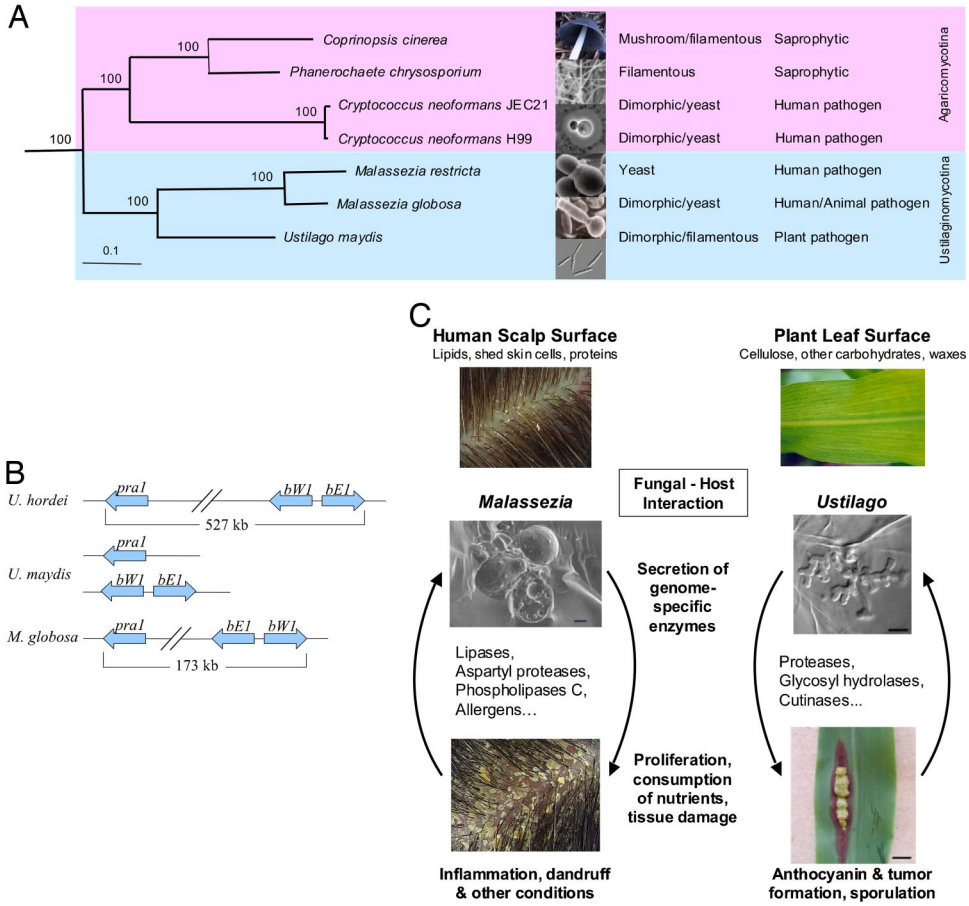

Fig. 3. Similarities of phylogeny, mating type gene organization, and pathogenesis between *Malassezia* and *Ustilago*. (*A*) The phylogenetic tree based on 28 concatenated single-copy orthologous proteins present in 34 fungal genomes was prepared by using Maximum Likelihood (PHYML) by using JTT amino acid model substitution (SI Table 11). This tree depicts a small number of fungi; a more inclusive tree is shown (SI Fig. 4). (*B*) Schematic representation of the mating loci of *U. hordei*, *U. maydis*, and *M. globosa*. *pra1*: pheromone receptor 1; *bW1*: homeodomain transcription factor b west 1; *bE1*: homeodomain transcription factor b east 1. *M. globosa* mating-related genes: putative pheromone, MGL_0963; *pra1*, MGL_0964; *bE1*, MGL_0884; *bW1*, MGL_0883. (*C*) Illustration of the adaptations by *Malassezia* and *Ustilago* to take advantage of distinct host environments for survival. The model emphasizes the most prominent protein families based on the available data. The downward-pointing arrows represent the damage to the host substrate inflicted by secreted proteins, and the upward-pointing arrows represent the damaged host substrate that provides resources to support the growth and persistence of the fungal population. [Scale bars: 2 μm (*Malassezia*), 10 μm (*Ustilago*), and 1 cm (tumor formation).]

subset of these allergen genes (30). Our discoveries of multicopy genes, *in vitro* protein production and secretion, and (in some cases) gene expression on the scalp provide support for the importance of several *M. globosa* gene families. These include the gene families whose importance has already been suggested, such as lipase, phospholipase C, and allergens, but also newly discovered families, such as aspartyl protease, glucose methanol choline oxidoreductase, and acid sphingomyelinase. These discoveries should facilitate the testing of the role of these protein families in skin diseases such as dandruff and seborrheic dermatitis.

Our report of *Malassezia* genome sequences is part of a growing understanding of the adaptations of human fungal pathogens. Because *Cryptococcus* and *Malassezia* are the only two basidiomycete fungi to infect humans commonly, there is interest in further comparison of their virulence traits. Of additional interest is the study of adaptation to animal skin that has apparently emerged independently in three lineages: the basidiomycete *Malassezia* spp. (Ustilaginomycotina), the ascomycete *C. albicans* (Saccharomycotina), and the true dermatophyte ascomycetes such as *Trichophyton* and *Microsporum* spp. (Pezizomycotina).

Summary

Overall, our genomic comparisons provide insights into the mechanisms by which fungi adapt to the mammalian skin environment. The phylogenetic position of *Malassezia* indicates they are closely related to plant pathogens, implying an ancestral shift from plant to animal host preference. This report describes the features of the *M. globosa* genome and some characteristics that may have facilitated this proposed host shift. The *M. globosa* genome is among the smallest of genomes of free-living fungi (SI Tables 2–4 and 14). One key insight is the apparent absence of a fatty acid synthase gene and compensation by the abundance of genes encoding hydrolases that could provide fatty acids. An additional feature of this adaptation to human skin is the presence of a set of secreted hydrolases similar to the set produced by the distantly related human pathogen *C. albicans*. Moreover, the secreted hydrolases are quite different between *M. globosa* and the phylogenetically related plant pathogen *U. maydis*.

Because *Malassezia*-associated skin diseases are relatively accessible for study, these diseases could become models to study the interaction of host and pathogen. With this report of the genome sequence of *M. globosa* and partial genome sequence of

M. restricta, coupled with the human genome sequence, this interaction can be studied at the molecular level, an approach formerly impossible because of the shortage of sequence information for *Malassezia* spp. For example, there are only four protein-coding genes or gene segments from each of *M. globosa* and *M. restricta* in the current National Center for Biotechnology Information nr database. A deeper understanding of these interactions may well lead to an understanding of the conditions that lead to skin disease and new treatment paradigms to alleviate the major impact that *Malassezia* species have on human and animal health.

Materials and Methods

Strains. *M. globosa* CBS 7966 and *M. restricta* CBS 7877 were obtained from the Centraalbureau voor Schimmelcultures Fungal Biodiversity Centre.

EST Sequencing, Genome Sequencing, and Annotation. An EST library was created from RNA isolated from *M. globosa* cells grown in mDixon medium (17). *M. globosa* DNA was sequenced by using a whole-genome shotgun approach (7×), and 2,000 finishing reactions were performed to fill gaps and fix low-quality regions. The genome was assembled by using Phrap. Gene models were constructed by combining sequencing information from 1,536 EST sequences and *in silico* predictions from TWINSCAN (31), GeneID (32), and GeneWise (33) (*SI Text*). Putative gene function assignments were obtained through homology searches against the National Center for Biotechnology Information nr, PFAM (34), and TIGRFAM (35) databases. *M. restricta* DNA was sequenced to 1× coverage and assembled by using Phrap. Gene models were constructed by aligning predicted *M. globosa* proteins and proteins from the nr database to the *M. restricta* genome by using EXONERATE (36). More details are provided in *SI Text*.

Prediction of Secreted Proteins. SignalP 3.0 (37) was used to predict N-terminal secretion signals. The intersection (347 genes) between the hidden Markov and neural network prediction was used to indicate candidates for secreted proteins.

Isolation of Cell Surface-Associated Proteins. Cells grown in mDixon media (17) were pelleted by centrifugation at 4,000 × *g* and washed three times with TE buffer [50 mM Tris·HCl (pH 8.0)/5 mM EDTA containing protease inhibitors (Roche, Indianapolis, IN)]. The cell pellet was resuspended in TE buffer, and cells were washed further by pelleting through a cushion of 20% sucrose at 10,000 × *g*. The washed cell pellets were resuspended in a buffer containing 100 mM EDTA, 300 mM 2-mercaptoethanol (one part cells to two parts buffer) and then gently mixed for 1 h at room temperature. Cells were removed by centrifugation, and the extracted proteins in the supernatant were dialyzed [Snake-Skin dialysis tubing with 7000 molecular weight cutoff (Pierce, Rockford, IL)] for 24 h at 4°C against 0.1% acetic acid, lyophilized, and solubilized in isoelectric focusing sample buffer. Proteins were focused by using pH 4–7 strips (GE Healthcare, Piscataway, NJ) and separated on 10–20% gradient gels (Invitrogen, Carlsbad, CA). Details of the 2D gel electrophoresis procedure and the identification of proteins by mass spectrometry have been described (17). The amino acid sequences were used to identify the corresponding genes.

We thank Gil Cloyd, Melissa Monich, Bob LeBoeuf, Jeff Hamner, Eric Armstrong, Bob Lindenschmidt, Chris Kelling, Pete Ellingson, John McIver, Charlie Bascom, Kenny Morand, Mike Manhart, Cathy Oppenheimer, and Noelle Fitzgerald for support; Eveline Guého for photographs of *Malassezia* (Fig. 3*A*); Jana Klose for photographs of *Ustilago*; George Smulian, Fred Dietrich, Pamela Trotter, Aditya Gupta, Jacques Guillot, Carlos Echavarri, and Jim Schwartz for technical discussions; Lisa Bosch and Diane Vonderheight for help with manuscript preparation; Vincent Robert for use of parts of the Biolomics software; Integrated Genomics for *M. globosa* sequencing reactions; and Agencourt (Beverly, MA) for sequencing the *M. restricta* genomic DNA and the EST library. Eiko Kuramae was supported by a KNAW renewal fund grant.

1. Chen TA, Hill PB (2005) *Vet Dermatol* 16:4–26.
2. Schwartz JR, Cardin CW, Dawson TL (2004) in *Textbook of Cosmetic Dermatology*, eds Baran R, Maibach HI (Martin Dunitz, Ltd, London), pp 259–272.
3. Ashbee HR, Evans EG (2002) *Clin Microbiol Rev* 15:21–57.
4. Gupta AK, Batra R, Bluhm R, Boekhout T, Dawson TL, Jr (2004) *J Am Acad Dermatol* 51:785–798.
5. Midgley G (2000) *Med Mycol 38 Suppl* 1:9–16.
6. Batra R, Boekhout T, Gueho E, Cabanes FJ, Dawson TL, Jr, Gupta AK (2005) *FEMS Yeast Res* 5:1101–1113.
7. Chang HJ, Miller HL, Watkins N, Arduino MJ, Ashford DA, Midgley G, Aguero SM, Pinto-Powell R, von Reyn CF, Edwards W, *et al.* (1998) *N Engl J Med* 338:706–711.
8. Crespo Erchiga V, Guého E (2005) in *Topley and Wilson's Microbiology and Microbial Infections, Medical Mycology*, eds Merz WG, Hay RJ (Hodder Arnold, London), pp 202–243.
9. Outerbridge CA (2006) *Clin Tech Small Anim Pract* 21:128–134.
10. Schmid-Grendelmeier P, Scheynius A, Crameri R (2006) *Chem Immunol Allergy* 91:98–109.
11. Gemmer CM, DeAngelis YM, Theelen B, Boekhout T, Dawson TL, Jr (2002) *J Clin Microbiol* 40:3350–3357.
12. Boekhout T, Kamp M, Gueho E (1998) *Med Mycol* 36:365–372.
13. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, *et al.* (2004) *Science* 304:304–307.
14. Hermida L, Brachat S, Voegeli S, Philippsen P, Primig M (2005) *Nucleic Acids Res* 33:D348–D352.
15. Kämper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O, *et al.* (2006) *Nature* 444:97–101.
16. Gaitatzis N, Kunze B, Muller R (2001) *Proc Natl Acad Sci USA* 98:11136–11141.
17. Deangelis YM, Saunders CW, Johnstone KR, Reeder NL, Coleman CG, Kaczvinsky JR, Jr, Gale C, Walter R, Mekel M, Lacey MP, *et al.* (2007) *J Invest Dermatol* 127:2138–2146.
18. Begerow D, Stoll M, Bauer R (2006) *Mycologia* 98:906–916.
19. Kronstad JW, Staben C (1997) *Annu Rev Genet* 31:245–276.
20. Bakkeren G, Jiang G, Warren RL, Butterfield Y, Shin H, Chiu R, Linning R, Schein J, Lee N, Hu G, *et al.* (2006) *Fungal Genet Biol* 43:655–666.
21. Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, Heitman J (2002) *Eukaryot Cell* 1:704–718.
22. Bardwell L (2005) *Peptides* 26:339–350.
23. Heitman J (2006) *Curr Biol* 16:R711–R725.
24. Hull CM, Johnson AD (1999) *Science* 285:1271–1275.
25. Lachke SA, Lockhart SR, Daniels KJ, Soll DR (2003) *Infect Immun* 71:4970–4976.
26. Braun BR, van Het Hoog M, d'Enfert C, Martchenko M, Dungan J, Kuo A, Inglis DO, Uhl MA, Hogues H, Berriman M, *et al.* (2005) *PLoS Genet* 1:36–57.
27. Calderone RA, Fonzi WA (2001) *Trends Microbiol* 9:327–335.
28. Naglik JR, Challacombe SJ, Hube B (2003) *Microbiol Mol Biol Rev* 67:400–428, table of contents.
29. Cafarchia C, Otranto D (2004) *J Clin Microbiol* 42:4868–4869.
30. Andersson A, Scheynius A, Rasool O (2003) *Med Mycol* 41:479–485.
31. Korf I, Flicek P, Duan D, Brent MR (2001) *Bioinformatics* 17(Suppl 1):S140–S148.
32. Guigo R (1998) *J Comput Biol* 5:681–702.
33. Birney E, Clamp M, Durbin R (2004) *Genome Res* 14:988–995.
34. Eddy SR (1998) *Bioinformatics* 14:755–763.
35. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O (2001) *Nucleic Acids Res* 29:41–43.
36. Slater GS, Birney E (2005) *BMC Bioinformatics* 6:31.
37. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) *J Mol Biol* 340:783–795.