RESEARCH ARTICLE

Genomic Insights into the Atopic Eczema-Associated Skin Commensal Yeast *Malassezia sympodialis*

Anastasia Gioti,[a] Björn Nystedt,[b] Wenjun Li,[c] Jun Xu,[d] Anna Andersson,[e] Anna F. Averette,[c] Karin Münch,[f] Xuying Wang,[c] Catharine Kappauf,[c] Joanne M. Kingsbury,[c] Bart Kraak,[g] Louise A. Walker,[h] Henrik J. Johansson,[i] Tina Holm,[e] Janne Lehtiö,[i] Jason E. Stajich,[j,k] Piotr Mieczkowski,[l] Regine Kahmann,[f] John C. Kennell,[m] Maria E. Cardenas,[c] Joakim Lundeberg,[n] Charles W. Saunders,[d] Teun Boekhout,[g,o] Thomas L. Dawson,[d] Carol A. Munro,[h] Piet W. J. de Groot,[p] Geraldine Butler,[q] Joseph Heitman,[c] Annika Scheynius[e]

Science for Life Laboratory, Translational Immunology Unit, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden[a]; Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden[b]; Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, USA[c]; Procter & Gamble Co., Miami Valley Innovation Center, Cincinnati, Ohio, USA[d]; Translational Immunology Unit, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden[e]; Max Planck Institute for Terrestrial Microbiology, Marburg, Germany[f]; CBS-Fungal Biodiversity Centre, Utrecht, The Netherlands[g]; Institute of Medical Sciences, University of Aberdeen, Aberdeen, United Kingdom[h]; Cancer Proteomics Mass Spectrometry, Science for Life Laboratory, Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden[i]; Department of Plant Pathology & Microbiology[j] and Institute for Integrative Genome Biology,[k] University of California, Riverside, California, USA; Department of Genetics, Carolina Center for Genome Science, School of Medicine, University of North Carolina, at Chapel Hill, Chapel Hill, North Carolina, USA[l]; Department of Biology, Saint Louis University, St. Louis, Missouri, USA[m]; Science for Life Laboratory School of Biotechnology, Division of Gene Technology, Royal Institute of Technology, Stockholm, Sweden[n]; Department of Internal Medicine and Infectious Diseases, University Medical Center, Utrecht, The Netherlands[o]; Regional Center for Biomedical Research, Albacete Science & Technology Park, University of Castilla-La Mancha, Albacete, Spain[p]; and School of Biomolecular and Biomedical Science, University College Dublin, Dublin, Ireland[q]

ABSTRACT *Malassezia* commensal yeasts are associated with a number of skin disorders, such as atopic eczema/dermatitis and dandruff, and they also can cause systemic infections. Here we describe the 7.67-Mbp genome of *Malassezia sympodialis*, a species associated with atopic eczema, and contrast its genome repertoire with that of *Malassezia globosa*, associated with dandruff, as well as those of other closely related fungi. Ninety percent of the predicted *M. sympodialis* protein coding genes were experimentally verified by mass spectrometry at the protein level. We identified a relatively limited number of genes related to lipid biosynthesis, and both species lack the fatty acid synthase gene, in line with the known requirement of these yeasts to assimilate lipids from the host. *Malassezia* species do not appear to have many cell wall-localized glycosylphosphatidylinositol (GPI) proteins and lack other cell wall proteins previously identified in other fungi. This is surprising given that in other fungi these proteins have been shown to mediate interactions (e.g., adhesion and biofilm formation) with the host. The genome revealed a complex evolutionary history for an allergen of unknown function, Mala s 7, shown to be encoded by a member of an amplified gene family of secreted proteins. Based on genetic and biochemical studies with the basidiomycete human fungal pathogen *Cryptococcus neoformans*, we characterized the allergen Mala s 6 as the cytoplasmic cyclophilin A. We further present evidence that *M. sympodialis* may have the capacity to undergo sexual reproduction and present a model for a pseudobipolar mating system that allows limited recombination between two linked *MAT* loci.

IMPORTANCE *Malassezia* commensal yeasts are associated with a number of skin disorders. The previously published genome of *M. globosa* provided some of the first insights into *Malassezia* biology and its involvement in dandruff. Here, we present the genome of *M. sympodialis*, frequently isolated from patients with atopic eczema and healthy individuals. We combined comparative genomics with sequencing and functional characterization of specific genes in a population of clinical isolates and in closely related model systems. Our analyses provide insights into the evolution of allergens related to atopic eczema and the evolutionary trajectory of the machinery for sexual reproduction and meiosis. We hypothesize that *M. sympodialis* may undergo sexual reproduction, which has important implications for the understanding of the life cycle and virulence potential of this medically important yeast. Our findings provide a foundation for the development of genetic and genomic tools to elucidate host-microbe interactions that occur on the skin and to identify potential therapeutic targets.

Address correspondence to Annika Scheynius, annika.scheynius@ki.se.

*M*alassezia is a dominant member of the normal human cutaneous microbial flora which belongs to the subphylum Ustilaginomycotina, phylum *Basidiomycota*, of fungi and thus is more closely related to the plant pathogen *Ustilago maydis* than to the ascomycetous fungi, such as the dermatophytes and *Candida* yeasts that infect humans. *Malassezia* colonizes human skin soon after birth (1) and is also associated with skin diseases such as atopic eczema/dermatitis, pityriasis versicolor, pityrosporum fol-

TABLE 1 Data and assembly and annotation statistics for the *M. sympodialis* genome

Data	454 Titanium (SE)[a]	Illumina HiSeq (MP)[b]	Contigs[c]	Scaffolds Nuclear[d]	Mitochondrial
No.	1,278,053 (reads)	32×10^6 (reads)	156	65	1
Read length (bp)	433 (avg)	50			
Read coverage (fold)	60	200			
Total assembly size (bp)			7,682,651	7,669,689	38,622
N_{50} (bp)[e]			186,342	513,493	
GC content (%)				59	32
No. of protein-coding genes				3,517	19
No. of rRNA genes				NE	2
No. of tRNA genes				NE	25

[a] SE, single-end reads.
[b] MP, 3-kb mate-pair reads.
[c] Contigs come from assembly of the 454 data.
[d] Scaffolds come from assembly of the 454 and Illumina data. NE, not estimated.
[e] N_{50}, weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value.

liculitis, dandruff, and seborrheic dermatitis and even with systemic infections (2, 3). Currently there are 14 recognized species of *Malassezia* that have been isolated from humans or other warm-blooded animals (4). Various studies show differences in the *Malassezia* species found on human skin (2), suggesting that there may be geographic variation in the commensal flora and also the species associated with disease. All *Malassezia* species except *Malassezia pachydermatis* require exogenous lipids for growth. They are frequently associated with sebum-rich areas of the skin, where they obtain fatty acids to fulfill their lipid requirements. Because of their unique nutritional requirements, specialized media such as Dixon's medium (or Leeming and Notman agar) are required for their *in vitro* growth (2). Another unique characteristic of *Malassezia* is the cell wall, which is very thick (~0.12 μm) compared to other yeasts and consists of ~70% sugars, ~10% protein, and 15 to 20% lipids (2). The cell wall is surrounded by a lipid-rich capsule-like structure (5), which may be involved in interactions of this yeast with its host. Thus far, no sexual cycle has been observed for *Malassezia*. A region corresponding to the mating-type locus (*MAT*) and genes encoding key proteins required for meiosis have, however, been identified in the genome of *Malassezia globosa* (6). Because infection of the plant host by the phylogenetically closely related species *U. maydis* is coupled to the sexual cycle (7), it will be of interest to explore in future studies whether a similar pathogenic mechanism may also operate in *Malassezia*.

Among the *Malassezia* species, *Malassezia sympodialis* is one of the most frequently isolated from both atopic eczema patients and healthy individuals (2). Atopic eczema is a common chronic inflammatory skin disease, and the prevalence of this disorder has doubled or tripled in industrialized countries during the past three decades, with 15 to 30% of children and 2 to 10% of adults being afflicted (8). Approximately 50% of adult patients with atopic eczema are sensitized to *M. sympodialis*, as reflected by allergen-specific IgE and/or T cell reactivity to the yeast (9). Reactivity is rarely observed in other allergic diseases (10) indicating a specific link between atopic eczema and *Malassezia*. The pathogenesis of atopic eczema likely results from the combination of a disturbed skin barrier and genetic and environmental factors such as lifestyle, stress, allergens, and microbes (11). The altered skin barrier

provides an environment that leads to elevated skin pH, which enhances the release of IgE-binding proteins (allergens) from *M. sympodialis* (12). Ten allergens have been identified in *M. sympodialis* so far (9). Interestingly, several of the identified allergens are homologous to host proteins, suggesting the possibility of cross-reactive immune responses, whereas others are proteins of unknown function with no sequence homology to characterized proteins (9).

In this study, we focused on *M. sympodialis* as a model to advance our understanding of how the normal skin microbiota interacts with the host and contributes to atopic eczema pathogenesis. We sequenced the genome of *M. sympodialis* to a high-coverage with the aims of (i) exploring genomic features related to the biology of the yeast, (ii) investigating the function and molecular evolution of allergens related to atopic eczema, and (iii) elucidating the presence of a potential sexual cycle. The *M. sympodialis* genome was compared to the published genome of *M. globosa* (6), a species associated with pityriasis versicolor and dandruff, as well as to genomes of other fungi, such as those found on human skin and *U. maydis*, which is found on plants.

RESULTS AND DISCUSSION

The genome of *M. sympodialis*. The draft high-coverage genome of *M. sympodialis* ATCC 42132 was assembled from a shotgun 454 data set and was extended and scaffolded using a 3-kb insert Illumina HiSeq mate-pair data set (Table 1) to a total of 65 scaffolds (L_{50} = 511 kb), corresponding to a nuclear genome of 7.67 Mbp (Table 1). According to CEGMA (13) analysis, this assembly shows a high degree of completeness: 88.3 to 93.1%, comparable to 89.5 to 93.5% for the genome of the closely related species *M. globosa* (6). The estimated genome size of *M. sympodialis* is smaller than that of the *M. globosa* genome (8.96 Mbp), while both are in line with previously reported genome sizes from electrophoretic karyotyping experiments (14). The *Malassezia* genomes are among the smallest in the fungal kingdom, a feature probably related to their dependence on warm-blooded animals. Whole-genome alignments showed extensive synteny between the *M. sympodialis* and *M. globosa* scaffolds (Fig. 1A).

We predicted a total of 3,517 protein-coding genes in *M. sympodialis* using multiple lines of evidence (*ab initio* predictors, ex-
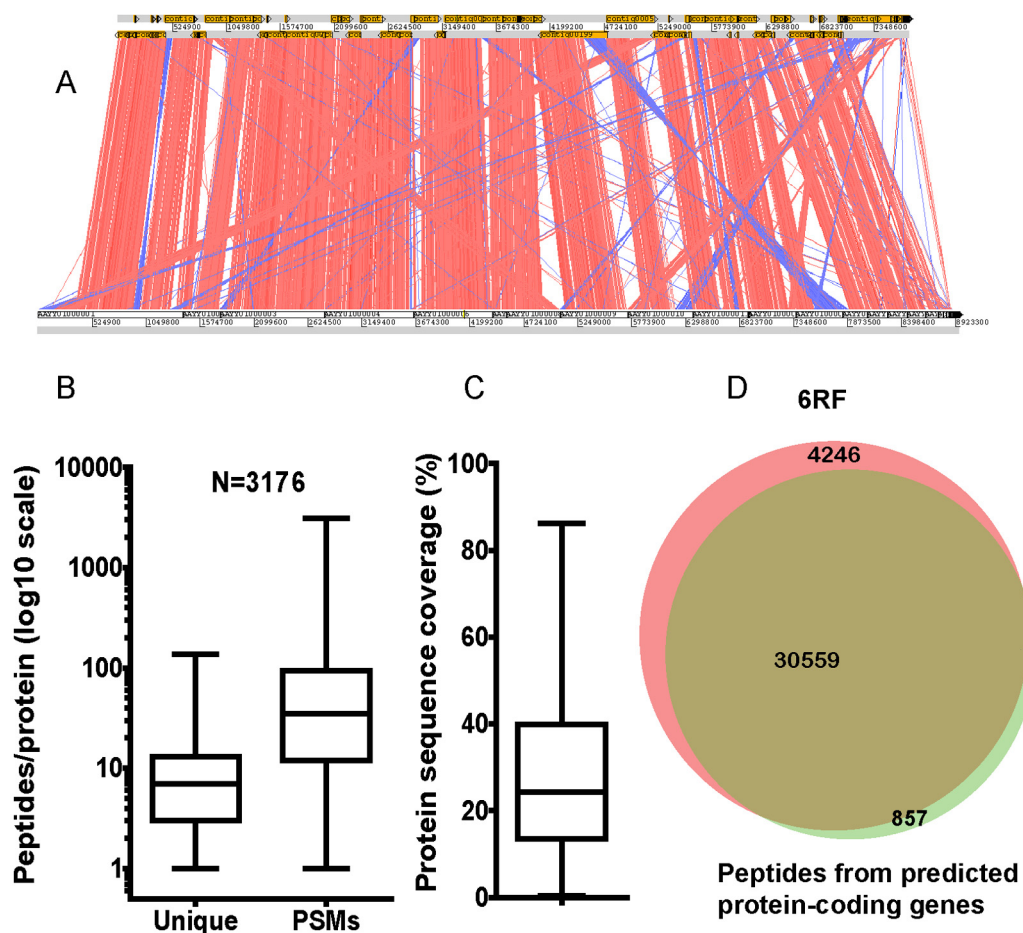
FIG 1 Nuclear genome and proteomics analyses of *M. sympodialis* ATCC 42132. (A) BLASTN alignment between assembly scaffolds from *M. globosa* and contigs of *M. sympodialis* (454 data assembly) indicates a globally conserved synteny. Red and blue bands indicate syntenic and inverted syntenic regions. Note that the contigs of *M. sympodialis* were ordered according to the *M. globosa* scaffold configuration, but their true order in both species is unknown. The alignment was visualized with the ACT tool (100). Mass spectrometry based proteomics (B to D). (B) Boxplot of the number of unique peptides per protein and the number of peptide spectrum matches (PSMs) per protein. (C) Boxplot presenting protein sequence coverage of identified peptides per protein. (D) Venn diagram showing overlap (30,559; 86%) of unique peptides identified by mass spectrometry both from predicted protein coding genes and peptides generated by searching the 6-reading-frame (6RF) translation of *M. sympodialis*.

pressed sequence tags (ESTs) from *M. globosa*, and protein and nucleotide alignments) and manual annotations for specific genes. We applied mass spectrometry (MS) proteomics using peptide isoelectric focusing (IEF) on a broad and a narrow pH range to achieve high proteome coverage of *M. sympodialis* protein extracts and confirmed 3,176 (90%) of the predicted proteins. Of the MS-identified proteins, 98% have 2 or more peptide spectrum matches (PSMs) supporting their identification and 90% have 2 or more unique peptides identified (Fig. 1B) with sequence coverage of 13.5% or more for 75% of the proteins (Fig. 1C). The completeness of the annotation of protein-coding genes was estimated using unique peptides identified by mass spectrometry from both predicted protein coding genes and peptides generated by searching the theoretical tryptic peptidome of 6-reading-frame (6RF) translation of the *M. sympodialis* genome (Fig. 1D). A high proportion of identified peptides (30,559, i.e., 86% of the total number) overlapped between predicted protein-coding genes and 6RF translation of the genome. From the predicted proteome, tryptic peptides can be derived from sequence reaching over exon bound-

aries and other possible variants not present in the 6RF translation of the genome. In this study, 857 peptides were identified only in the predicted-protein database. In addition to the annotated proteome, 4,246 unique peptides were identified only in the 6RF translation. Thus, the comprehensive MS-based proteomics data confirm the majority of the predicted proteins and suggest that future analysis incorporating experimental peptide evidence has the potential to complement and refine the protein-coding-genome annotation.

When inspecting the synteny of orthologous proteins shared between *M. globosa* and *M. sympodialis*, we observed a tendency for genes which are distinct in the *M. globosa* genome to be fused into the same open reading frame (ORF) in the *M. sympodialis* genome. These discrepancies, as well as the difference between the two species regarding the total number of genes, mainly reflect differences between annotation platforms, the lack of RNA evidence for gene predictions in *M. sympodialis*, and, to a smaller extent, errors in sequencing. The annotation of the genomes of *Malassezia* species is particularly challenging, given the limited
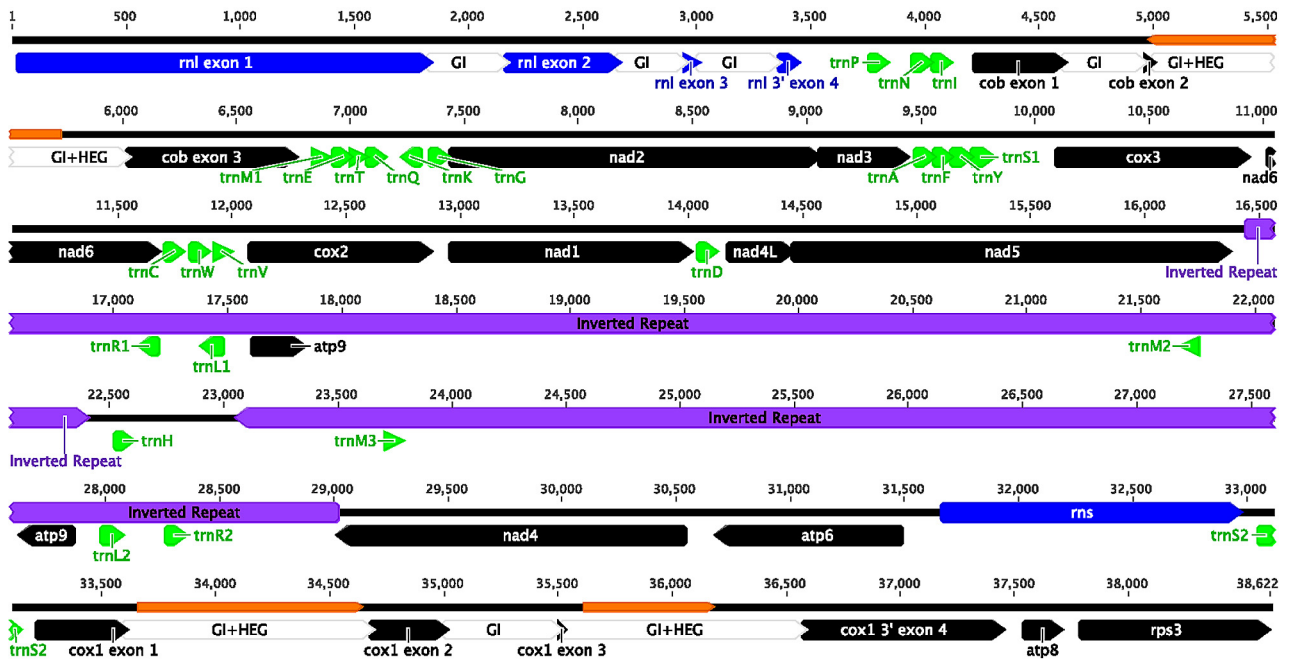
Gioti et al.

FIG 2 Physical map of the mitochondrial genome of *M. sympodialis* ATCC 42132. The 38,622-bp mtDNA maps as a circular molecule and is displayed in a linear form beginning with the *RNL* gene. Black bars represent genes or exons of highly conserved protein-encoding regions, with the orientation indicated by the pointed end. Other bars represent rRNAs (blue), tRNAs (green), introns (gray), and a large inverted repeat (purple). The *RNL*, *COB*, and *COX1* genes are interrupted by group I introns (GI), with three of these introns containing open reading frames belonging to the LAGLI-DADG family of homing endonuclease genes (GI + HEG). The intron-located HEG of the second intron of *COB* is immediately adjacent to and in-frame with the upstream exon, while the HEGs in the first and third introns of *COX1* are free standing.

availability of transcript evidence (only 1,392 ESTs are available for *M. globosa*). Furthermore, the genomes of only a few species from this phylogenetic clade have been sequenced to date, such as *U. maydis* (15), *Ustilago hordei* (16), and *Sporisorium reilianum* (17). This renders the task of gene predictions via homology searches challenging, especially for fast-evolving and species- and genus-specific genes. For *M. sympodialis*, high-coverage sequencing of RNA extracted from distinct growth conditions is now under way to allow better annotation of the genome in a future update (A. Scheynius and T. L. Dawson, personal communication).

The mitochondrial genome of *M. sympodialis*. The mitochondrial genome was assembled from the 454 data and maps as a circular fragment of 38,622 bp with an estimated GC content of 32% (Table 1). Overall it is syntenic with the mitochondrial genome of *M. globosa* (Fig. 2). The genome contains all 15 expected protein-coding genes, 2 rRNA genes, and 25 tRNAs representing all 20 amino acids, with Ser, Leu, and Arg in two copies and Met in three copies. This coding content is the same as that of the related species *U. maydis* (accession number NC_008368). Although there is no synteny between *Malassezia* and *Ustilago* regarding gene order, genes are present on both strands in each species. In contrast to the mitochondrial genome of *M. globosa*, that of *M. sympodialis* has several group I introns (8 in total, including 3 that encode putative homing endonucleases).

A conspicuous feature of the *M. sympodialis* mitochondrial DNA (mtDNA) is a large (5.9-kb) inverted repeat containing the *ATP9* gene and tRNA genes for Met, Leu, and Arg (Fig. 2). The inverted repeat is also present in the *M. globosa* mitochondrial genome, although it is shorter and poorly conserved between the two species, apart from regions corresponding to *ATP9* and

tRNAs. Large inverted repeats are common in chloroplast genomes (18), yet they occur infrequently in fungal mtDNAs and have been reported for only a few genera. In *Candida* species, homologous recombination between large inverted repeats in mtDNAs of *Candida* species has been proposed to play a role in replication and is associated with genome rearrangements (19, 20). Strand invasion structures in the inverted repeats of *C. albicans* mtDNA support a recombination-driven mechanism of DNA initiation (20), while comparative studies of the mitochondrial genomes of *Candida* species indicate that large inverted repeats are involved in conversions between circular and linear forms of the genome, as well as the formation of multipartite linear forms (19). Comparative analysis of mtDNAs of additional *Malassezia* species and clinical isolates is necessary to assess the functional significance of the inverted repeats and to determine whether mitochondria play a role in virulence, as has been reported with other fungal pathogens (21–23).

Differences between *Malassezia* species regarding metabolism. We investigated the metabolic pathways for potential changes that could explain the *in vitro* nutritional requirements of *Malassezia* species. Comparing the genomes of *M. globosa* and *M. sympodialis* for genes involved in lipid metabolism, we found that both genomes had a similar complement of genes. Neither genome encodes a recognizable fungal-type fatty acid synthase, while each genome encodes a plethora of lipid-hydrolyzing enzymes, such as lipases, phospholipases C, and acid sphingomyelinases (Table 2). *M. sympodialis* has a slightly reduced number of lipases and phospholipases C compared to *M. globosa*. It is not clear whether this contributes to any physiological differences or if this pattern is simply due to the potentially incomplete predicted

TABLE 2 Main lipid-hydrolyzing enzymes in *M. globosa* and *M. sympodialis*

Gene family[a]	No. of enzymes in:	
	M. globosa	*M. sympodialis*
M. globosa LIP lipase like	6	4
M. globosa LIP1 lipase like	6	5
Acid sphingomyelinase	4	4
Phospholipase C	6	4

[a] From reference 6.

proteome of *M. sympodialis*. No $\Delta 2,3$-enoyl coenzyme A isomerase gene was found in either genome, suggesting defects in utilizing unsaturated fatty acids. Notably, a $\Delta 9$ desaturase gene, absent in both the *M. globosa* and *Malassezia restricta* genomes, was found in the *M. sympodialis* genome (MSY001_2159). This suggests that *M. sympodialis*, in contrast to *M. globosa*, may be able to add a double bond to generate oleic acid if provided with stearic acid in the culture medium. However, since oleic acid is included in the mDixon medium used to grow *Malassezia*, the difference between the two species regarding the $\Delta 9$ desaturase is unlikely to explain the reported better *in vitro* growth of *M. sympodialis* (24), unless oleic acid uptake is limiting in *M. globosa* and *M. restricta*.

A second explanation for the differences in growth between the two species could relate to genes involved in sugar assimilation, such as those encoding β-glucosidases. In contrast to *M. globosa*, *M. sympodialis* is positive in the β-glucosidase enzymatic assay (25). In fungi these enzymes belong to two families, glycoside hydrolase 1 and 3 (GH1 and 3); the number of genes belonging to each family varies in basidiomycetes, with 0 to 2 copies for GH1 and 3 to 7 for GH3 (15). Similar to the *U. maydis* and *Cryptococcus neoformans* genomes, *Malassezia* genomes do not code for a GH1-type fungal β-glucosidase but, notably, show even greater compactness, with only one gene coding for a putative GH3-type enzyme (see Fig. S1 in the supplemental material). As the GH3-type gene is present in both *Malassezia* genomes, differences in expression may explain why *M. globosa* does not show a β-glucosidase enzymatic activity.

Cell wall genes in *M. sympodialis*. A unique characteristic of *Malassezia* is the cell wall, which is very thick compared to the cell walls of other yeasts (2). The *M. sympodialis* genome contains representatives of the major polysaccharide biosynthesis genes (Table 3). These include genes encoding six chitin synthases, proteins associated with β-1,6-glucan synthesis, and only one β-1,3-glucan synthase catalytic subunit. Other basidiomycetes also have one Fks-like β-1,3-glucan synthase catalytic subunit, whereas ascomycetes such as *S. cerevisiae* generally harbor a number of alternative Fks subunits (Table 3). There are six predicted chitin deacetylases, which is similar to the number found in other basidiomycetes but larger than the number generally seen in ascomycetes. This suggests that a large proportion of cell wall chitin may be converted to chitosan, the deacetylated form of chitin. The chromosomal location of this gene family suggests that there have been three separate gene duplication events leading to expansion of the chitin deacetylase family. Studies in *C. neoformans* have shown that chitosan helps to maintain cell wall integrity (26), suggesting that chitin deacetylases and the chitosan made by them may prove to be excellent antifungal targets. There is also an abundance of putative exoglucanases with similarity to *S. cerevisiae* Exg1 (Table 3). The classical cell wall integrity or protein kinase C (PKC) pathway seems to be highly conserved in *M. sympodialis* (see Table S1 in the supplemental material). Putative enzymes involved in *O*-glycosylation are represented in the genome such as the O protein mannosyltransferase (Pmt) family, but there is little evidence of orthologs of *N*-glycosylation enzymes, in particular those that add sugars to the outer chains to *N*-glycan (see Table S1).

In yeasts such as *S. cerevisiae* and *Candida albicans*, the major class of cell wall-localized proteins comprises proteins that are modified by the addition of a glycosylphosphatidylinositol (GPI) anchor (27). The posttranslational addition of a GPI anchor targets proteins to the plasma membrane. Through a poorly understood mechanism, the GPI anchor of a subset of GPI proteins is cleaved and the proteins are translocated to the wall and covalently attached to β-1,6-glucan. Only ten *M. sympodialis* proteins (and 20 in *M. globosa*) were predicted to become GPI anchored

TABLE 3 Cell wall genes

Gene class[a]	No. of genes in[b]:				
	Msym	Mglo	Umay	Cneo	Scer
Chitin synthase	6	7	8	9	3
Chitin deacetylase	6[c]	4	6	4	2
Chitinase (class IV)	1	1	2	4	2
Catalytic subunit of β-1,3-glucan synthase (*FKS*)	1	1	1	1	3
Exo-β-1,3-glucanase (*EXG1*)	8	6	8	8	3
Transglycosylase (GH16, *CRH*)	1	1	2	2	3
Transglucosylase (GH72, *GAS*)	0	0	1	1	5
Mixed-linked glucanase (*MLG1*)	2[d]	2	4	5	0
Putative β-1,6-glucan transglycosylase (*KRE6*)	4[c]	4	8	6	2
ER chaperone involved in protein N- and O-glycosylation (*ROT1*)	1	2	4	1	1
Predicted GPI proteins	10	20	55[e]	63	59[f]/66[g]

[a] Homologous gene or gene families in *S. cerevisiae* (except for *MLG1*, which is from *Cochliobolus carbonum*).
[b] Msym, *M. sympodialis*; Mglo, *M. globosa*; Umay, *U. maydis*; Cneo, *C. neoformans*; Scer, *S. cerevisiae*.
[c] One of the gene models seems incorrect and comprises two paralogous genes.
[d] Second paralog identified by tBLASTn but no gene model available.
[e] Data from reference 101.
[f] Data from reference 102.
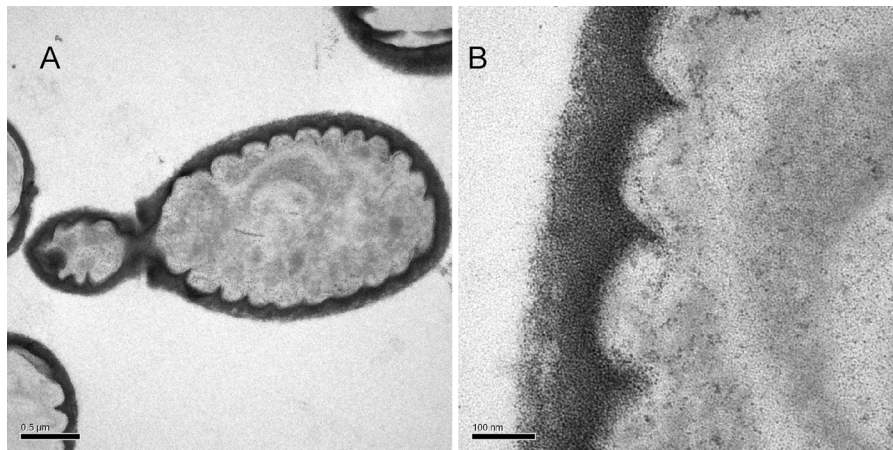[g] Data from reference 30.

FIG 3 *M sympodialis* cell wall architecture revealed by HPF-TEM (high-pressure freezing--transmission electron microscopy). *M. sympodialis* was grown on mDixon agar at 32°C for 4 days. (A) Transmission electron micrograph of a budding *M. sympodialis* yeast cell. (B) Ultrastructure of the *M. sympodialis* cell wall. Bars: 0.5 μm (magnification, ×25,000) (A) and 100 nm (magnification, ×130,000) (B).

(Table 3), and surprisingly, they are likely to be associated with the membrane rather than the cell wall. This is a very small number compared to that in other fungal species; pathogens of the genus *Candida* can have over ten times more predicted GPI-anchored proteins (28). In *C. albicans*, these GPI-modified proteins include important adhesins that are involved in adhesion to host epithelial and endothelial cells as well as to innate surfaces, such as indwelling catheters, and thereby also contribute to biofilm formation (27). Both adhesion and biofilm formation play a role in virulence in invading pathogens that cause systemic and bloodstream infections. GPI-modified proteins also include carbohydrate-active enzymes with important roles in cell wall construction and maintenance of cell wall integrity. These proteins (transglucosidases and chitin-glucan cross-linkers) act by modulating the cell wall polysaccharides chitin and β-1,3-glucan. Only one putative chitin-glucan cross-linker gene, a Utr2 homolog, was identified in the *M. sympodialis* genome. No Dfg5/Dcw1 family members were identified. These putative mannosidases are generally assumed to be involved in the cleavage of GPI anchors, consistent with the lack of GPI cell wall proteins. This prompted us to look for homologs of the proteins that synthesize the GPI anchor itself, and we found representatives of the enzymes that synthesize most steps in the pathway in *S. cerevisiae* (see Table S1 in the supplemental material).

From the above, we can surmise that there are β-1,3-glucan, β-1,6 glucan (probably β-1,3/β-1,6 glucan), chitin, and chitosan in the cell wall of *M. sympodialis*, and glycosylation may be limited primarily to *O*-glycosylation. Analysis of the genome provides no evidence of α-glucan synthesis and no or very few "classical" fungal cell wall proteins. These *in silico* results are in agreement with a previous cell wall carbohydrate analysis that revealed that the *M. sympodialis* cell wall is composed primarily of β-1,6 glucan, with trace amounts of branched β-1,6/β-1,3-glucan and mannan (29). We further performed high-pressure freezing--transmission electron microscopy (HPF-TEM) to corroborate our observations on the absence of cell wall proteins and proteins for outer N-mannosylation, which participate in forming a fibrillar layer. *M. sympodialis* indeed lacks the extensive outer fibrillar layer (Fig. 3) that is evident on the wall of *S. cerevisiae* and *C. albicans* (30, 31).

Molecular evolution and function of allergens. Genes coding for all 10 allergens previously cloned from *M. sympodialis* (Mala s 1 and s 5 to s 13) and for the three allergens from *M. furfur* (Mala f 2 to 4) (9) were identified in the *M. sympodialis* genome (Table 4). For some of the allergens, where the previously described sequence was incomplete (Mala s 13) or contained sequencing errors at the 5′ end (Mala s 11), the availability of the genome sequence and protein alignments allowed identification of the full coding sequence, including the start codon. Sequence identity between *M. sympodialis* and *M. globosa* at both the nucleotide and amino acid levels is generally high (Table 4). Despite the high degree of protein identity between putative orthologs, a molecular evolution analysis indicated high levels of nucleotide substitutions ($dS > 1$) for all of them (Table 4). A population of 56 clinical *M. sympodialis* isolates from healthy individuals and atopic eczema patients (see Table S2 in the supplemental material) was further analyzed for the presence of two major allergens, Mala s 1, an allergen of unknown function (32), and Mala s 12, showing sequence similarity to the GMC oxidoreductase family (33). The partial genes for Mala s 1 and Mala s 12 (encoding 81% and 45% of the mature proteins, respectively) were amplified (primers are listed in Table S3 in the supplemental material) in all clinical isolates and showed strong conservation. This finding suggests that the regions of the genes we examined are under high selective constraints, possibly reflecting the maintenance of essential roles in the interaction with the host.

The function of Mala s 1 is still an enigma despite the availability of its three-dimensional (3-D) structure. The 3-D structure indicates that Mala s 1 is a β-propeller-folded protein. This novel fold among allergens has structural similarity in the potential homologs Q4P4P8 and Tri 14, from the plant pathogens *U. maydis* and *Gibberella zeae*, respectively (34), suggesting that Mala s 1 and the plant pathogen proteins may have similar functions. Because gene deletion approaches have not been established for *Malassezia*, we investigated the role of the *Mala s 1* ortholog in the related smut fungus *U. maydis*. Quantitative real-time PCR revealed that expression of the *Mala s 1* ortholog in *U. maydis* (um04915), encoding a protein predicted to be secreted, is induced during colonization of maize seedlings (see Fig. S2A in the supplemental ma-

TABLE 4 Allergens encoded in the *M. sympodialis* ATCC 42132 genome and putative orthologs in *M. globosa*

Allergen[a]	Accession no. (reference)	Predicted function	Prediction for secretion (both species)[b]	*M. sympodialis* gene	*M. globosa* ortholog[c]	% identity Amino acid	Nucleotide	dN	dS
Mala s 1	X96486 (32)	Unknown; similarity to *G. zeae* Tri14	Secreted	MSY001_0607	MGL_1303	69	66	0.23	4.09
Mala f 2	AB011804 (103)	Peroxisomal protein	No	MSY001_2163	MGL_4042	55	61	0.39	3.45
Mala f 3	AB011805 (103)	Peroxisomal protein	No	MSY001_2163	MGL_4042	55	61	0.39	3.45
Mala f 4	AF084828 (104)	Malate dehydrogenase	No	MSY001_0149	MGL_2703	84	77	0.10	2.74
Mala s 5	AJ011955 (43)	Peroxisomal protein	No	MSY001_2163	MGL_4042	55	61	0.39	3.45
Mala s 6	AJ011956 (43)	Cytoplasmic cyclophilin	No	MSY001_1373	MGL_3612	93	85	0.04	1.61
Mala s 7	AJ011957 (39)	Unknown	Secreted	MSY001_3348	MGL_0968[d]	NA[d]	NA[d]	NA[d]	NA[d]
Mala s 8	AJ011958 (39)	Unknown	Secreted	MSY001_0606	MGL_1304	71	68	0.22	4.52
Mala s 9	AJ011959 (39)	Unknown	No	MSY001_1912	MGL_2179	77	74	0.19	1.95
Mala s 10	AJ428052 (105)	Heat shock protein	No	MSY001_0570	MGL_0201	89	78	0.07	4.60
Mala s 11	AJ548421 (105)	Manganese superoxide dismutase	No	MSY001_2804	MGL_3190	71	73	0.18	3.23
Mala s 12	AJ871960 (33)	GMC oxidoreductase	Secreted	MSY001_2108	MGL_0750	64	65	0.29	2.52
Mala s 13	AJ937746 (106)	Thioredoxin	No	MSY001_0904	MGL_1781	85	81	0.10	1.76

[a] Isolated from ATCC 42132 except for Mala f 2, 3, and 4, which come from isolate 2782 (Teikyo Institute for Medical Mycology, Tokyo, Japan).
[b] Evidence for no secretion is absence of signal peptides, transmembrane domains, and GPI-anchoring peptides.
[c] Single-copy orthologs between the two species were identified with a bidirectional best-hit BLASTP approach (E value = 1E − 50).
[d] The gene shows weak similarity to the gene encoding Mala s 7, and due to gene family amplification [see the text], it cannot be safely assigned as its ortholog; therefore, dN/dS analysis between the two copies is not applicable (NA).

terial). The gene *um04915* was deleted from the genome of the solopathogenic strain SG200 (see the supplemental methods in the supplemental material). With respect to SG200, *um04915* mutants were unaltered in growth sensitivity to various stressors, including H_2O_2, sorbitol, Calcofluor white, and Congo red, or in filamentation on charcoal-containing plates (data not shown). Virulence was determined in seedling infections by comparing four independent mutant strains and SG200. No significant differences in disease symptoms were noted in comparison to SG200 (see Fig. S2B in the supplemental material). Therefore, the *um04915* gene is not directly related to *U. maydis* pathogenicity in seedling infections; however, we cannot exclude the possibility that the gene may be required for disease in different maize organs (35) or has evolved different roles in *Malassezia*, associated with a different host.

Enhanced release of Mala s allergens and particularly Mala s 12 has been observed when *M. sympodialis* is cultured at a higher pH, which reflects that of the skin of atopic eczema patients (12). Here we predicted four of the known allergens to be secreted proteins (Table 4). Combining this observation with proteomics experiments on the *M. globosa* orthologs (6) and results from a previous study that showed that Mala s 1 and 12 are expressed on the cell surface of *Malassezia* (36), we suggest that these allergens may be exported and/or loosely associated with the cell wall, for example, via disulfide bonds (27) or, for Mala s 1, via binding to phosphoinositides involved in membrane trafficking (34). Notably, the characterized ortholog of Mala s 1 from the wheat pathogen *Fusarium graminearum* (Tri 14) was proposed to be functionally associated (either as a regulator or as a transporter) with an adjacent gene cluster involved in biosynthesis of a mycotoxin (37). This observation is of interest, as in both the *M. sympodialis* and *M. globosa* genomes, the gene encoding Mala s 1 is located adjacent to the gene encoding *Mala s 8*. Furthermore, the first 10 genes located on the 5′ end of *Mala s 8* code for proteins potentially involved in secondary metabolism, such as a putative monooxygenase, a permease, a taurine dioxygenase, and a cobalamin-independent methionine synthase. A compelling hypothesis that merits further investigation is that Mala s 1 and Mala s 8 are involved in cell wall or postsecretory modifications of an as-yet-unidentified secondary metabolite. A few pathways of secondary metabolism have been observed in *Malassezia*, with some evidence for contributions to pathogenesis (4). Another potentially secreted allergen is Mala s 7 (Table 4), an allergen of unknown function (38, 39), which we identified here as a member of a novel family. Indeed, most of the 13 allergens represent single-copy genes in *M. sympodialis*, with a few belonging to small gene families with 2 to 5 gene copies each. The *M. sympodialis* genome has four genes predicted to encode proteins similar to the protein sequence of Mala s 7, in contrast to three genes in *M. globosa*. The identified sequences of all Mala s 7-like proteins from *M. sympodialis* and *M. globosa* bear signal peptides and do not have transmembrane domains or GPI anchors. Two of the genes in *M. sympodialis* are highly similar: *Mala s 7a*, which codes for the published allergen (38, 39) and a second, termed here *Mala s 7b*; both are located at the ends of relatively short scaffolds, showing a nucleotide identity of ~90% over a region of 4 kb, including both the complete *Mala s 7* genes as well as the surrounding intergenic sequences. We confirmed by PCR (primers are listed in Table S3 in the supplemental material) that the observed duplication does not represent an assembly artifact (data not shown). Additional sequencing and mapping to

chromosomes is required to resolve whether this is a segmental duplication or whether the duplicated fragments lie on distinct chromosomes. A segmental duplication event could also be at the origin of the other two *Mala s 7* copies in *M. sympodialis* and *M. globosa*, as the respective genes in both species are located only 1.5 kb apart. These genes are *Mala s 7c* and *7d* in *M. sympodialis* and *MGL_0968* and a gene missed in previous annotations (accession number JX857443) in *M. globosa*. Gene conversion often occurs between copies lying in close proximity in the genome, and the species-based clustering of genes for Mala s 7-like proteins in the gene tree (see Fig. S3 in the supplemental material) is in line with this scenario. Another interesting observation is that the Mala s 7 proteins in *M. globosa* (MGL_0968, MGL_2673, and JX857443) are shorter and have low similarity with their *M. sympodialis* counterparts at both the C and N termini, which could indicate that Mala s 7 in *M. sympodialis* comes from a fusion of two smaller Mala s 7-like proteins. A molecular evolution analysis using branch models in PAML (40, 41) did not further elucidate this family's complex history, due to inconclusively high *dS* values. Overall, the genome sequence revealed a gene family amplification for the Mala s 7 allergen in *M. sympodialis*, which merits further investigation. Gene duplication is a major force driving evolution of new traits, including virulence, and thus it will be of interest to determine the roles of the secreted Mala s 7-like proteins in the interaction with the host and their evolutionary history.

We next addressed the function of Mala s 6, which is a member of the cyclophilin panallergen family (42). The sequence of the *M. sympodialis* Mala s 6 protein exhibited highest similarity to the cytoplasmic form of cyclophilin A, which is in agreement with the predicted absence of a secretion signal peptide (Table 4). Mala s 6 was found to be the most conserved protein in pairwise comparisons between *M. globosa* and *M. sympodialis* (Table 4), in line with its reported conservation across the tree of life. Using the *C. neoformans* model system, we investigated the functions of Mala s 6 using immunological, enzymatic, and drug inhibition assays. Recombinant Mala s 6 (43) reacted with a Cpa1-specific antisera (Fig. 4A) that successfully recognizes both the Cpa1 and Cpa2 cyclophilin A proteins in *C. neoformans* (44). Using a chymotrypsin-coupled assay that measures the *cis*-to-*trans* isomerization of a synthetic peptide (45), we found that the Mala s 6 recombinant protein shows robust *cis-trans* peptidyl-prolyl isomerase activity (Fig. 4B). Furthermore, the recombinant Mala s 6 protein is sensitive to inhibition by cyclosporine A (Fig. 4B), an effective immunosuppressive natural product that targets cyclophilin A and calcineurin (44). Cyclosporine A shows beneficial effects in atopic eczema, but due to side effects, its use is limited to patients with severe refractory disease (11). In conclusion, we demonstrated that Mala s 6 is a bona fide cyclophilin A and targeted by cyclosporine A.

The mating-type (*MAT*) locus of *M. sympodialis* corresponds to a pseudobipolar mating system. We identified in the *M. sympodialis* genome assembly two scaffolds corresponding to the *A* (or pheromone/receptor [P/R]) and *B* (or homeodomain [HD]) mating type (*MAT*) loci based on sequence similarity shared with *MAT* genes of the closely related species *M. globosa* (6). The *A* and *B* mating-type loci in *M. sympodialis* were not linked in our assemblies; however, aligning them to the *M. globosa MAT* locus, where the two alleles are linked (6), identified three additional scaffolds which may lie between these loci and which could define a contiguous mating-type locus in *M. sympodialis* (Fig. 5A). Alignments of
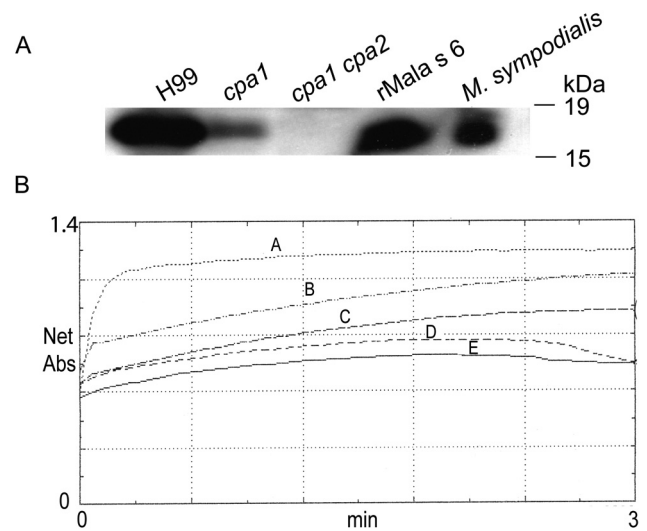

FIG 4 Functional characterization of Mala s 6. (A) Western blot detection of Mala s 6 antigen using a polyclonal antiserum against *C. neoformans* cyclophilin A. Total protein extracts from *C. neoformans* strains, including wild-type H99 and *cpa1* and *cpa1 cpa2* mutants, were separated in parallel with a protein extract from *M. sympodialis* and recombinant Mala s 6 (rMala s 6). (B) Mala s 6 catalyzes *cis-trans* peptidyl-prolyl isomerization, as shown by a chymotrypsin-coupled assay. *x* axis, time (in minutes); *y* axis, net absorbance measured in the spectrophotometer. Curve A, rMala s 6; curve B, *C. albicans* cyclophilin A (Cyp1); curve C, rMala s 6 + 1 μM cyclosporine A; curve D, *C. albicans* cyclophilin A (Cyp1) +1 μM cyclosporine A; curve E, control reaction mixture without enzyme.

raw Illumina reads onto these scaffolds and PCR experiments (see the supplemental methods) provided further evidence for this contiguous assembly. Thus, similar to *M. globosa* (6), the *M. sympodialis A* and *B* loci appear to be physically linked and lie about ~141 kb apart.

In basidiomycetes, linkage of the *A* and *B* loci, along with strict biallelism, indicative of absence of recombination between these alleles, commonly defines bipolar mating systems, while tetrapolar mating systems contain unlinked and multiallelic *A* and *B* loci (46). However, in contrast to expectations for bipolar species, both the *A* and *B MAT* locus alleles of *M. sympodialis* showed extended flanking synteny with the corresponding *M. globosa MAT* locus regions (Fig. 5A). In comparison, in the bipolar species *U. hordei*, where the ancestral *A* and *B MAT* regions lie 430 to 500 kb apart, separated by a region that is highly rearranged between the two mating types, there is synteny on the 5′ end flanking the *B* locus and on the 3′ end flanking the *A* locus but not on their other flanks (47). Sequencing of the *M. sympodialis A* and *B MAT* alleles in a population sample of isolates (see Table S2 and the supplemental methods in the supplemental material) further indicated that the mating system of this species might not fit either of the traditional bipolar and tetrapolar systems, similar to what was previously reported for the pseudobipolar species *Sporidiobolus salmonicolor* (48). Below we present evidence suggesting that *M. sympodialis* has an intermediate mating system, termed the pseudobipolar mating system, where multiallelism might occur despite physical linkage (for a comparison of mating systems, see Fig. S4 in the supplemental material).

The *A* locus present in the genome of ATCC 42132 encodes a pheromone and a pheromone receptor arranged as two adjacent
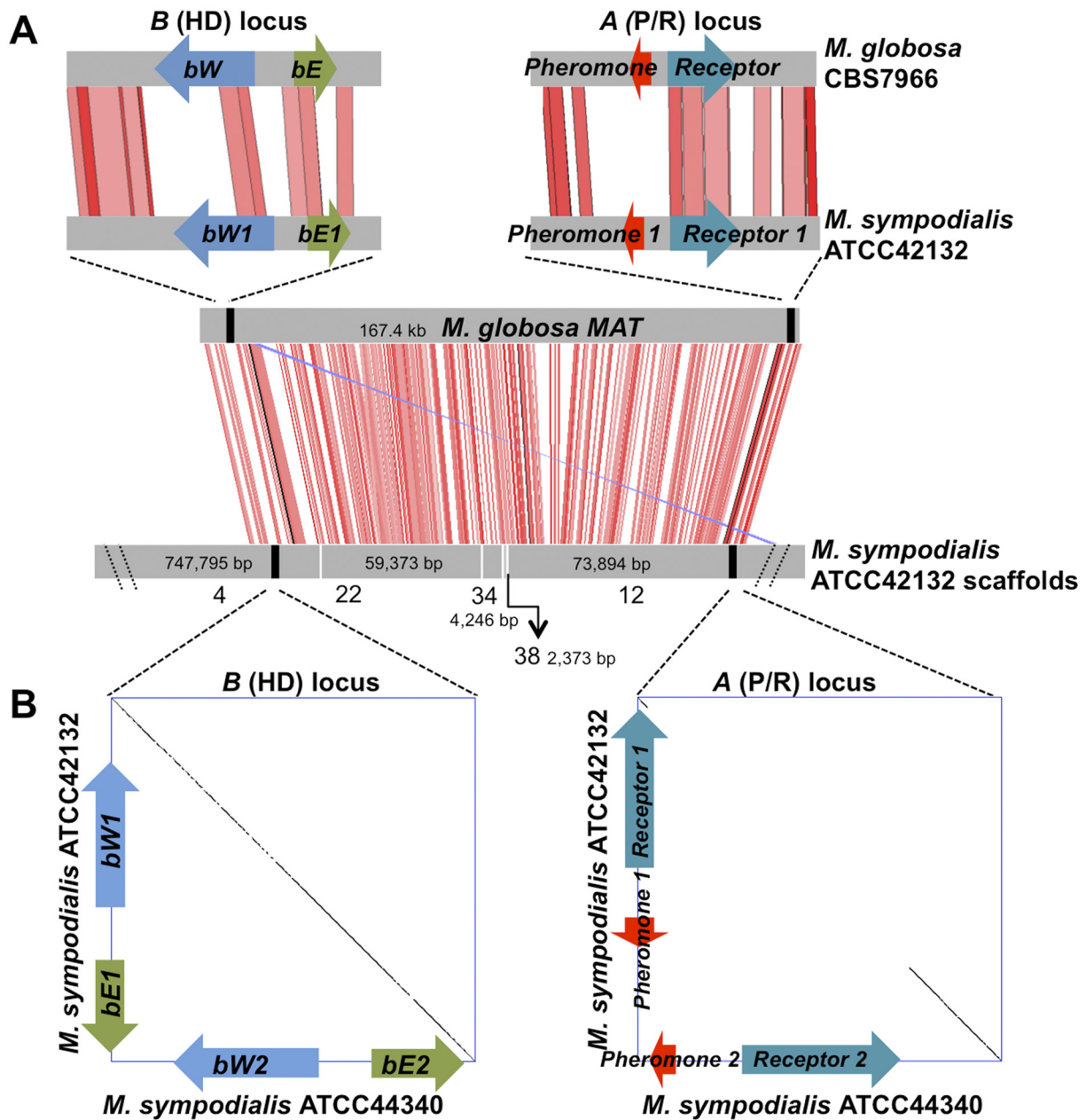
FIG 5 Organization of the *MAT* locus in *Malassezia*. (A) Comparison of the *MAT* locus of *M. globosa* and *M. sympodialis*. As seen in the lower comparison, the *MAT* locus of *M. globosa* CBS 7966 (accession no. AYY01000003.1) maps to five scaffolds of the *M. sympodialis* isolate ATCC 42132. Scaffolds 12 and 4 correspond to the *A* (pheromone/receptor [P/R]) and *B* (homeodomain [HD]) loci, comprising the pheromone and the pheromone receptor gene and the transcription factor genes *bW* and *bE*, respectively. The *A* and *B* loci are ~167.4 kb apart in *M. globosa* and ~141 kb apart in *M. sympodialis*, with scaffolds 22, 34, and 38 linked to scaffolds 4 and 12, based on analysis of Illumina reads and PCR and sequence analysis spanning each gap (see the supplemental methods). Alignments between the two species were done with tBLASTx and visualized with ACT (100). (B) Dot plot comparison of the two alleles of the *MAT* locus between *M. sympodialis* isolates ATCC 42132 (sequenced isolate) and *M. sympodialis* ATCC 44340 (sequence determined by PCR and sequencing; see Materials and Methods). Sequences were aligned using Dnadot (http://www.vivo.colostate.edu/molkit/dnadot/index.html) with a window size of 15. The pheromone and pheromone receptor genes in the *A* locus (right) have different sequences and orientations in *M. sympodialis* ATCC 44340 (accession no. JX964849) and ATCC 42132 (accession no. JX964848), and the flanking regions are highly conserved. The *bW* and *bE* genes in the *B* (HD) locus (left) share high similarity between these two isolates (ATCC 44340, accession no. JX964801; ATCC 42132, accession no. JX964802), and the flanking regions are highly conserved.

divergently oriented genes (Fig. 5A) and was designated *a1*. A *MAT A* locus allele that was distinct in terms of both sequence conservation and orientation of the genes was identified by sequencing the same region in a collection of isolates (see Table S2 and supplemental methods in the supplemental material) and

designated *a2*. Both *a1* and *a2* are embedded in highly conserved, syntenic regions (Fig. 5B). In contrast to the *a* locus in the closely related species *U. maydis*, which is biallelic, the *a1* and *a2 MAT* alleles from *M. sympodialis* do not contain a pheromone pseudo-gene, which in *U. maydis* is thought to descend from a tri-allelic

P/R mating type system shared with the last common ancestor with *S. reilianum*, which is also tri-allelic (49–51). Moreover, the *M. sympodialis A MAT* locus alleles both lack the *lga2* and *rga2* genes that are present in the *U. maydis a2* and *S. reilianum a2* alleles, and which are involved in governing uniparental mitochondrial inheritance in those species (52, 53).

Similarly, the *M. sympodialis B MAT* locus of the reference strain ATCC 42132 (Fig. 5A) was designated allele *b1* and contains divergently oriented genes encoding the homeodomain transcription factors bE and bW (by analogy with *U. maydis*). Again, PCR with flanking primers and sequence analysis revealed the presence of two additional alleles, *b2* and *b3*, distinguished from *b1* by a series of substitutions, many of which lie in the N-terminal region or in the protein-protein interaction domains of the two homeodomain factors and are nonsynonymous (Fig. 5B; also, see Fig. S5 in the supplemental material). The differences observed in the *M. sympodialis B* locus alleles may be sufficient to represent different mating types based on two lines of reasoning (54, 55). First, comparison to similar amino acid changes that are naturally occurring in *U. maydis B* locus alleles of different mating types reveals a similar degree of substitution and in similar regions of the proteins. Second, amino acid substitutions found in *U. maydis* mutants that exhibit compatibility with different-partner homeodomain proteins also exhibit a similar pattern to the changes observed between the *b1*, *b2*, and *b3* mating type alleles of *M. sympodialis*. The finding of biallelism for the *A* locus and triallelism for the *B MAT* locus suggests that recombination could occur between these regions. Further evidence for this is based on an allele compatibility test that revealed five of the six possible allelic configurations predicted for two unlinked *MAT* loci (*a1b1*, *a1b2*, *a2b1*, *a2b2*, and *a2b3*) in a collection of *M. sympodialis* isolates (see Table S2 in the supplemental material). In contrast, if the three *B* alleles observed were simply the result of drift, we would have expected linkage between the *A* and *B* alleles (*a1b1* and *a2b2*), but not the recombinant *a1b2* or *a2b1* combinations. These results suggest that the ~141 kb region separating the *A* and *B MAT* loci does not suppress recombination as in the biallelic *MAT* locus of *U. hordei* (56).

Overall, our comparative genomic and polymorphism analyses are consistent with a pseudobipolar mating system for *M. sympodialis*, distinct from tetrapolar mating systems such as the one observed in *U. maydis* and from strict biallelic bipolar systems, such as in the species *U. hordei* (47, 48, 54, 56–59). The pseudobipolar mating system of *M. sympodialis* may have arisen recently from a tetrapolar ancestor, so that large-scale rearrangements between the two linked loci that erase flanking synteny have not yet occurred, in contrast to other derived bipolar species, such as *U. hordei*. Multiple independent transitions from an ancestral tetrapolar state to a bipolar state and possibly also a pseudobipolar derived state in the *Basidiomycota* have frequently been reported (47, 60). In this view, *M. globosa* could also represent a pseudobipolar state derived from a tetrapolar ancestor; however, evidence for multiallelism in isolates of this species is needed to confirm this hypothesis.

Genes involved in mating and meiosis. Pheromone response during mating in fungi is regulated through a MAP kinase cascade, coupled to a heterotrimeric G protein consisting of α, β, and γ subunits. The genes of the MAP kinase module (e.g., *FUS3*, *STE7*, *STE11*) and the receptor for a-factor pheromone (*STE3*) are generally conserved in *Malassezia* (Table 5; for a full list, see Table S4

in the supplemental material). In contrast, the G protein subunits show more diverged profiles: of the three (or four in *U. maydis* [61] and *S. reilianum*) Gα subunits present in filamentous ascomycetes and many basidiomycetes (GPA1-4) (62), only one is present in *Malassezia* species (Table 5). This protein is most closely related to Gpa3 from *U. maydis*, required for mating (61). The Gβ subunit, associated with mating in many fungi (for examples, see reference 63), is an ortholog of Ste4 of *S. cerevisiae* (64); homologs of this protein are present in *U. maydis*, *S. reilianum*, and *M. globosa* but not in the *M. sympodialis* genome (Table 5). However, mating is regulated in *U. maydis* by a WD-40 protein related to Gβ subunits, Rak1 (65); Rak1 is conserved in all the *Ustilagomycotina*, including *Malassezia* species (see Table S4). It is possible that this protein interacts with the single Gα (Gpa3) in *Malassezia* as part of the mating signaling response. We could not identify a Gγ subunit (Ste18) ortholog in either of the *Malassezia* species. However, Gγ proteins are small and poorly conserved and can be difficult to find in genomic sequences.

Of the 29 genes defined as "core" for meiosis in eukaryotes (66, 67), only 19 are unequivocally present in the *Malassezia* genomes (Table 5). However, many losses are not specific to these species. For example, in most organisms, strand invasion is promoted following formation of double-strand breaks (DSB) by the activity of two proteins that arose from a gene duplication event that preceded the evolution of eukaryotes: Rad51, required during recombination in meiosis and for repairing DSBs in somatic cells, and Dmc1, which functions only in meiosis (68). Similarly to the genomes of other *Ustilagomycotina* species (*M. globosa*, *U. maydis*, and *S. reilianum*), the yeasts *Candida guilliermondii* and *Candida lusitaniae* (28, 69), the microsporidian species *Encephalitozoon cuniculi*, *Caenorhabditis elegans*, and *Drosophila melanogaster*, the *M. sympodialis* genome contains only a Rad51 homolog, no Dmc1. The loss of the *DMC1* gene is correlated with the absence of genes (see Table S4 in the supplemental material) coding for the assembly factors Sae3 and Mei5 (70, 71). It is therefore possible that Rad51 alone is required for recombination in *Malassezia*, as previously proposed for *C. elegans* and *D. melanogaster* (72). Initiation of meiotic recombination in eukaryotes requires the formation of double-strand breaks in DNA by a complex containing Spo11. Spo11 orthologs are conserved in almost all eukaryotes, apart from one lineage of protists (73). It was initially difficult to identify the *SPO11* gene in *M. sympodialis* (MSY001_2221). However, preliminary RNA-Seq data (T. Holm and A. Scheynius, unpublished data) provided evidence for a gene structure that contains six introns, one of which has a noncanonical splice site. *SPO11* is expressed at a low level in two-day cultures of *M. sympodialis*, suggesting that the organism has retained the capacity to undergo meiosis. A putative multi-intron *SPO11* candidate is also found in the *M. globosa* genome (J. Xu and C. W. Saunders, unpublished data).

Most eukaryotic genomes contain two paralogs of the essential cohesion complex, namely, *REC8* and *RAD21*; the corresponding proteins are required for cohesion of sister chromatids during mitosis and meiosis (66). Rec8 is a meiosis-specific component (74). Surprisingly, the *Malassezia* genomes contain only one paralog gene, and this is more closely related to *RAD21* than to *REC8* (see Fig. S6 in the supplemental material). Both paralogs are present in the other *Ustilagomycotina* (*U. maydis* and *S. reilianum*) and in other basidiomycetes (e.g., *C. neoformans*) (see Table S4 in the supplemental material). The loss of Rec8 is unusual in fungi; how-

TABLE 5 Genes involved in mating signaling and in basidiomycetes

Process	Gene[a]	Presence[b] of gene in:				
		Msym	Mglo	Umay	Sreil	Cneo
Mating signaling	*ASC1*	1	1	1	1	1
	FUS3	1	1	1	1	1
	GPA1	0	0	1	1	1
	GPA2	0	0	1	1	1
	GPA3	1	1	1	1	1
	GPA4	0	0	1	1	0
	STE3[c]	1	1	1	1	1
	STE4	0	1	1	1	1
	STE7	1	1	1	1	1
	STE11	1	1	1	1	1
	STE18	?[d]	?[d]	1	1	1
Recombination and crossing over	**DMC1**	0	0	0	0	1
	HOP1	0	0	0	0	1
	HOP2	0	0	0	0	1
	MND1	0	0	0	0	1
	MRE11	1	1	1	1	1
	RAD50[e]	1	1	1	1	1
	RAD51	1	1	1	1	1
	RAD52[c]	1	1	1	1	1
	SPO11	1	1	1	1	1
Crossover resolution	**MER3**	0	0	1	1	1
	MSH2	1	1	1	1	1
	MSH4	0	0	1	1	1
	MSH5	0	0	1	1	1
	MSH6	1	1	1	1	1
	RAD1[c]	1	1	1	1	1
Cohesin complex	**PDS5**[f]	1	1	1	1	1
	RAD21	1	1	1	1	1
	REC8	0	0	1	1	1
	SCC3	1	1	1	1	1
	SMC1	1	1	1	1	1
	SMC2	1	1	1	1	1
	SMC3	1	1	1	1	1
	SMC4	1	1	1	1	1
	SMC5	1	1	1	1	1
	SMC6	1	1	1	1	1
Mismatch repair	**MLH1**	1	1	1	1	1
	MLH2	0	0	0	0	0
	MLH3	0	0	1	1	1
	PMS1[g]	1	1	1	1	1

[a] Core meiosis genes are in bold.

[b] 1, presence; 0, absence. Msym, *M. sympodialis*; Mglo, *M. globosa*; Umay, *U. maydis*; Sreil, *S. reilianum*; Cneo, *C. neoformans*.

[c] The gene is present in the genome based on tBLASTn, but there is no model available.

[d] *STE18* homologs were not found, but this might be due to the fact that they are small and poorly conserved.

[e] The *RAD50* homolog in *M. globosa* is incorrectly split into two genes, *MGL_0431* and *MGL_0432*.

[f] The *PDS5* homolog is split into *MGL_3630* and *MGL_3631*.

[g] The *M. sympodialis* gene model (*MSY001_1319*) might be an incorrect fusion of two genes, corresponding to *M. globosa MGL_0016* and *MGL_0017*.

ever, this protein is also missing from the protists (66). It is possible that Rad21, which plays a role in meiosis in mammals (75, 76), may substitute for Rec8. Other meiotic subunits of the cohesion complex (Smc1, Smc3, Scc3, and Pds5) are present in *Malassezia* and related basidiomycetes (see Table S4 in the supplemental material).

Overall, the *Malassezia* species have lost a substantial number of genes associated with mating and meiosis (Table 5; also, see Table S4 in the supplemental material). However, few of the gene losses are unique; rather, most are shared with other species with an extant sexual cycle. Examples include genes encoding proteins involved in formation of the synaptonemal complex (SC) (*HOP1*, *ZIP2*, *ZIP3*, and *RED1*). These genes are absent not only in

Malassezia spp. (see Table S4) but also in *U. maydis* (77, 78), a species in which sexual reproduction is well documented, which may suggest that the *Ustilagomycotina* do not form SCs at all, similar to *Schizosaccharomyces pombe* and some *Candida* species (28, 79). The genomic evidence is consistent with the possibility that there may be a sexual cycle in *Malassezia* resembling that of other *Ustilagomycotina*.

Sexual reproduction in *Malassezia*? In this study, we identified independent lines of evidence for sexual reproduction in the *Malassezia* genus: (i) the presence of a *MAT* locus with apparently intact *MAT* alleles in the genomes of three *Malassezia* species (Fig. 5A) (6); (ii) evidence for recombination in a population of isolates of *M. sympodialis*, as shown by the discovery of segregating poly-

morphisms in the *A* and *B* regions of the *MAT* locus (Fig. 5B; also, see Fig. S3 in the supplemental material); and (iii) conservation of genes required for meiosis and signaling in the mating process in both *M. globosa* and *M. sympodialis* (Table 5; also, see Table S4 in the supplemental material). Although some of the core meiotic genes are absent, these are not consistently defined as "core" genes in the literature (for example, see reference 72), and their absence does not necessarily suggest an absence of sex (80). For example, genes absent from *Malassezia* but present in the other *Ustilagomycotina* include *MSH4*, *MSH5*, and *MER3*, required for the resolution of Holliday junctions during recombination in *S. cerevisiae*. All three are also missing from sexual fungi such as *C. guilliermondii*, *C. lusitaniae* (28), and *S. pombe* and from other eukaryotes with intact sexual cycles, including *Plasmodium* and *Drosophila* (66).

One way to confirm sexual reproduction and establish the role of *MAT* alleles in this process would be to conduct mating assays for fertility. In an attempt to detect an extant sexual cycle, a collection of *M. sympodialis* isolates (see the supplemental methods and Table S2 in the supplemental material) have been cocultured in pairwise and more complex mixtures under a variety of different media and environmental conditions (see the supplemental methods). However, despite these efforts, to date none of the morphological features associated with basidiomycete sexual reproduction (e.g., dikaryotic hyphae, clamp connections, basidia, and basidiospores) have been observed. Thus, the right combination of strains or conditions to detect an extant sexual cycle, if one exists, has not yet been established. The *Malassezia* sexual cycle might differ morphologically from that of other fungi and could require genetic approaches with marked strains to detect responses to pheromones, cell-cell fusion, or genetic recombination. No sexual cycle is known for any *Malassezia* species, but previous studies provide evidence for *M. furfur* hybrids that may result from mating between varieties or cryptic species (81; T. Boekhout, personal communication). Furthermore, recombination was observed in allozyme studies of the species *M. pachydermatis* (82).

Concluding remarks. In summary, the genome of *M. sympodialis* reported here, combined with previous studies of the *M. restricta* and *M. globosa* genomes, provides a rich foundation for future studies to elucidate the unique features of these ubiquitous commensals of human skin associated with multiple disease states. Moreover, apparent transitions in the mating type locus suggest that much remains to be learned about how, when, and where sexual reproduction might occur. Sexual reproduction may occur on human skin, with implications for antigens presented by the yeast, which might provoke immune reactions, leading to disease. This study provides insight into a number of hypotheses related to the life cycle of these fungi. The development of *M. sympodialis* transformation protocols for gene replacement studies will be a critical next step toward assessing the roles of genes potentially required for *Malassezia* species to become specialized to live on the skin as commensals but also to provoke disease.

MATERIALS AND METHODS

DNA extraction. DNA was isolated from *M. sympodialis* ATCC 42132 cultured on Dixon agar (24) modified to contain 1% (vol/vol) Tween 60, 1% (wt/vol) agar, and no oleic acid (mDixon) at 32°C for 4 days using the QIAamp DNA minikit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions with small modifications. Briefly, glass beads were added to the cell suspension, which was vortexed for 4 min prior to the lysing incubation at 56°C for 3 h with additional vortexing during the incubation period.

Genome sequencing, assembly, and CEGMA analysis. Pyrosequencing was performed using 454 Titanium chemistries (Roche/454 Life Sciences, Branford, CT). Illumina libraries were made with the Illumina mate pair kit (3-kb insert) according to the manufacturer's instructions, followed by 50-bp paired-end sequencing on one lane of an Illumina HiSeq instrument. Genome assembly of the 454 data was accomplished with the GS De Novo assembler, v. 2.3 (Roche Diagnostics, Basel, Switzerland). Contig extension and scaffolding were based on the Illumina data using SSPACE, v. 1.0 (83). We noted that a large fraction (~50%) of the Illumina mate-pair data in reality represented standard noncircularized paired ends, but no attempts to confirm or extend the scaffold structure by long-range PCRs were done at this point. The mitochondrial assembly was performed with Newbler 2.6, using a random subset of 40,000 reads from the 454 shotgun sequencing, representing ~25× coverage of the mitochondrial genome. Scaffolding and identification of the inverted repeat were aided by mapping of the Illumina 3-kb jumping library on the mitochondrial contigs. Finally, the genome was fully completed by manual identification and addition of 454 reads spanning the assembly gaps flanking the inverted repeat. CEGMA analysis was run using a set of 248 core eukaryotic genes (CEGs) as queries against the assemblies of *M. sympodialis* and *M. globosa*, with the completeness reported as a percentage reflecting the number of CEGs found as complete or partial genes, respectively.

Gene predictions. The genome of *M. sympodialis* was annotated using the program MAKER versions 2.10 and 2.25 (84, 85). As evidence for gene annotations we used (i) protein alignments to a set of 67,086 publicly available proteins derived from the species *M. globosa*, *U. maydis*, *C. neoformans*, *Fusarium graminearum*, *Magnaporthe grisea*, *Neurospora discreta*, *Neurospora tetrasperma*, and *Sordaria macrospora* and clustered using Cd-hit (86) version 4.02 and a protein identity threshold of 90%; (ii) nucleotide alignments to 1,392 EST sequences coming from an *M. globosa* sequenced library (accession number LIBEST_028020) previously used for gene predictions in *M. globosa* (6); and (iii) the *ab initio* predictors Augustus (87), using a *U. maydis* model, Genemark-ES (88), trained with the *M. sympodialis* scaffolds, and SNAP (89), trained within MAKER as follows. MAKER was run four consecutive times, and each annotation output file from MAKER (genome gff file) was converted into a model using the instructions from the SNAP documentation and provided as an input model in the next run. We complemented the predicted genes from the fourth MAKER run with (i) a set of 542 models, identified by protein alignments of the *M. globosa* proteome against *ab initio* models not retained by MAKER using the "pass-through" method implemented in MAKER version 2.25, and (ii) a few genes manually retrieved (e.g., *SPO11*, genes for Mala s 7-like proteins, and a pheromone gene) using tBLASTn and manual curation. The gene models from the above analyses have not been further curated. Mitochondrial assembly was performed with Newbler 2.6, using a random subset of 40,000 reads from the 454 shotgun sequencing, representing ~25× coverage of the mitochondrial genome. Scaffolding and identification of the inverted repeat were aided by mapping of the Illumina 3-kb jumping library on the mitochondrial contigs. The genome was completed by manual identification and addition of 454 reads spanning the assembly gaps flanking the inverted repeat. The mitochondrial DNA was annotated following steps described in reference 90; the mitochondrial map was generated using Geneious Pro, v.5.6.4 (Biomatters, Auckland, New Zealand). Details on identification of specific genes presented in tables are found in the supplemental methods.

Mass spectrometry. Four replicates of *M. sympodialis* (ATCC 42132) were cultured on mDixon agar (see above) at 32°C for 2 and 15 days. Cells were harvested and washed twice with PBS (phosphate-buffered saline) by centrifugation at 1,200 × *g* for 5 min. Pellets were frozen at −80°C. To extract proteins for mass spectrometry analyses, 30 mg of every pellet was dissolved in 200 μl PBS and transferred to tubes containing 200 μl 425 to

600 μm acid-washed glass beads (Sigma-Aldrich, Sweden). The cells were disrupted by homogenization in a Precellys 24 tissue homogenizer (Bertin Technologies, France). Approximately 100 μl cell suspension from every sample was removed and transferred to a new tube and a 1:1 volume of lysis buffer was added to obtain a final concentration of 4% (wt/vol) SDS, 1 mM DTT (dithiothreitol), 25 mM HEPES (pH 7.6). The samples were heated at 95°C for 5 min followed by sonication twice for 30 s each time. Protein concentration was determined with a DC protein assay (Bio-Rad, Sweden). Samples were subsequently reduced by dithiothreitol and alkylated by iodoacetamide followed by overnight trypsination (Promega). The four biological replicates of each sample from the 2- and 15-day cultures were pooled and separated by immobilized pH gradient-isoelectric focusing (IPG-IEF) on a narrow-range pH 3.7 to 4.9 strip and a 3- to 10-gel strip as described previously (91). Extracted fractions from the IPG-IEF were separated using an Agilent 1200 Nano-LC system coupled to a Thermo Scientific LTQ Orbitrap Velos. Proteome discoverer 1.3 with Sequest-Percolator (Thermo Scientific) was used to search predicted proteins or a 6-reading frame translation of the *M. sympodialis* genome merged with the *Bos taurus* database (UniProt canonical sequences, 120,727) for protein identification, limited to a false discovery rate of <1%. Peptide matches to the *B. taurus* database were considered remnants from the culture medium and removed.

High-pressure freezing (HPF)-transmission electron microscopy (TEM). HPF of *Malassezia* isolates was carried out as described previously (31). Briefly, samples were prepared by high-pressure freezing with an EMPACT2 high-pressure freezer and rapid transport system (Leica Microsystems Ltd., Milton Keynes, United Kingdom). After freezing, cells were freeze-substituted in substitution reagent (1% [wt/vol] OsO_4 in acetone) with a Leica EMAFS2. Samples were then embedded in Spurr resin and additional infiltration was provided under a vacuum at 60°C before embedding in Leica FSP specimen containers and polymerizing at 60°C for 48 h. Semithin survey sections, 0.5 μm thick, were stained with 1% toluidine blue to identify areas containing cells. Ultrathin sections (60 nm) were prepared with a Diatome diamond knife on a Leica UC6 ultramicrotome and stained with uranyl acetate and lead citrate for examination with a Philips CM10 transmission microscope (FEI UK Ltd., Cambridge, United Kingdom) and imaging with a Gatan Bioscan 792 (Gatan United Kingdom, Abingdon, United Kingdom).

Molecular evolution analyses. Orthologous genes were aligned using ClustalW 2.1 (92) or Muscle (93). Nonsynonymous (dN) and synonymous (dS) substitution frequencies were calculated using the method described in reference 94, as implemented in yn00 in the PAML package (94), except for the Mala s 7 family, where branch models were tested in PAML (40, 41). Gene trees for this analysis (see Fig. S3 in the supplemental material) and for the β-glucosidase (see Fig. S1 in the supplemental material) were constructed with PhyML software (95) using the LG model for amino acid equilibrium frequencies, allowing estimation of invariable sites and setting the rate categories number to 4. An optimized tree topology search was performed with a starting BioNJ tree and using the best-of-NNI and SPR method. Bootstrapping analysis was performed with 1,000 datasets.

M. sympodialis **isolates.** The native clinical *M. sympodialis* isolates utilized for amplification of allergens, mating-type genes, and mating assays (see Table S2 in the supplemental material) were obtained from healthy individuals and from patients with moderate to severe atopic eczema at the Dermatology Unit, Karolinska University Hospital, Stockholm, Sweden, and the protocol was approved by the local ethics committee. The participants were instructed not to wash their upper back on the day of isolation. Samples were taken by holding a contact plate containing modified Leeming and Notman agar medium (96) against the skin of the upper back for 15 s. The contact plates were incubated at 32°C for 6 days. One colony from each plate was transferred to mDixon agar plates (see above) and cultured for 4 days at 32°C. The isolates were identified as *M. sympodialis* using the primers listed in Table S3 in the supplemental material.

PCR amplification and sequencing of allergens in *M. sympodialis*. For 56 clinical isolates (see Table S2 in the supplemental material), DNA corresponding to the partial gene sequences of *Mala s 1* (915 bp) and *Mala s 12* (913 bp) was amplified by PCR using primers (see Table S3 in the supplemental material) designed according to published sequences (32, 33). The PCR amplifications were carried out with Phusion high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA) under the following cycling conditions: an initial denaturation at 98°C for 2 min followed by 30 cycles of 10 s at 98°C, 20 s at 64°C, and 15 s at 72°C and a final elongation step at 72°C for 7 min. The PCR products were purified using the QIAquick PCR purification kit (Qiagen) and sequenced at a verified core facility (KIGene, Karolinska Institutet, Stockholm, Sweden) using the same primers as the PCR amplification. Sequencing reactions were performed on an ABI 3730 Prism DNA analyzer (Applied Biosystems, Foster City, CA) using the BigDye terminator, v.3.1 (Applied Biosystems). The retrieved forward and reverse sequences were aligned using Geneious software, version 5.5.7 (Biomatters Ltd.), and the resulting consensus sequences were used for further analysis. The copies of *Mala s 7a* and *7b* were confirmed in *M. sympodialis* ATCC 42132 by PCR amplification using the polymerase noted above and primers listed in Table S3 in the supplemental material. The following cycling conditions were used: an initial denaturation at 98°C for 1 min followed by 35 cycles of 15 s at 98°C, 15 s at 64°C, and 2.5 min at 72°C and a final elongation step at 72°C for 10 min. The product was analyzed by electrophoresis in a 1% (wt/vol) agarose gel (Invitrogen, Groningen, Netherlands) with SYBRSafe DNA gel staining (Invitrogen) in 1× TAE (Tris-acetate-EDTA).

Western blotting and biochemical assays of Mala s 6. Total proteins (50 μg) from *C. neoformans* strains, including wild-type H99 and *cpa1* and *cpa1 cpa2* mutants (44), were fractionated by 18% (wt/vol) SDS-PAGE in parallel with a protein extract from *M. sympodialis* ATCC 42132 (97) and recombinant Mala s 6 (43). Proteins were transferred to PVDF (polyvinylidene difluoride) membranes (Bio-Rad) and incubated with an anti-Cpa1 rabbit antiserum diluted 1:2,000 (44). Membranes were developed using the enhanced chemiluminescence (ECL) advanced detection kit (Amersham). Peptidylprolyl isomerization activity of Mala s 6 was assayed by an improved method described previously (45). Briefly, 1 ml reaction buffer containing 0.5 mg/ml chymotrypsin (Sigma) and 500 ng of the recombinant Mala s 6 protein (43) or 500 ng of the *C. albicans* Cyp1 cyclophilin A (Y. L. Chen and M. E. Cardenas-Corona, unpublished data) was pre-equilibrated to 10°C and then rapidly mixed in chilled cuvettes containing 10 μl of the substrate peptide (from a stock solution of 0.5 mM *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide [Sigma] dissolved in trifluoroethanol containing 470 mM LiCl). The cuvettes were immediately placed in the spectrophotometer, and the release of *p*-nitroanilide was monitored at 395 nm and 10°C with a Beckman DU-600 spectrophotometer. Cyclosporine A (LC Laboratories) from a 100 mM stock solution in methanol was added to 1-ml reaction mixtures at a final concentration of 1 μM prior to the addition of substrate peptide to test for inhibition of the enzyme activity.

Identification of the *M. sympodialis* mating type (*MAT*) locus and sequencing of the region. The genomic scaffolds containing the *MAT* locus were retrieved using tBLASTn using as queries the *M. globosa* genes *pra1* (MGL_0964), a putative pheromone gene (MGL_0963), *bW1* (MGL_0883), and *bE1* (MGL_0884). Whole-genome alignments between *M. globosa* and *M. sympodialis* genomes with Mummer (98) allowed identification of the three additional scaffolds located between the *A* (P/R) and *B* (HD) loci in *M. globosa*, which were not linked in the original *M. sympodialis* assembly. To further characterize the *M. sympodialis* MAT locus, we designed primers (see Table S3 in the supplemental material) to amplify the *A* and *B* regions from *M. sympodialis* isolates (see Table S2) cultured on modified Dixon agar (24) to containing 1% (wt/vol) peptone, 1% (wt/vol) desiccated ox bile, 1% (wt/vol) Tween 60, 2% (wt/vol) agar, and no oleic acid at 30°C for 4 days. Primers were designed based on alignments with *M. globosa* using Primer3 (99). The PCR products (~3 kb for the *A* locus and ~4 kb for the *B* locus) were purified on a 1% (wt/vol)

agarose gel using a Qiaquick gel DNA extraction kit and used for direct DNA sequencing by primer walking. Sequencing reactions were carried out at the Genome Sequencing & Analysis Core Facility at the Duke Institute for Genome Sciences & Policy (IGSP), at Eton BioScience (Research Triangle Park, NC) and at CBS Fungal Biodiversity Centre, Utrecht, The Netherlands. For details on linkage analysis of *MAT* loci, mating assays, and mating type identification, see the supplemental methods.

Nucleotide sequence accession numbers. The nuclear and mitochondrial genomes of *M. sympodialis* have been deposited in the EMBL database and were assigned accession numbers HE999549-HE999613 and HF558646, respectively. The EST data from *M. globosa* are accessible under accession number LIBEST_028020. ITS (internal transcribed spacer) and *MAT* locus sequences of different *M. sympodialis* isolates were deposited in GenBank under accession numbers JX964840 to JX964847, JX964800 to JX964802, and JX964848 to JX964850 (see Table S2 in the supplemental material).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org /lookup/suppl/doi:10.1128/mBio.00572-12/-/DCSupplemental.

Text S1, PDF file, 0.6 MB.
Figure S1, PDF file, 0.1 MB.
Figure S2, PDF file, 0.3 MB.
Figure S3, PDF file, 0.1 MB.
Figure S4, PDF file, 0.3 MB.
Figure S5, PDF file, 0.1 MB.
Figure S6, PDF file, 0.2 MB.
Table S1, XLSX file, 0.1 MB.
Table S2, XLSX file, 0.1 MB.
Table S3, XLSX file, 0.1 MB.
Table S4, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. **Nagata R, et al.** 2012. Transmission of the major skin microbiota, *Malassezia*, from mother to neonate. Pediatr. Int. **54**:350–355.

2. **Ashbee HR, Scheynius A.** 2010. *Malassezia*, p 209–230. *In* Ashbee HR, Bignell EM (ed), Pathogenic yeasts. The yeast handbook. Springer-Verlag, Berlin, Germany.

3. **Saunders CW, Scheynius A, Heitman J.** 2012. *Malassezia* fungi are specialized to live on skin and associated with dandruff, eczema, and other skin diseases. PLoS Pathog. **8**:e1002701. http://dx.doi.org/10.1371 /journal.ppat.1002701.

4. **Hort W, Mayser P.** 2011. *Malassezia* virulence determinants. Curr. Opin. Infect. Dis. **24**:100–105.

5. **Mittag H.** 1995. Fine structural investigation of *Malassezia furfur*. II. The envelope of the yeast cells. Mycoses **38**:13–21.

6. **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.

7. **Feldbrügge M, Kämper J, Steinberg G, Kahmann R.** 2004. Regulation of mating and pathogenic development in *Ustilago maydis*. Curr. Opin. Microbiol. **7**:666–672.

8. **Bieber T.** 2008. Mechanisms of disease: atopic dermatitis. N. Engl. J. Med. **358**:1483–1494.

9. **Scheynius A, Crameri R.** 2010. *Malassezia* in atopic eczema/dermatitis, p 212–228. *In* Boekhout T, Guého-Kellerman E, Mayser P, Velegraki A (ed), *Malassezia* and the skin. Springer-Verlag, Berlin, Germany.

10. **Casagrande BF, et al.** 2006. Sensitization to the yeast *Malassezia sympodialis* is specific for extrinsic and intrinsic atopic eczema. J. Invest. Dermatol. **126**:2414–2421.

11. **Akdis CA, et al.** 2006. Diagnosis and treatment of atopic dermatitis in children and adults: European Academy of Allergology and Clinical Immunology/American Academy of Allergy, Asthma and Immunology/ PRACTALL Consensus report. J. Allergy Clin. Immunol. **118**:152–169.

12. **Selander C, Zargari A, Möllby R, Rasool O, Scheynius A.** 2006. Higher pH level, corresponding to that on the skin of patients with atopic eczema, stimulates the release of *Malassezia sympodialis* allergens. Allergy **61**:1002–1008.

13. **Parra G, Bradnam K, Korf I.** 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23**: 1061–1067.

14. **Sugita T, et al.** 2010. Epidemiology of *Malassezia*-related skin diseases, p 65–119. *In* Boekhout T, Guého-Kellerman E, Mayser P, Velegraki A (ed), *Malassezia* and the skin. Springer-Verlag, Berlin, Germany.

15. **Kämper J, et al.** 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature **444**:97–101.

16. **Laurie JD, et al.** 2012. Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. Plant Cell **24**:1733–1745.

17. **Schirawski J, et al.** 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. Science **330**:1546–1548.

18. **Bock R.** 2007. Structure, function, and inheritance of plastid genomes, p 29–63. *In* Bock R (ed), Cell and molecular biology of plastids. Springer-Verlag, Berlin, Germany.

19. **Valach M, et al.** 2011. Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. Nucleic Acids Res. **39**:4202–4219.

20. **Gerhold JM, Aun A, Sedman T, Jõers P, Sedman J.** 2010. Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. Mol. Cell **39**:851–861.

21. **Ma H, et al.** 2009. The fatal fungal outbreak on Vancouver Island is characterized by enhanced intracellular parasitism driven by mitochondrial regulation. Proc. Natl. Acad. Sci. U. S. A. **106**:12980–12985.

22. **Olson A, Stenlid J.** 2001. Plant pathogens: mitochondrial control of fungal hybrid virulence. Nature **411**:438.

23. **Qu Y, et al.** 2012. Mitochondrial sorting and assembly machinery subunit Sam37 in *Candida albicans*: insight into the roles of mitochondria in fitness, cell wall integrity, and virulence. Eukaryot. Cell **11**:532–544.

24. **Guého E, Midgley G, Guillot J.** 1996. The genus *Malassezia* with description of four new species. Antonie Van Leeuwenhoek **69**:337–355.

25. **Guého-Kellerman E, Boekhout T, Begerow D.** 2010. Biodiversity, phylogeny and ultrastructure, p 17–76. *In* Boekhout T, Guého-Kellerman E, Mayser P, Velegraki A (ed), *Malassezia* and the skin. Springer-Verlag, Berlin, Germany.

26. **Baker LG, Specht CA, Donlin MJ, Lodge JK.** 2007. Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*. Eukaryot. Cell **6**:855–867.

27. **De Groot PW, Ram AF, Klis FM.** 2005. Features and functions of covalently linked proteins in fungal cell walls. Fungal Genet. Biol. **42**: 657–675.

28. **Butler G, et al.** 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature **459**:657–662.

29. **Kruppa MD, et al.** 2009. Identification of (1→6)-beta-D-glucan as the major carbohydrate component of the *Malassezia sympodialis* cell wall. Carbohydr. Res. **344**:2474–2479.

30. **De Groot PW, Hellingwerf KJ, Klis FM.** 2003. Genome-wide identification of fungal GPI proteins. Yeast **20**:781–796.

31. **Walker CA, et al.** 2010. Melanin externalization in *Candida albicans* depends on cell wall chitin structures. Eukaryot. Cell **9**:1329–1342.

32. **Schmidt M, et al.** 1997. The complete cDNA sequence and expression of the first major allergenic protein of *Malassezia furfur*, Mal f 1. Eur. J. Biochem. **246**:181–185.

33. **Zargari A, et al.** 2007. Mala s 12 is a major allergen in patients with atopic eczema and has sequence similarities to the GMC oxidoreductase family. Allergy **62**:695–703.

34. **Vilhelmsson M, et al.** 2007. Crystal structure of the major *Malassezia sympodialis* allergen Mala s 1 reveals a beta-propeller fold: a novel fold among allergens. J. Mol. Biol. **369**:1079–1086.

35. **Skibbe DS, Doehlemann G, Fernandes J, Walbot V.** 2010. Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. Science **328**:89–92.

36. **Zargari A, Emilson A, Halldén G, Johansson S, Scheynius A.** 1997. Cell surface expression of two major yeast allergens in the *pityrosporum* genus. Clin. Exp. Allergy **27**:584–592.

37. **Dyer RB, Plattner RD, Kendra DF, Brown DW.** 2005. *Fusarium graminearum* TRI14 is required for high virulence and DON production on wheat but not for DON synthesis *in vitro*. J. Agric. Food Chem. **53**:9281–9287.

38. **Andersson A, Scheynius A, Rasool O.** 2003. Detection of Mala f and Mala s allergen sequences within the genus *Malassezia*. Med. Mycol. **41**:479–485.

39. **Rasool O, et al.** 2000. Cloning, characterization and expression of complete coding sequences of three IgE binding *Malassezia furfur* allergens, Mal f 7, Mal f 8 and Mal f 9. Eur. J. Biochem. **267**:4355–4361.

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

41. **Yang Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**:555–556.

42. **Glaser AG, et al.** 2006. Analysis of the cross-reactivity and of the 1.5 Å crystal structure of the *Malassezia sympodialis* Mala s 6 allergen, a member of the cyclophilin pan-allergen family. Biochem. J. **396**:41–49.

43. **Lindborg M, et al.** 1999. Selective cloning of allergens from the skin colonizing yeast *Malassezia furfur* by phage surface display technology. J. Invest. Dermatol. **113**:156–161.

44. **Wang P, Cardenas ME, Cox GM, Perfect JR, Heitman J.** 2001. Two cyclophilin A homologs with shared and distinct functions important for growth and virulence of *Cryptococcus neoformans*. EMBO Rep. **2**:511–518.

45. **Heitman J, Koller A, Cardenas ME, Hall MN.** 1993. Identification of immunosuppressive drug targets in yeast. Methods **5**:176–187.

46. **Kües U, James TY, Heitman J.** 2011. Mating type in basidiomycetes: unipolar, bipolar, and tetrapolar patterns of sexuality, p 97–160. *In* Pöggeler S, Wöstemeyer J (ed), Evolution of fungi and fungal-like organisms, vol **14**. Springer Verlag, Berlin, Germany.

47. **Bakkeren G, Kämper J, Schirawski J.** 2008. Sex in smut fungi: structure, function and evolution of mating-type complexes. Fungal Genet. Biol. **45**(Suppl 1):S15–S21.

48. **Coelho MA, Sampaio JP, Gonçalves P.** 2010. A deviation from the bipolar-tetrapolar mating paradigm in an early diverged basidiomycete. PLoS Genet. **6**:e1001052. http://dx.doi.org/10.1371/journal.pgen.1001052.

49. **Kellner R, Vollmeister E, Feldbrügge M, Begerow D.** 2011. Interspecific sex in grass Smuts and the genetic diversity of their pheromone-receptor system. PLoS Genet. **7**:e1002436. http://dx.doi.org/10.1371/journal.pgen.1002436.

50. **Schirawski J, Heinze B, Wagenknecht M, Kahmann R.** 2005. Mating type loci of *Sporisorium reilianum*: novel pattern with three a and multiple b specificities. Eukaryot. Cell **4**:1317–1327.

51. **Urban M, Kahmann R, Bölker M.** 1996. The biallelic a mating type locus of *Ustilago maydis*: remnants of an additional pheromone gene indicate evolution from a multiallelic ancestor. Mol. Gen. Genet. **250**:414–420.

52. **Fedler M, Luh KS, Stelter K, Nieto-Jacobo F, Basse CW.** 2009. The *a2* mating-type locus genes *lga2* and *rga2* direct uniparental mitochondrial

53. **Mahlert M, Vogler C, Stelter K, Hause G, Basse CW.** 2009. The *a2* mating-type-locus gene *lga2* of *Ustilago maydis* interferes with mitochondrial dynamics and fusion, partially in dependence on a Dnm1-like fission component. J. Cell Sci. **122**(Pt 14):2402–2412.

54. **Schulz B, et al.** 1990. The *b* alleles of *U. maydis*, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif. Cell **60**:295–306.

55. **Kämper J, Reichmann M, Romeis T, Bölker M, Kahmann R.** 1995. Multiallelic recognition: nonself-dependent dimerization of the bE and bW homeodomain proteins in *Ustilago maydis*. Cell **81**:73–83.

56. **Lee N, Bakkeren G, Wong K, Sherwood JE, Kronstad JW.** 1999. The mating-type and pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. Proc. Natl. Acad. Sci. U. S. A. **96**:15026–15031.

57. **Bölker M, Urban M, Kahmann R.** 1992. The a mating type locus of *U. maydis* specifies cell signaling components. Cell **68**:441–450.

58. **Bakkeren G, et al.** 2006. Mating factor linkage and genome evolution in basidiomycetous pathogens of cereals. Fungal Genet. Biol. **43**:655–666.

59. **Bakkeren G, Kronstad JW.** 1994. Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. Proc. Natl. Acad. Sci. U. S. A. **91**:7085–7089.

60. **Hsueh YP, Heitman J.** 2008. Orchestration of sexual reproduction and virulence by the fungal mating-type locus. Curr. Opin. Microbiol. **11**:517–524.

61. **Regenfelder E, et al.** 1997. G proteins in *Ustilago maydis*: transmission of multiple signals? EMBO J. **16**:1934–1942.

62. **Li L, Wright SJ, Krystofova S, Park G, Borkovich KA.** 2007. Heterotrimeric G protein signaling in filamentous fungi. Annu. Rev. Microbiol. **61**:423–452.

63. **Li L, et al.** 2007. Canonical heterotrimeric G proteins regulating mating and virulence of *Cryptococcus neoformans*. Mol. Biol. Cell **18**:4201–4209.

64. **Whiteway M, et al.** 1989. The *STE4* and *STE18* genes of yeast encode potential β and γ subunits of the mating factor receptor-coupled G protein. Cell **56**:467–477.

65. **Wang L, Berndt P, Xia X, Kahnt J, Kahmann R.** 2011. A seven-WD40 protein related to human RACK1 regulates mating and virulence in *Ustilago maydis*. Mol. Microbiol. **81**:1484–1498.

66. **Schurko AM, Logsdon JM.** 2008. Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. Bioessays **30**:579–589.

67. **Malik SB, Pightling AW, Stefaniak LM, Schurko AM, Logsdon JM, Jr.** 2008. An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. PLoS One **3**:e2879. http://dx.doi.org/10.1371/journal.pone.0002879.

68. **Masson JY, West SC.** 2001. The Rad51 and Dmc1 recombinases: a non-identical twin relationship. Trends Biochem. Sci. **26**:131–136.

69. **Reedy JL, Floyd AM, Heitman J.** 2009. Mechanistic plasticity of sexual reproduction and meiosis in the *Candida* pathogenic species complex. Curr. Biol. **19**:891–899.

70. **Hayase A, et al.** 2004. A protein complex containing Mei5 and Sae3 promotes the assembly of the meiosis-specific RecA homolog Dmc1. Cell **119**:927–940.

71. **Tsubouchi H, Roeder GS.** 2004. The budding yeast Mei5 and Sae3 proteins act together with Dmc1 during meiotic recombination. Genetics **168**:1219–1230.

72. **Villeneuve AM, Hillers KJ.** 2001. Whence meiosis? Cell **106**:647–650.

73. **Malik SB, Ramesh MA, Hulstrand AM, Logsdon JM.** 2007. Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. Mol. Biol. Evol. **24**:2827–2841.

74. **Buonomo SB, et al.** 2000. Disjunction of homologous chromosomes in meiosis I depends on proteolytic cleavage of the meiotic cohesin Rec8 by separin. Cell **103**:387–398.

75. **Xu H, et al.** 2004. A new role for the mitotic RAD21/SCC1 cohesin in meiotic chromosome cohesion and segregation in the mouse. EMBO Rep. **5**:378–384.

76. **Prieto I, et al.** 2002. STAG2 and Rad21 mammalian mitotic cohesins are implicated in meiosis. EMBO Rep. **3**:543–550.

77. **Donaldson ME, Saville BJ.** 2008. Bioinformatic identification of *Ustilago maydis* meiosis genes. Fungal Genet. Biol. **45**(Suppl 1):S47–S53.

78. **Holloman WK, Schirawski J, Holliday R.** 2008. The homologous re-

DNA (mtDNA) inheritance and constrain mtDNA recombination during sexual development of *Ustilago maydis*. Genetics **181**:847–860.

combination system of *Ustilago maydis*. Fungal Genet. Biol. **45**(Suppl 1):S31–S39.

79. **Loidl J.** 2006. *S. pombe* linear elements: the modest cousins of synaptonemal complexes. Chromosom **115**:260–271.

80. **Ramesh MA, Malik SB, Logsdon JM.** 2005. A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. Curr. Biol. **15**:185–191.

81. **Theelen B, Silvestri M, Guého E, van Belkum A, Boekhout T.** 2001. Identification and typing of *Malassezia* yeasts using amplified fragment length polymorphism (AFLPTm), random amplified polymorphic DNA (RAPD) and denaturing gradient gel electrophoresis (DGGE). FEMS Yeast Res. **1**:79–86.

82. **Midreuil F, et al.** 1999. Genetic diversity in the yeast species *Malassezia pachydermatis* analysed by multilocus enzyme electrophoresis. Int. J. Syst. Bacteriol. **49**:1287–1294.

83. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**:578–579.

84. **Cantarel BL, et al.** 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. **18**:188–196.

85. **Holt C, Yandell M.** 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics **12**:491.

86. **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.

87. **Stanke M, Waack S.** 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19**(Suppl 2):ii215–ii225.

88. **Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M.** 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Res. **18**:1979–1990.

89. **Korf I.** 2004. Gene finding in novel genomes. BMC Bioinformatics **5**:59.

90. **Al-Reedy RM, Malireddy R, Dillman CB, Kennell JC.** 2012. Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. Fungal Genet. Biol. **49**:2–14.

91. **Eriksson H, et al.** 2008. Quantitative membrane proteomics applying narrow range peptide isoelectric focusing for studies of small cell lung cancer resistance mechanisms. Proteomics **8**:3008–3018.

92. **Larkin MA, et al.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947–2948.

93. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

94. **Yang Z, Nielsen R.** 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. **17**:32–43.

95. **Guindon S, et al.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.

96. **Leeming JP, Notman FH.** 1987. Improved methods for isolation and enumeration of *Malassezia furfur* from human skin. J. Clin. Microbiol. **25**:2017–2019.

97. **Zargari A, Midgley G, Bäck O, Johansson SG, Scheynius A.** 2003. IgE-reactivity to seven *Malassezia* species. Allergy **58**:306–311.

98. **Kurtz S, et al.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**:R12.

99. **Rozen S, Skaletsky H.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. **132**:365–386.

100. **Carver TJ, et al.** 2005. ACT: the Artemis comparison tool. Bioinformatics **21**:3422–3423.

101. **Ruiz-Herrera J, Ortiz-Castellanos L, Martínez AI, León-Ramírez C, Sentandreu R.** 2008. Analysis of the proteins involved in the structure and synthesis of the cell wall of *Ustilago maydis*. Fungal Genet. Biol. **45**(Suppl 1):S71–S76.

102. **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.

103. **Yasueda H, et al.** 1998. Identification and cloning of two novel allergens from the lipophilic yeast, *Malassezia furfur*. Biochem. Biophys. Res. Commun. **248**:240–244.

104. **Onishi Y, et al.** 1999. Two-dimensional electrophoresis of *Malassezia* allergens for atopic dermatitis and isolation of Mal f 4 homologs with mitochondrial malate dehydrogenase. Eur. J. Biochem. **261**:148–154.

105. **Andersson A, et al.** 2004. Cloning, expression and characterization of two new IgE-binding proteins from the yeast *Malassezia sympodialis* with sequence similarities to heat shock proteins and manganese superoxide dismutase. Eur. J. Biochem. **271**:1885–1894.

106. **Limacher A, et al.** 2007. Cross-reactivity and 1.4-Ångstrom crystal structure of *Malassezia sympodialis* thioredoxin (Mala s 13), a member of a new pan-allergen family. J. Immunol. **178**:389–396.

107. **Jain E, et al.** 2009. Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics **10**:136.

108. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. **27**:221–224.