Whole-Genome Analysis Illustrates Global Clonal Population Structure of the Ubiquitous Dermatophyte Pathogen *Trichophyton rubrum*

Gabriela F. Persinoti,*,[1] Diego A. Martinez,†,[1,2] Wenjun Li,‡,[1,3] Aylin Döğen,‡,§ R. Blake Billmyre,‡,[4] Anna Averette,‡ Jonathan M. Goldberg,†,[5] Terrance Shea,† Sarah Young,† Qiandong Zeng,†,[6] Brian G. Oliver,** Richard Barton,†† Banu Metin,‡‡ Süleyha Hilmioğlu-Polat,§§ Macit Ilkit,*** Yvonne Gräser,††† Nilce M. Martinez-Rossi,* Theodore C. White,‡‡‡ Joseph Heitman,‡,[7] and Christina A. Cuomo†,[7]

*Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Brazil 14049-900, †Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, ‡Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710, §Department of Pharmaceutical Microbiology, Faculty of Pharmacy, University of Mersin, Turkey 33110, **Center for Infectious Disease Research, Seattle, Washington 98109, ††School of Molecular and Cellular Biology, University of Leeds, United Kingdom LS2 9JT, ‡‡Department of Food Engineering, Faculty of Engineering and Natural Sciences, Istanbul Sabahattin Zaim University, Turkey, §§Department of Microbiology, Faculty of Medicine, University of Ege, Izmir, Turkey 35100, ***Division of Mycology, Department of Microbiology, Faculty of Medicine, University of Çukurova, Adana, Turkey 01330, †††Institute of Microbiology and Hygiene, University Medicine Berlin - Charité, Germany 12203, and ‡‡‡School of Biological Sciences, University of Missouri–Kansas City, Missouri 64110

ORCID IDs: 0000-0002-0975-7283 (G.F.P.); 0000-0002-8518-8502 (D.A.M.); 0000-0002-0388-306X (A.D.); 0000-0003-4866-3711 (R.B.B.); 0000-0002-1174-4182 (M.I.); 0000-0002-5723-4276 (N.M.M.-R.); 0000-0001-6369-5995 (J.H.); 0000-0002-5778-960X (C.A.C.)

ABSTRACT Dermatophytes include fungal species that infect humans, as well as those that also infect other animals or only grow in the environment. The dermatophyte species *Trichophyton rubrum* is a frequent cause of skin infection in immunocompetent individuals. While members of the *T. rubrum* species complex have been further categorized based on various morphologies, their population structure and ability to undergo sexual reproduction are not well understood. In this study, we analyze a large set of *T. rubrum* and *T. interdigitale* isolates to examine mating types, evidence of mating, and genetic variation. We find that nearly all isolates of *T. rubrum* are of a single mating type, and that incubation with *T. rubrum* "morphotype" *megninii* isolates of the other mating type failed to induce sexual development. While the region around the mating type locus is characterized by a higher frequency of SNPs compared to other genomic regions, we find that the population is remarkably clonal, with highly conserved gene content, low levels of variation, and little evidence of recombination. These results support a model of recent transition to asexual growth when this species specialized to growth on human hosts.

KEYWORDS *Trichophyton rubrum*; *Trichophyton interdigitale*; dermatophyte; genome sequence; MLST; mating; recombination; LysM

D ERMATOPHYTE species are the most common fungal species causing skin infections. Of the > 40 different species infecting humans, *Trichophyton rubrum*, the major cause of athlete's foot, is the most frequently observed (Achterman and White 2013; White *et al.* 2014). Other species are more often found on other skin sites, such as those found on the head, including *T. tonsurans* and *Microsporum canis*. Some dermatophyte species only cause human infections, including *T. rubrum*, *T. tonsurans*, and *T. interdigitale*. Other species, including *T. benhamiae*, *T. equinum*, *T. verrucosum*, and *M. canis*, infect mainly animals and occasionally

humans, while others such as *M. gypseum* [*Nannizzia gypsea* (G. S. de Hoog *et al.* 2017)] are commonly found in soil and rarely infect animals. In addition to the genera *Trichophyton* and *Microsporum*, *Epidermophyton* and *Nannizzia* are other genera of dermatophytes that commonly cause infections in humans (S. de Hoog *et al.* 2017). The species within these genera are closely related phylogenetically and are within the Ascomycete order Onygenales, family Arthrodermataceae (White *et al.* 2008; G. S. de Hoog *et al.* 2017).

The *T. rubrum* species complex includes several morphotypes, many of which rarely cause disease, and *T. violaceum*, a

species that causes scalp infections (Gräser *et al.* 2000; G. S. de Hoog *et al.* 2017). Some morphotypes display phenotypic variation, though these differences can be modest. For example, the *T. rubrum* morphotype *raubitscheckii* differs from *T. rubrum* in production of urease, and in colony pigmentation and colony appearance under some conditions (Kane *et al.* 1981). *T. rubrum* morphotype *megninii*, which is commonly isolated in Mediterranean countries, requires L-histidine for growth unlike other *T. rubrum* isolates (Gräser *et al.* 2000). However, little variation has been observed between these and other morphotypes in the sequence of individual loci, such as the internal transcribed spacer (ITS) ribosomal DNA (rDNA) locus; additionally, some of the morphotypes do not appear to be monophyletic (Gräser *et al.* 2000, 2007; G. S. de Hoog *et al.* 2017), complicating any simple designation of all types as separate species. Combining morphological and multilocus sequence typing (MLST) data has helped to clarify relationships of the major genera of dermatophytes and resolved polyphyletic genera initially assigned by morphological or phenotypic data.

Mating has been observed in some dermatophyte species, although not to date in strict anthropophiles including *T. rubrum* (Metin and Heitman 2017). Mating type in dermatophytes, as in other Ascomycetes, is specified by the presence of one of two idiomorphs at a single mating type (*MAT*) locus; each idiomorph includes either an α-box domain or a high mobility group (HMG) domain transcription factor gene (Li *et al.* 2010). In the geophilic species *M. gypseum*, isolates of opposite mating type (*MAT1-1* and *MAT1-2*) undergo mating and produce recombinant progeny (Li *et al.* 2010). In the zoophilic species *T. benhamiae*, both mating types are detected in the population and mating assays produced fertile cleistothecia (Symoens *et al.* 2013), structures that contain meiotic ascospores. In a study examining 600 isolates of *T. rubrum*, only five appeared to produce structures similar to cleistothecia (Young 1968), suggesting inefficient development of the spores required for mating. Sexual reproduction experiments of *T. rubrum* with tester strains of *T. simii*, a skin-infecting species that is closely related to *T. mentagrophytes*,

have been reported and one recombinant isolate was characterized, consistent with a low frequency of mating of *T. rubrum* (Anzawa *et al.* 2010). Further, sexual reproduction of *T. rubrum* may be rare in natural populations, as a single mating type (*MAT1-1*) has been noted in Japanese isolates (Kano *et al.* 2013), matching that described in the *T. rubrum* reference genome of CBS 118892 (Li *et al.* 2010).

Here, we describe genome-wide patterns of variation in *T. rubrum*, revealing a largely clonal population. This builds on prior work to produce reference genomes for *T. rubrum* (Martinez *et al.* 2012) and other dermatophytes (Burmester *et al.* 2011; Martinez *et al.* 2012). Genomic analysis of two divergent morphotypes of *T. rubrum*, *T. megninii*, and *T. soudanense* reveal hotspots of variation linked to the mating type locus suggestive of recent recombination. While nearly all *T. rubrum* isolates are of a single mating type (*MAT1-1*), the sequenced *megninii* morphotype isolate contains a *MAT1-2* locus, suggesting the capacity for infrequent mating in the population. Additionally, we examine variation in gene content across dermatophyte genomes including the first representatives of *T. interdigitale*.

Materials and Methods

Isolate selection, growth conditions, and DNA isolation

Isolates analyzed are listed in Supplemental Material, Table S1, including the geographic origin, site of origin, and mating type for each. Isolates selected for whole-genome sequencing were chosen to maximize diversity by covering the main known groups. For whole-genome sequencing, 10 *T. rubrum* isolates and 2 *T. interdigitale* isolates were selected, including representatives of the major morphotypes of *T. rubrum* (Table S2). Growth and DNA isolation for whole-genome sequencing were performed as previously described (Martinez *et al.* 2012).

For MLST analysis, a total of 80 *T. rubrum* isolates and 11 *T. interdigitale* isolates were selected for targeted sequencing. Isolates were first grown on potato dextrose agar (PDA) medium (Difco, Detroit, MI) for 10 days at 25°. Genomic DNA was extracted using an Epicentre Masterpure Yeast DNA purification kit (catalog number MPY08200). Fungal isolates were harvested from solid medium using sterile cotton swabs, transferred to microcentrifuge tubes, and washed with sterile PBS. Glass beads (2 mm) and 300 μl yeast cell lysis solution (Epicentre) were added to the tube to break down fungal cells, and the protocol provided by Epicentre was then followed. The contents of the tube were mixed by vortexing and incubated at 65° for 30 min, followed by addition of 150 μl Epicentre MPC Protein Precipitation Solution. After vortexing, the mixture was centrifuged for 10 min, followed by isopropanol precipitation and washing with 70% ethanol. The DNA pellet was dissolved in TE buffer.

For mating assays, we investigated 55 *T. rubrum* and 9 *T. interdigitale* isolates recovered from Adana and Izmir, Turkey. *T. simii* isolates CBS 417.65 MT −, CBS 448.65 MT +, and morphotype *megninii* isolates CBS 389.58, CBS 384.64, and CBS

[1]These authors contributed equally to this work.
[2]Present address: Veritas Genetics, Danvers, MA 01923.
[3]Present address: National Center for Biotechnology Information (NCBI), 8600 Rockville Pike, Bethesda, MD 20894.
[4]Present address: Stowers Institute for Medical Research, Kansas City, MO 64110.
[5]Present address: Harvard T. H. Chan School of Public Health, Boston, MA 02115.
[6]Present address: LabCorp, Westborough, MA 01581.
[7]Corresponding authors: Department of Molecular Genetics and Microbiology, 322 CARL Building, Box 3546, Duke University Medical Center, Durham, NC 27710. E-mail: heitm001@duke.edu; and Broad Institute, 415 Main Street, Cambridge, MA 02142. E-mail: cuomo@broadinstitute.org

417.52 were also used in mating assays. DNA extraction was performed according to the protocol described by Turin *et al.* (2000). These isolates were typed by ITS sequence analysis. rDNA sequences spanning the ITS region were PCR-amplified using the universal fungal primers ITS1 (5′-TCCGTAGGT GAACCTGCGG-3′) and ITS4 (5′-CCTCCGCTTATTGATATGC-3′), and sequenced on an ABI PRISM 3130XL genetic analyzer at Refgen Biotechnologies using the same primers (Ankara, Turkey). CAP contig assembly software, included in the BioEdit Sequence Alignment Editor 7.0.9.0 software package, was used to edit the sequences (Hall 1999). Assembled DNA sequences were characterized using BLAST (Basic Local Alignment Search Tool) in GenBank.

MLST

A total of 108 isolates were subjected to MLST analysis (Table S3). For each isolate, three loci [the *TruMDR1* ABC transporter (Cervelatti *et al.* 2006), an intergenic region (IR), and an α-1,3-mannosyltransferase (CAP59 protein domain)], with high sequence diversity between *T. rubrum* CBS 118892 (GenBank accession: NZ_ACPH00000000) and *T. tonsurans* CBS 112818 (GenBank accession: ACPI00000000), were selected as molecular markers in MLST. The following conditions were used in the PCR amplification of the three loci: an initial 2 min of denaturation at 98°, followed by 35 cycles of denaturation for 10 sec at 98°, an annealing time of 15 sec at 54°, and an extension cycle for 1 min at 72°. The amplification was completed with an extension period of 5 min at 72°. PCR amplicons were sequenced using the same PCR primers on an ABI PRISM 3130XL genetic analyzer by Genewiz, Inc. (Table S4). Electropherograms of Sanger sequencing were examined and assembled using Sequencher 4.8 (Gene Codes). Alternatively, sequences were obtained from genome assemblies (Table S5).

To confirm the species typing for four isolates (MR857, MR827, MR816, and MR897), the ITS1, 5.8S, and ITS2 regions were amplified using the ITS5 (5′-GAAGTAAAAGTCGTAACAAGG-3′) and Mas266 (5′-GCATTCCCAAACAACTCGACTC-3′) primers with initial denaturation at 94° for 4 min, 35 cycles of denaturation at 94° for 30 sec, annealing at 60° for 30 sec, extension at 72° for 1 min, and final extension at 72° for 10 min. The reactions were carried out using a Bio-Rad C1000 Touch thermocycler (Hercules, CA). ABI sequencing reads were compared to the dermatophyte database of the Westerdijk Fungal Biodiversity Institute. The sequences of MR857 and MR827 isolates were 99.6% identical to that of the isolate RV 30,000 of the African race of *T. benhamiae* (GenBank AF170456).

Mating type determination

To identify the mating type of each isolate, primers were designed to amplify either the α or HMG domain of *T. rubrum* (Table S6). For most isolates, PCR amplification was performed using an Eppendorf epGradient Mastercycler and reactions were carried out using the following conditions for amplification: initial denaturation at 94° for 4 min, 35 cycles of denaturation at 94° for 30 sec, annealing at 55° for 30 sec,

and extension at 72° for 1 min, with a final extension at 72° for 7 min. For isolates from Turkey, PCR amplifications were performed with the same primers using a Bio-Rad C1000 Touch Thermal Cycler and slightly modified conditions were used for amplification: initial denaturation at 94° for 5 min, 35 cycles of denaturation at 95° for 45 sec, annealing at 55° for 1.5 min, and extension at 72° for 1 min, with final extension at 72° for 10 min. The presence of the α-box gene, which is indicative of the *MAT1-1* mating type, or the HMG domain, which is indicative of the *MAT1-2* mating type, was identified using primers JOHE21771/WL and JOHE21772/WL, creating a 500-bp product, and JOHE21773/WL and JOHE21774/ WL, creating a 673-bp product, respectively. *T. rubrum* MR 851 was used as a positive control for *MAT1-1*, and morphotype *megninii* CBS 389.58, CBS 384.64, CBS 417.52, and *T. interdigitale* MR 8801 were used as positive controls for *MAT1-2*. The mating type was assigned based on the presence or absence of PCR products on 1.5% agarose gels. For the whole-genome sequenced isolates, mating type was determined by analysis of assembled and annotated genes.

Mating assays

Mating assays were performed using both Medium E [12 g/ liter oatmeal agar (Difco), 1 g/liter MgSO$_4$0.7H$_2$O, 1 g/liter NaOH$_3$, 1 g/liter KH$_2$PO$_4$, and 16 g/liter agar (Weitzman and Silva-Hutner 1967)] and Takashio medium (1/10 Sabouraud containing 0.1% neopeptone, 0.2% dextrose, 0.1% MgSO$_4$ 0.7H$_2$O, and 0.1%. KH$_2$PO$_4$). *MAT1-1* and *MAT1-2* isolates grown on Sabouraud Dextrose Agar for 1 week were used to inoculate both Medium E and Takashio medium plates pairwise 1 cm apart from each other. The plates were incubated at room temperature without parafilm in the dark for 4 weeks. The petri dishes were examined under light microscopy for sexual structures.

Genome sequencing, assembly, and annotation

For genome sequencing, we constructed a 180-base fragment library from each sample, by shearing 100 ng of genomic DNA to a median size of ∼250 bp using a Covaris LE instrument and preparing the resulting fragments for sequencing as previously described (Fisher *et al.* 2011). Each library was sequenced on the Illumina HiSeq 2000 platform. Roughly 100× of 101-bp Illumina reads were assembled using ALLPATHS-LG (Gnerre *et al.* 2011) run with an assisting mode utilizing *T. rubrum* CBS118892 as a reference. For most genomes, assisting mode 2 was used (ASSISTED_PATCHING = 2) with version R42874; for *T. interdigitale* H6 and *T. rubrum* MR1463, version R44224 was used. For *T. rubrum* morphotype *megninii* CBS 735.88 and *T. rubrum* morphotype *raubitschekii* CBS 202.88, mode 2.1 was used (ASSISTED_PATCHING = 2.1) with version R47300. Assemblies were evaluated using GAEMR (http://software. broadinstitute.org/software/gaemr/); contigs corresponding to the mitochondrial genome or contaminating sequence from other species were removed from assemblies.

The *Trichophyton* assemblies were annotated using a combination of expression data, conservation information, and *ab*

initio gene-finding methods as previously described (Haas *et al.* 2011). Expression data included Illumina reads (SRX123796) from one RNA-sequencing (RNA-Seq) study (Ren *et al.* 2012) and all EST data available in GenBank as of 2012. RNA-Seq reads were assembled into transcripts using Trinity (Grabherr *et al.* 2011). PASA (Haas *et al.* 2003) was used to align the assembled transcripts and ESTs to the genome and identify open reading frames (ORFs); gene structures were also updated in the previously annotated *T. rubrum* CBS118892 assembly (Martinez *et al.* 2012). Conserved loci were identified by comparing the genome with the UniRef90 database (Wu *et al.* 2006) (updated in 2012) using BLAST (Altschul *et al.* 1997). The BLAST alignments were used to generate gene models using Genewise (Birney *et al.* 2004). The *T. rubrum* CBS 118892 genome was aligned with the new genomes using NUCmer (Kurtz *et al.* 2004). These alignments were used to map gene models from *T. rubrum* to conserved loci in the new genomes.

To predict gene structures, GeneMark, which is self-training, was applied first; GeneMark models matching GeneWise ORF predictions were used to train the other *ab initio* programs. *Ab intio* gene-finding methods included GeneMark (Borodovsky *et al.* 2003), Augustus (Stanke *et al.* 2004), SNAP (Korf 2004), and Glimmer (Majoros *et al.* 2004). Next, EvidenceModeler (EVM) (Haas *et al.* 2008) was used to select the optimal gene model at each locus. The input for EVM included aligned transcripts from Trinity and ESTs, gene models created by PASA and GeneWise, mapped gene models, and *ab initio* predictions. Rarely, EVM failed to produce a gene model at a locus likely to encode a gene. If alternative gene models existed at such loci, they were added to the gene set if they encoded proteins longer than 100 amino acids, or if the gene model was validated by the presence of a Pfam domain or expression evidence. Finally, PASA was run again to improve gene model structure, predict splice variants, and add UTRs.

Gene model predictions in repetitive elements were identified and removed from gene sets if they overlapped TransposonPSI predictions (http://transposonpsi.sourceforge.net), contained Pfam domains known to occur in repetitive elements, or had BLAST hits against the Repbase database (Jurka *et al.* 2005). Additional repeats were identified using a BLAT (Kent 2002) self-alignment of the gene set to the genomic sequence (requiring $\geq$ 90% nucleotide identity over 100 bases aligned); genes that hit the genome more than eight times using these criteria were removed. Genes with Pfam domains not found in repetitive elements were retained in the gene set, even if they met the above criteria for removing likely repetitive elements from the gene set.

Lastly, the gene set was inspected to address systematic errors. Gene models were corrected if they contained in-frame stop codons, had coding sequence overlaps with coding regions of other gene models, predicted transfer or ribosomal RNAs, contained exons spanning sequence gaps, had incomplete codons, or had UTRs overlapping the coding sequences of other genes. Transfer RNAs were predicted using tRNAscan (Lowe and Eddy 1997), and ribosomal RNAs were predicted with RNAmmer (Lagesen *et al.* 2007).

All annotated assemblies and raw sequence reads are available at the National Center for Biotechnology Information (NCBI) database (Table S5).

SNP identification and classification

To identify SNPs within the *T. rubrum* group, Illumina reads for each *T. rubrum* isolate were aligned to the *T. rubrum* CBS 118892 reference assembly using BWA-MEM (Li 2013); reads from H6 *T. interdigitale* were also aligned to the *T. interdigitale* MR 816 assembly. The Picard tools (http://picard.sourceforge.net) AddOrReplaceReadGroups, MarkDuplicates, CreateSequenceDictionary, and ReorderSam were used to preprocess read alignments. To minimize false-positive SNP calls near insertion/deletion (indel) events, poorly aligned regions were identified and realigned using GATK RealignerTargetCreator and IndelRealigner (GATK version 2.7-4 [McKenna *et al.* (2010), page 201]. SNPs were identified using the GATK UnifiedGenotyper (with the haploid genotype likelihood model) run with the SNP genotype likelihood models (GLM). We also ran BaseRecalibrator and PrintReads for base quality score recalibration on sites called using GLM SNP and recalled variants with UnifiedGenotyper emitting all sites. VCFtools (Danecek *et al.* 2011) was used to count SNP frequency in windows across the genome (–SNPdensity 5000) and to measure nucleotide diversity (–site-pi), which was normalized for the assembly size. For comparison, the nucleotide diversity was calculated for the SNPs identified in a set of 159 isolates of *Cryptococcus neoformans* var. *grubii*, a fungal pathogen that undergoes frequent recombination (Rhodes *et al.* 2017).

SNPs were mapped to genes using VCFannotator (http://vcfannotator.sourceforge.net/), which annotates whether a SNP results in a synonymous or nonsynonymous change in the coding region. The total number of synonymous and nonsynonymous sites across the *T. rubrum* CBS 118892 and *T. interdigitale* MR 816 gene sets were calculated across all coding regions using codeml in PAML (version 4.8) (Yang 2007); these totals were used to normalize the ratios of nonsynonymous to synonymous SNPs.

Copy number variation

To identify regions of *T. rubrum* that exhibit copy number variation between the isolates, we identified windows showing significant variation in normalized read depth using CNVnator (Abyzov *et al.* 2011). The realigned read files used for SNP calling were input to CNVnator version 0.2.5, specifying a window size of 1 kb. Regions reported as deletions or duplications were filtered requiring P-val1 $<$ 0.01.

Phylogenetic and comparative genomic analysis

To infer the phylogenetic relationship of the sequenced isolates, we identified single-copy genes present in all genomes using OrthoMCL (Li *et al.* 2003). Individual orthologs were aligned with MUSCLE (Edgar 2004) and then the alignments

were concatenated and input to RAxML (Stamatakis 2006), version 7.3.3 with 1000 bootstrap replicates and model GTRCAT. RAxML version 7.7.8 was used for phylogenetic analysis of SNP variants in seven *T. rubrum* isolates, with the same GTRCAT model.

For each gene set, HMMER3 (Eddy 2011) was used to identify Pfam domains using release 27 (Finn *et al.* 2014); significant differences in gene counts for each domain were identified using Fisher's exact test, with *P*-values corrected for multiple testing (Storey and Tibshirani 2003). Proteins with LysM domains were identified using a revised hidden Markov model (HMM) as previously described (Martinez *et al.* 2012); this HMM includes conserved features of fungal LysM domains, including conserved cysteine residues not represented in the Pfam HMM model, and identified additional genes with this domain.

Construction of paired allele compatibility matrix

To construct SNP profiles, SNPs shared by at least two members of the *T. rubrum* data set were selected. Private SNPs are not informative for a paired allele compatibility test because they can never produce a positive result. These profiles were then counted across the genome to construct SNP profiles via a custom Perl script. We required profiles to be present at least twice, to minimize the signal from homoplasic mutations. Pairwise tests were then conducted between each of the profiles to look for all four possible allele combinations, which would only occur via either mating or homoplasic mutations.

Linkage disequilibrium calculation

Linkage disequilibrium was calculated for *T. rubrum* SNPs in 1-kb windows of all scaffolds with VCFtools version 1.14 (Danecek *et al.* 2011), using the–hap-r2 option with a minimum minor allele frequency of 0.2.

Data availability

All genomic data are available in NCBI under the Umbrella BioProject PRJNA186851 and can be accessed via the accession numbers in Table S2. The NCBI GenBank accession numbers of the three MLST loci are listed in Table S3.

Results

Relationship of global Trichophyton isolates using MLST

To examine the relationship of global isolates of *T. rubrum*, we sequenced three loci in each of 104 *Trichophyton* isolates and carried out phylogenetic analysis. The typed isolates included 91 *T. rubrum* isolates, 11 *T. interdigitale* isolates, and 2 *T. benhamiae* isolates (Table S1). In addition, data from the genome assemblies of additional dermatophyte species (*T. verrucosum*, *T. tonsurans*, *T. equinum*, and *M. gypseum*) were also included. Three loci—the *TruMDR1* ABC transporter (Cervelatti *et al.* 2006), an IR, and an α-1,3-mannosyltransferase (CAP59 protein domain)—were sequenced in each isolate.

Phylogenetic analysis of the concatenated loci can resolve species boundaries between the seven species (Figure 1). A large branch separates a *T. benhamiae* isolate (MR857) from the previously described genome sequenced isolate (CBS 112371) (Figure 1), and the sequences of two loci of a second *T. benhamiae* isolate (MR827) were identical to those of MR857 (Table S3). Sequencing of the ITS region of the MR857 and MR827 isolates revealed high sequence similarity to isolates from the *T. benhamiae* African race (*Materials and Methods*), which is more closely related to *T. bullosum* than isolates of *T. benhamiae* Americano-European race including CBS112371 (Heidemann *et al.* 2010). Otherwise, the species relationships and groups are consistent between studies.

MLST analysis demonstrated that the *T. rubrum* isolates were nearly identical at the three sequenced loci. Remarkably, of the 84 *T. rubrum* isolates sequenced at all three loci, 83 were identical at all positions of the three loci sequenced (genotype 2, Table S3). Only one isolate, 1279, displayed a single difference at one site in the *TruMDR1* gene (genotype 3, Table S3). For the remaining six isolates, sequences at a subset of the loci were generated and matched that of the predominant genotype. Thus, MLST was not sufficient to discern the phylogenetic substructure in the *T. rubrum* population that included six isolates representing different morphotypes (Table S3). Similarly, the 11 *T. interdigitale* isolates were highly identical at these three loci; two groups were separated by a single-nucleotide difference in the IR and the third group contained a six-base deletion overlapping the same base of the IR (genotypes 1, 5, and 6, Table S3). Although most species can be more easily discriminated based on the MLST sequence, *T. equinum* and *T. tonsurans* isolates differed only by a single transition mutation in the IR, which illustrates the remarkable clonality of these species.

Genome sequencing and refinement of phylogenetic relationships

As MLST analysis was insufficient to resolve the population substructure of the *T. rubrum* species complex, we sequenced the complete genomes of *T. rubrum* isolates representing worldwide geographical origins and five morphotypes: *fischeri*, *kanei*, *megninii*, *raubitschekii*, and *soudanense*. We generated whole-genome Illumina sequences for 10 *T. rubrum* and 2 *T. interdigitale* isolates (Table S2). The sequence of each isolate was assembled and utilized to predict gene sets. The *T. rubrum* assembly size was very similar across isolates, ranging from 22.5 to 23.2 Mb (Table S5). The total predicted gene numbers were also similar across the isolates, with between 8616 and 9064 predicted genes in the 10 *T. rubrum* isolates, and 7993 and 8116 predicted genes in the two *T. interdigitale* isolates (Table S5).

To infer the phylogenetic relationship of these isolates and other previously sequenced *Trichophyton* isolates, we identified 5236 single-copy orthologs present in all species and estimated a phylogeny with RAxML (Stamatakis 2006) (Figure 2A). This phylogeny more precisely delineates the species
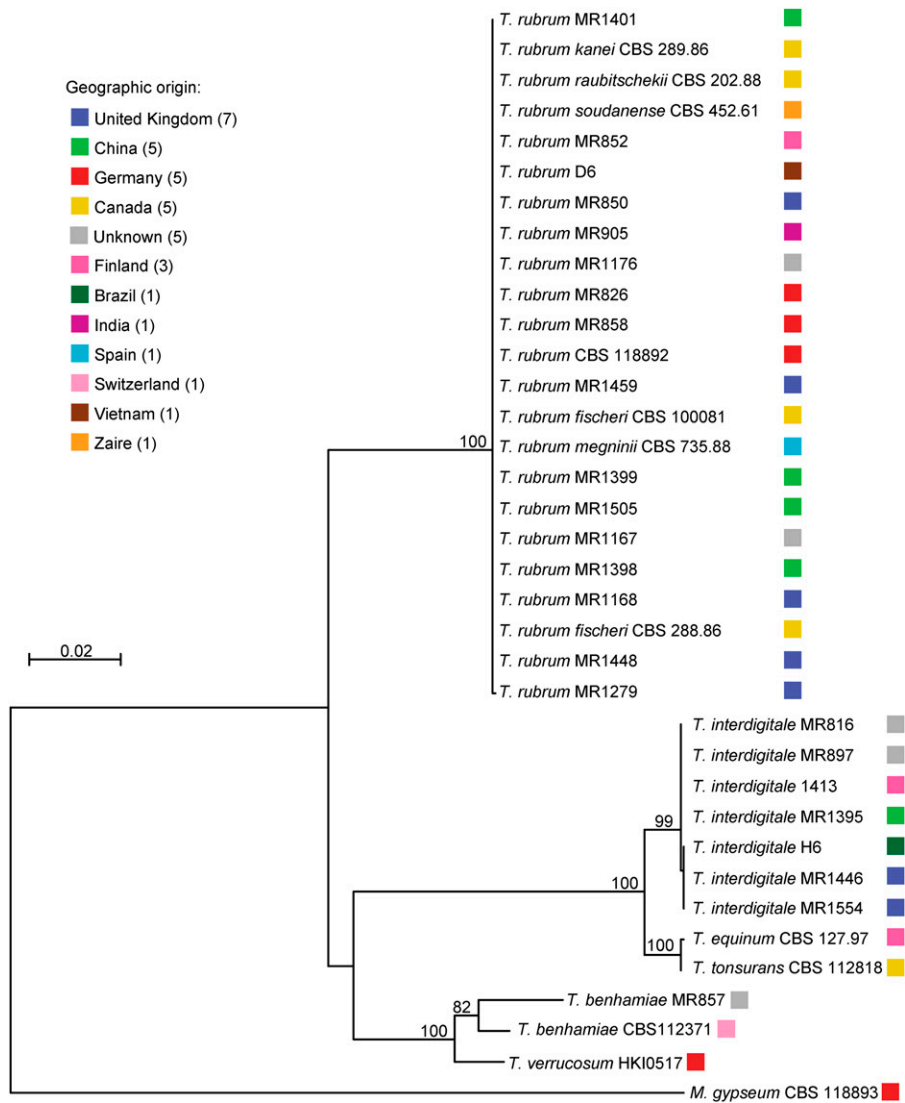
Figure 1 Phylogeny inferred from concatenated MLST sequences. Three MLST loci (ABC transporter, outer membrane protein, and CAP59 protein) were amplified and sequenced from 79 isolates and sequences were identified in an additional 19 assemblies. The concatenated sequence for each isolate was used to build a maximum likelihood tree using MEGA 5.2. Isolate MR1168 is representative of 73 *T. rubrum* isolates that have identical MLST sequences. MLST, multilocus sequence typing.

groups than that derived from the MLST loci and also illustrates the relationship between the *T. rubrum* isolates (Figure 2B). The results of this analysis suggest that the *fischeri* morphotype is not monophyletic, as one *fischeri* isolate (CBS100081) is more closely related to the *raubitschekii* isolate than to the other *fischeri* isolate (CBS 288.86). While a subset of seven *T. rubrum* isolates appear closely related, others show much higher divergence, including the *soudanense* isolate, the *megninii* isolate, the MR1459 isolate, and the CBS 118892 isolate representing the reference genome. The *soudanense* isolate (CBS 452.61) was placed as an outgroup relative to the other *T. rubrum* isolates; this is consistent with this isolate being part of a clade more closely related to *T. violaceum* than to *T. rubrum* (Gräser *et al.* 2000) and with the reestablishment of *soudanense* isolates as a separate species (G. S. de Hoog *et al.* 2017).

To further classify the two *T. interdigitale* isolates, we assembled the ITS region of the ribosomal DNA locus and compared the sequences to previously classified ITS sequences, as *T. interdigitale* isolates differ from *T. mentagrophytes* at the ITS locus (Gräser *et al.* 2008; G. S. de Hoog *et al.* 2017). For the two genomes of these species that we sequenced, MR816 was identical to *T. interdigitale* at the ITS1 locus, whereas the H6 isolate appears intermediate between *T. interdigitale* and *T. mentagrophytes*, containing polymorphisms specific to each group (Figure S1). Genomic analysis of allele sharing across a wider set of *T. interdigitale* and *T. mentagrophytes* isolates could be used to evaluate the extent of hybrid genotypes and genetic exchange between these two species.

MAT1-1 prevalence and clonality in *T. rubrum*

To address if the *T. rubrum* population is capable of sexual reproduction, we surveyed the *MAT* locus of all isolates. Using either gene content in assembled isolates or a PCR assay to assign mating type, we found that 79 of the 80 *T. rubrum* isolates contained the α-domain gene at the *MAT* locus (*MAT1-1*). In addition, a set of 55 isolates from Turkey were found to harbor the *MAT1-1* allele based on a PCR assay (Figure S2). However, the *T. rubrum* morphotype *megninii* isolate contained an HMG gene at the *MAT* locus (*MAT1-2*)
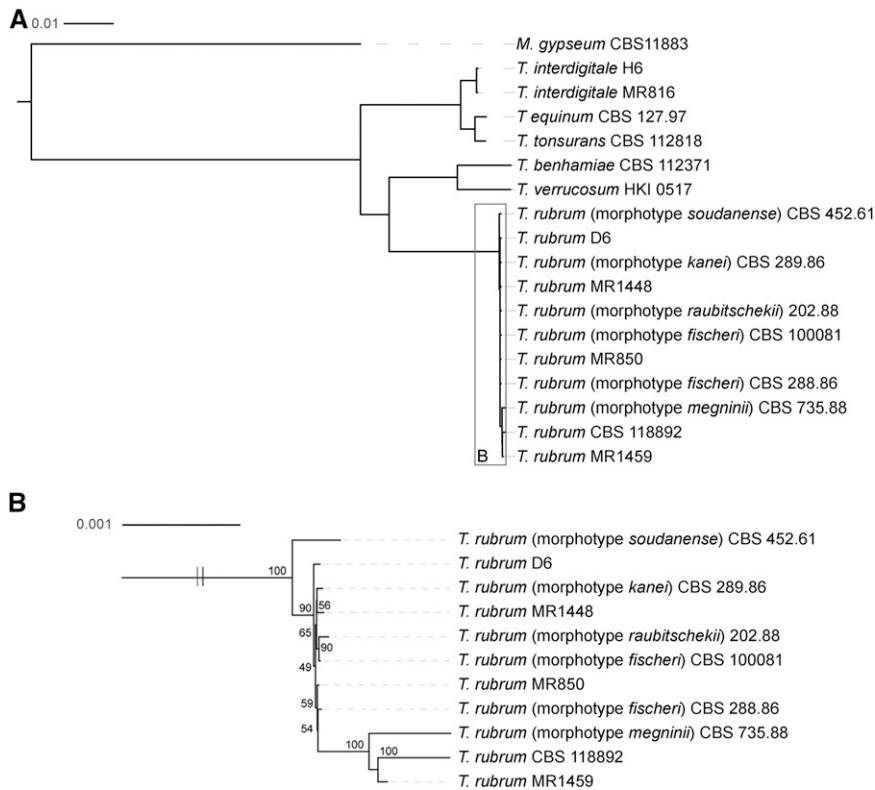
Figure 2 Phylogenetic relationship of *Trichophyton* isolates. A total of 5236 single-copy genes were each aligned with MUSCLE; the concatenated alignment was used to infer a species phylogeny with RAxML (GTRCAT model) with 1000 bootstrap replicates using either (A) all species including the outgroup *M. gypseum* or (B) only *T. rubrum* isolates.

(Figure 3 and Table S1). The presence of both mating types suggests that this species could be capable of mating under some conditions. However, the high frequency of a single mating type strongly suggests that *T. rubrum* largely undergoes clonal growth, although other interpretations are also possible (see *Discussion*). In further support of this, a study of 206 *T. rubrum* clinical isolates from Japan noted that all were of the *MAT1-1* mating type (Kano *et al.* 2013).

A closer comparison of the genome sequences of *T. rubrum* isolates also supports a clonal relationship of this population. Phylogenetic analysis of the seven most closely related *T. rubrum* isolates using SNPs between these isolates (see below) suggests that the isolates have a similar level of divergence from each other (Figure S3). This supports the idea that these *MAT1-1 T. rubrum* isolates have likely undergone clonal expansion.

To test for recombination that could reflect sexual reproduction within the *T. rubrum* population sampled here, we conducted a genome-wide paired allele compatibility test to look for the presence of all four products of meiosis (Figure 4). This test is a comparison between two paired polymorphic sites in the population. While the presence of three of the four possible allele combinations at two sites in a population is possible through a single mutation and identity-by-descent, the presence of all four combinations requires either recombination, or less parsimoniously, a second homoplasic mutation. Four positive tests resulted from this analysis (out of 21 possible), including allele combinations that occurred a minimum of 13 times. This may suggest that recombination

is a rare event arising through infrequent sexual recombination occurring in this population, although the same mutations and combinations arising via homoplasy (or selection) are difficult to exclude. Based on the number of triallelic sites in the data set (19), we would predict 9.5 homoplasic sites to have occurred by random chance, which is similar to the number of sites responsible for the positive signals in the compatibility test. In addition, linkage disequilibrium does not decay over increasing distance between SNPs in *T. rubrum* (Figure S4), which further supports a low level of recombination in this species; sequencing additional diverse isolates would help to address whether some isolates or lineages were more prone to recombination.

We also characterized the *MAT* locus of the newly sequenced *T. interdigitale* isolates (H6 and MR816) and found that both contain an HMG-domain gene. These *T. interdigitale* isolates were more closely related to *T. equinum* (*MAT1-2*) and *T. tonsurans* (*MAT1-1*) than *T. rubrum* (Figure 2A). To survey the mating type across a larger set of *T. interdigitale* isolates, a set of 11 additional isolates from Turkey were typed. Based on PCR analysis, all *T. interdigitale* isolates harbor the *MAT1-2* allele (Figure S2).

The mating abilities of the isolates were tested by conducting mating assays with potentially compatible isolates of *T. rubrum*, including the *megninii* morphotype, *T. interdigitale*, and *T. simii* (Table S7). These experiments were conducted using both Takashio and E medium at room temperature (~21–22°) without parafilm in the dark. Although the assay plates were incubated for longer than 5 months, ascomata or ascomatal initials
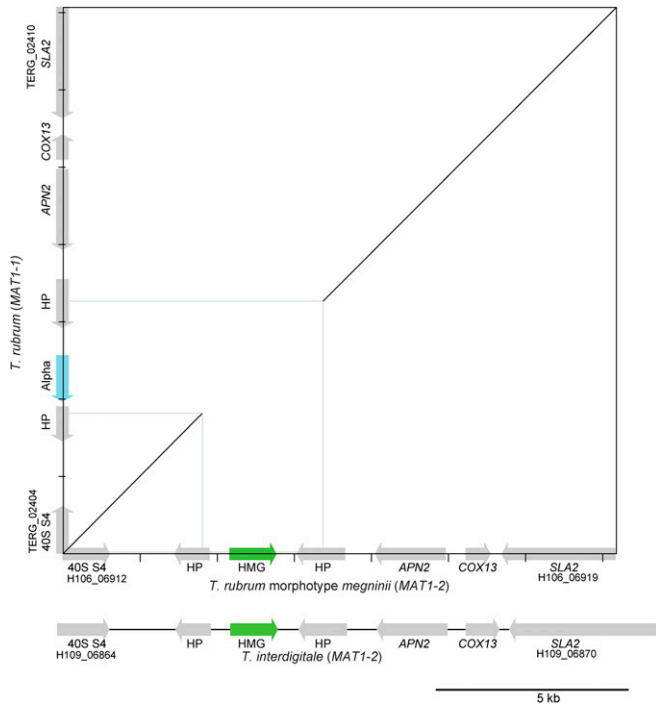
Figure 3 Alignment of the mating type locus of selected isolates. Mating type genes of *T. rubrum* morphotype *megninii* (CBS 735.88) and *T. rubrum* (CBS 188992) are shown along the *x*- and *y*-axes, respectively, with regions aligning by NUCmer show in the dot plot. The alignment extends into two hypothetical proteins (HP) immediately flanking the α- or high mobility group (HMG)-domain gene that specifies mating type. Most *T. rubrum* (MAT1-1) isolates contain an α-domain protein (blue) at the *MAT* locus. In contrast, the *T. rubrum* morphotype *megninii* isolate contains an HMG-domain protein (green) representing the opposite mating type (MAT1-2). All sequenced *T. interdigitale* isolates are also of MAT1-2 mating type including MR816. Gene locus identifiers are shown for the genes flanking each locus (prefix TERG, H106, and H109).

Figure 4 Paired allele compatibility test suggests limited evidence for sexual reproduction. (A) A single example of a positive paired allele compatibility test from the *T. rubrum* population. In this test, two loci are examined and typed across the population. To perform a meaningful test, at least two individuals in the population must share a variant allele at each site. Here, alternative SNPs are depicted in red and the reference in white. Evidence for recombination is provided by any pairwise comparison of two loci in which isolates are present where red–red, white–white, red–white, and white–red combinations are all found (AB, Ab, aB, and ab), satisfying the allele compatibility test and providing evidence for recombination. (B) Paired allele compatibility tests were performed for all isolates in the *T. rubrum* population across the entire genome. SNP profiles were grouped into unique and informative allele patterns and collapsed, with the number of occurrences of each profile across the genome listed. Thus, the larger the number, the more common that SNP distribution is in the population. Pairwise tests were then conducted for each combination of SNP profiles. Reference nucleotides are indicated by white and variants by red. The pairwise matrix displays the results of all of these tests; a green square in the pairwise matrix is indicative of a positive test for the pairwise comparison and thus provides potential evidence of recombination.

were not observed (Figure S5). While it is possible that mating may occur under cryptic conditions (Heitman 2010), this data suggests that the conditions tested are not sufficient for the initiation of mating structures in *T. rubrum*.

Genome-wide variation patterns in T. rubrum

SNP variants were identified between *T. rubrum* isolates to examine the level of divergence within this species complex (Table S8). On average, *T. rubrum* isolates contain 8092 SNPs compared to the reference genome of the CBS118892 isolate; this reflects a bimodal divergence pattern where most isolates, including three morphotypes (*fischeri*, *kanei*, and *raubitschekii*), have an average of 3930 SNPs and two more divergent isolates (morphotypes *megninii* and *soudanense*) have an average of 24,740 SNPs. The average nucleotide diversity (π) for all 10 *T. rubrum* isolates is 0.00054; excluding the two divergent morphotypes, the average nucleotide diversity is 0.00031. By comparison, the average nucleotide diversity of the fungal pathogen *C. neoformans* var. *grubii*, which is actively recombining as evidenced by low linkage disequilibrium (Desjardins *et al.* 2017; Rhodes *et al.* 2017), is 0.0074, a level ~24-fold higher than that in *T. rubrum* [Materials and Methods (Rhodes *et al.* 2017)]. Even
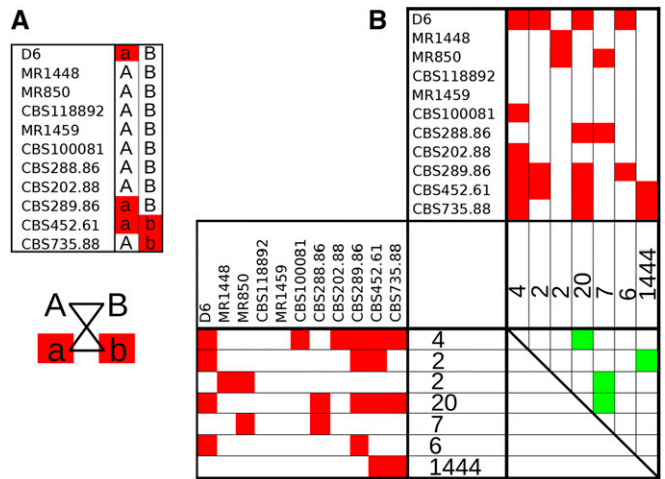
higher levels of nucleotide diversity have been reported in global populations of other fungi (see *Discussion*). A similar magnitude of SNPs separate the two *T. interdigitale* isolates; 22,568 SNPs were identified based on the alignment of H6 reads to the MR 816 assembly. Across all isolates, SNPs were predominantly found in IRs for both species, representing 76 and 81% of total variants, respectively (Table 1 and Table S8). Within genes, the higher ratio of nonsynonymous relative to synonymous changes among the closely related *T. rubrum* isolates (Table 1) is consistent with lower purifying selection over recent evolutionary time (Rocha *et al.* 2006).

Examining the frequency of SNPs across the *T. rubrum* genome revealed high-diversity regions that flank the mating type locus in the two divergent isolates. Across all isolates, some regions of the genome are overrepresented for SNPs, including the smallest scaffolds of the reference genome (Figure 5); these regions contain a high fraction of repetitive elements (Martinez *et al.* 2012). The largest high-diversity window unique to the *T. rubrum* morphotype *megninii* was found in an ~810-kb region encompassing the mating type locus on scaffold 2; a smaller high-diversity region spanning the

Table 1 Variation in _T. rubrum_ SNP rate and class

Isolate	Total number of SNPs	SNPs in CDS	SYN	NSY	pN/pS
T. rubrum MR1448	4,283	374	83	287	1.15
T. rubrum MR1459	2,188	436	103	317	1.02
T. rubrum MR850	4,203	387	88	289	1.09
T. rubrum D6	4,121	484	112	363	1.08
T. rubrum (morphotype _fischeri_) CBS 100081	4,199	409	94	307	1.09
T. rubrum (morphotype _fischeri_) CBS 288.86	4,147	375	84	283	1.12
T. rubrum (morphotype _kanei_) CBS 289.86	4,491	474	116	350	1.00
T. rubrum (morphotype _raubitschekii_) CBS 202.88	3,808	375	83	285	1.14
T. rubrum (morphotype _megninii_) CBS 735.88	26,406	7,328	3,069	4,185	0.45
T. rubrum (morphotype _soudanense_) CBS 452.61	23,073	6,253	2,377	3,808	0.53
T. interdigitale MR816	1,223,298	591,173	395,250	194,498	0.16
T. interdigitale H6	1,183,411	585,288	393,079	190,826	0.16

CDS, coding sequence; SYN, synonymous SNP sites; NSY, nonsynonymous SNP sites; pN/pS, (NSY/total NSY sites)/(SYN/total SYN sites).

mating type locus was found in the diverged _soudanense_ isolate (Figure 5). The higher diversity found in this location could reflect introgressed regions from recent outcrossing or could be associated with lower recombination proximal to the mating type locus, resulting in stratification of linked genes.

Gene content variation in _T. rubrum_ and _T. interdigitale_

To examine variation in gene content in the _T. rubrum_ species complex, we first measured copy number variation across the genome. Duplicated and deleted regions of the genome were identified based on significant variation in normalized read depth (_Materials and Methods_). We observed increased copy number only for two adjacent 26-kb regions of scaffold 4 in two isolates (MR850 and MR1448) (Figure S6). Both of these regions had nearly triploid levels of coverage (Table S9). While ploidy variation is a mechanism of drug resistance in fungal pathogens, none of the 25 total genes in these regions (Table S10) are known drug targets or efflux pumps. These regions include two genes classified as fungal zinc cluster transcription factors; this family of transcription factors was previously noted to vary in number between dermatophyte species (Martinez _et al._ 2012). A total of 12 deleted regions (CNVnator _P_-val < 0.01) ranging in size from 4 to 37 kb were also identified in a subset of genomes (Table S11). Two of these regions include genes previously noted to have higher copy number in dermatophyte genomes, a nonribosomal peptide synthase (NRPS) gene (TERG_02711) and a LysM gene (TERG_02813) (Table S12). Overall, this analysis suggests recent gain or loss in dermatophytes for a small set of genes including transcription factors, NRPS, and LysM-domain proteins.

We next examined candidate loss-of-function mutations in the _T. rubrum_ species complex. For the eight closely related _T. rubrum_ isolates, an average of 8.1 SNPs are predicted to result in new stop codons, disrupting protein-coding regions; in the _soudanense_ and _megninii_ isolates, an average of 58.5 SNPs result in new stop codons. These predicted loss-of-function mutations do not account for previously noted phenotypic differences between the morphotypes; no stop codons were found in the seven genes involved in histidine biosynthesis (_HIS1-HIS7_) in the histidine auxotroph _T. rubrum_ morphotype _megninii_ or in urease genes in _T. rubrum_ morphotype _raubitschekii_.

Comparison of the first representative genomes for _T. interdigitale_ (isolates MR816 and H6) to those of dermatophyte species highlighted the close relationship of _T. interdigitale_ to _T. tonsurans_ and _T. equinum_. These three species are closely related (Figure 2), sharing 7618 ortholog groups, yet there are also substantial differences in gene content. A total of 1253 ortholog groups were present only in _T. equinum_ and _T. tonsurans_, and 512 ortholog groups were present only in both _T. interdigitale_ isolates. However, there were no significant differences in functional groups between these species based on Pfam domain analysis, suggesting no substantial gain or loss of specific protein families. Two Pfam domains were unique to the _T. interdigitale_ isolates and present in more than one copy: PF00208, found in ELFV dehydrogenase family members and PF00187, a chitin-recognition protein domain. This chitin-binding domain is completely absent from the _T. equinum_ and _T. tonsurans_ genomes, while in _T. interdigitale_ this domain is associated with the glycosyl hydrolase family 18 (GH18) domain (Davies and Henrissat 1995). GH18 proteins are chitinases and some other members of this family also contain LysM domains. We also examined genes in the ergosterol pathway for variation, as this could relate to drug resistance; while this pathway is highly conserved in dermatophytes (Martinez _et al._ 2012), _T. interdigitale_ isolates had an extra copy of a gene containing the ERG4/ERG24 domain found in sterol reductase enzymes in the ergosterol biosynthesis pathway. The _ERG4_ gene encodes an enzyme that catalyzes the final step in ergosterol biosynthesis, and it is possible that an additional copy of this gene results in higher protein levels to help ensure that this step is not rate-limiting.

These comparisons also highlighted the recently discovered dynamics of the LysM family members, which bind bacterial peptidoglycan and fungal chitin (Buist _et al._ 2008). Dermatophytes contain high numbers of LysM-domain proteins ranging from the 10 genes found in _T. verrucosum_ to 31 copies found in _M. canis_ [Table S13 (Martinez _et al._ 2012)]. Both the class of LysM proteins with additional catalytic domains and the larger class consists of proteins with only LysM domains, many of which contain secretion signals and may represent candidate effectors (Martinez _et al._ 2012), vary in number across the dermatophytes. Isolates from the _T. rubrum_ species complex have 16–18 copies of
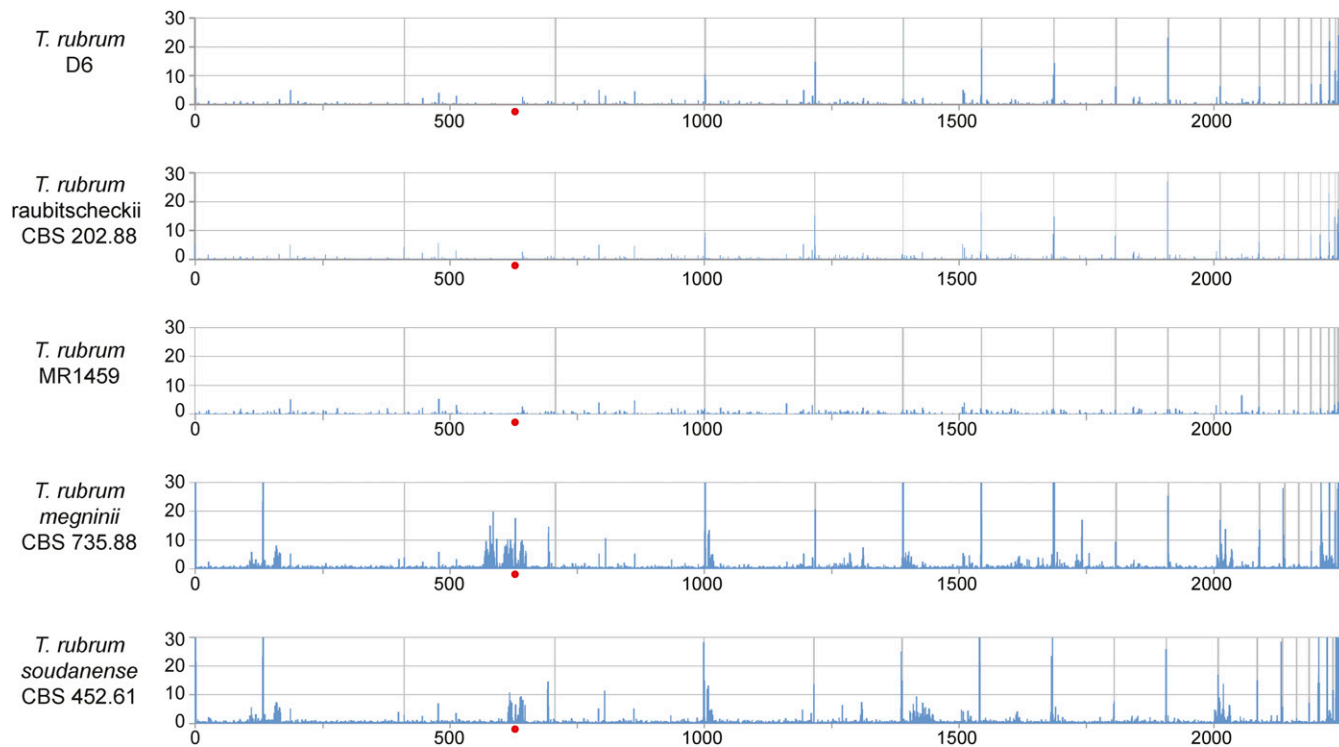
Figure 5 Genome-wide SNP frequency highlights hotspots. For each panel, the frequency of SNPs in 5-kb windows is shown across the genome. The genome assembly of isolate CBS 118892 was used for all comparisons, and scaffolds are ordered along the *x*-axis with gray lines representing scaffold boundaries. Red dots indicate the position of the mating type locus.

LysM proteins compared to the 15 found in the previously reported genome of the CBS 118892 isolate (Table S13). One of the additional LysM genes present in all of the newly sequenced isolates encodes a polysaccharide deacetylase domain involved in chitin catabolism. There is also an additional copy of a gene with only a LysM domain in 9 of the 10 new *T. rubrum* isolates (Table S13). The genomes of the *T. interdigitale* isolates have only 14 genes containing a LysM-binding domain and are missing a LysM gene encoding GH18 and Hce2 domains (Figure S7). Notably, this locus is closely linked to genes encoding additional LysM-domain proteins in some species (Figure S7). The variation observed in the LysM gene family suggests that recognition of chitin appears to be highly dynamic based on these differences in gene content and domain composition.

Discussion

In this study, we selected diverse *T. rubrum* isolates for genome sequencing, assembly, and analysis and surveyed a wider population sample using MLST analysis. These isolates include multiple morphotypes, which show noted phenotypic variation yet are assigned to the same species based on phylogenetic analyses (Gräser *et al.* 2008; G. S. de Hoog *et al.* 2017). The *T. rubrum* morphotype *soudanense* and *T. rubrum* morphotype *megninii* show higher divergence from a closely related subgroup that includes the *kanei*, *raubitschekii*, and *fischerii* morphotypes, as well as most other *T. rubrum* isolates.

Our MLST and whole-genome analyses provide strong support for the idea that *T. rubrum* is highly clonal and may be primarily asexual, or at least infrequently sexually reproducing. Across 135 isolates examined, 134 were from a single mating type (*MAT1-1*). Consistent with prior reports (Gräser *et al.* 2008; G. S. de Hoog *et al.* 2017), only the *T. rubrum* morphotype *megninii* isolates contain the opposite mating type (*MAT1-2*), while all other *T. rubrum* isolates are of *MAT1-1* type. Direct tests of mating between these and other species did not find evidence for mating and sexual development. While mating was not detected, studies in other fungi have required specialized conditions and long periods of time to detect sexual reproduction (O'Gorman *et al.* 2009). As genes involved in mating and meiosis are conserved in *T. rubrum* (Martinez *et al.* 2012), gene loss does not provide a simple explanation for the inability to mate. Sexual reproduction might occur rarely under specific conditions such as specific temperatures as found for *T. onychocola* (Hubka *et al.* 2015), may be geographically restricted, as the opposite mating type *megninii* morphotype is generally found in the Mediterranean (Sequeira *et al.* 1991), or could be unisexual as in some other fungi such as *C. neoformans* (Lin *et al.* 2005).

As MLST data provided no resolution of the substructure of the *T. rubrum* population, we examined whole-genome sequences for eight diverse isolates. Analysis of the sequence read depth revealed that while some small regions of the genome show amplification or loss, there is no evidence for aneuploidy of entire chromosomes. Most of these *T. rubrum* isolates contain an average of only 3930 SNPs (0.01% of the genome) and phylogenetic

analysis revealed little genetic substructure. Two isolates were more divergent with an average of 24,740 SNPs (0.06% of the genome); one of these was of the recently proposed separated species *T. soudanense* (G. S. de Hoog *et al.* 2017) and the other was the *T. rubrum* morphotype *megninii* isolate. While the similar level of divergence raises the question of whether morphotype *megninii* isolates could also be a separate species, this has not yet been proposed when considering additional phenotypic data in addition to molecular data; however, further study would help clarify species assignments. The low level of variation is remarkable in comparison to other fungal pathogens; for example, while *T. rubrum* isolates are identical at 99.97% of positions on average, isolates of *C. neoformans* var. *grubii* isolates are 99.36% identical on average (Desjardins *et al.* 2017; Rhodes *et al.* 2017). Global populations of *Saccharomyces* have even higher reported diversity (Liti *et al.* 2009). The low diversity and the dependency on the human host for growth suggests that *T. rubrum* may have a low effective population size impacted by the reduction of intraspecies variation by genetic drift. In addition, direct tests for recombination found a low level of candidate reassortment that was not in excess of the estimated number of homoplasmic mutations; further, as there was no apparent decay of linkage disequilibrium over genetic distance, our analyses support the overall clonal nature of this species. The high clonality observed in *T. rubrum* is also supported by MLST analysis of eight microsatellite markers in ~230 *T. rubrum* isolates, including morphotypes from diverse geographic origins (Gräser *et al.* 2007). With additional genome sequencing, geographic substructure may become more apparent; the fungal pathogen *Talaromyces marneffei* also displays high clonality yet isolates from the same country or region are more closely related (Henk *et al.* 2012). While low levels of diversity seem surprising in a common pathogen, this is similar to findings in some bacterial pathogens including *Mycobacterium tuberculosis* and *My. leprae* (Monot *et al.* 2005; Comas *et al.* 2013), which also display high clonality despite phenotypic variation.

LysM-domain proteins are involved in dampening host recognition of fungal chitin (de Jonge *et al.* 2010) and can also regulate fungal growth and development (Seidl-Seiboth *et al.* 2013), yet their specific function in dermatophytes and closely related fungi is not well understood. We also observed variation in genes containing the LysM domain across the sequenced isolates, both in the gene number and domain organization. LysM genes have higher copy numbers in dermatophytes than related fungi in the Ascomycete order Onygenales (Martinez *et al.* 2012). Recent sequencing of additional nonpathogenic species in this order related to *Coccidioides* revealed that most LysM copies found in dermatophytes have a homolog (Whiston and Taylor 2015). Although this analysis excluded *M. canis*—the dermatophyte species with the highest LysM count—this suggests that dermatophytes have retained rather than recently duplicated many of their LysM genes. However, changes in the domain composition of both genes with catalytic domains and those with only LysM domains, many of which represent candidate effectors, highlight the dynamic

evolution of the LysM family in the dermatophytes. Studies of LysM genes in dermatophytes are needed to determine whether these genes serve similar or different roles in these species.

T. rubrum is only found as a pathogen of humans, though this adaptation is more recent relative to the related species that infect other animals or grow in the environment. Unlike the obligate human fungal pathogen *Pneumocystis jirovecii* (Cissé *et al.* 2012; Ma *et al.* 2016), *T. rubrum* does not display widespread gene loss (Martinez *et al.* 2012) indicative of host dependency for growth; further, its genome size is also comparable to related dermatophyte species, supporting no overall reduction (Martinez *et al.* 2012). The presence of a single mating type in the vast majority of isolates and the limited evidence of recombination suggests that sexual reproduction of *T. rubrum* may have been recently lost or may be rarely occurring in specific conditions or geographic regions. This may be linked to the specialization as a human pathogen, as mating may be optimized during environmental growth in the soil (Gräser *et al.* 2008).

Literature Cited

Abyzov, A., A. E. Urban, M. Snyder, and M. Gerstein, 2011 CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21: 974–984. https://doi.org/10.1101/gr.114876.110

Achterman, R. R., and T. C. White, 2013 Dermatophytes. Curr. Biol. 23: R551–R552. https://doi.org/10.1016/j.cub.2013.03.026

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation

of protein database search programs. Nucleic Acids Res. 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Anzawa, K., M. Kawasaki, T. Mochizuki, and H. Ishizaki, 2010 Successful mating of *Trichophyton rubrum* with *Arthroderma simii*. Med. Mycol. 48: 629–634.

Birney, E., M. Clamp, and R. Durbin, 2004 GeneWise and genomewise. Genome Res. 14: 988–995. https://doi.org/10.1101/gr.1865504

Borodovsky, M., A. Lomsadze, N. Ivanov, and R. Mills, 2003 Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr. Protoc. Bioinformatics.* Chapter 4: Unit 4.6.1–Unit 4.6.10.

Buist, G., A. Steen, J. Kok, and O. P. Kuipers, 2008 LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol. Microbiol. 68: 838–847. https://doi.org/10.1111/j.1365-2958.2008.06211.x

Burmester, A., E. Shelest, G. Glockner, C. Heddergott, S. Schindler *et al.*, 2011 Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biol. 12: R7. https://doi.org/10.1186/gb-2011-12-1-r7

Cervelatti, E. P., A. L. Fachin, M. S. Ferreira-Nozawa, and N. M. Martinez-Rossi, 2006 Molecular cloning and characterization of a novel ABC transporter gene in the human pathogen *Trichophyton rubrum*. Med. Mycol. 44: 141–147. https://doi.org/10.1080/13693780500220449

Cissé, O. H., M. Pagni, and P. M. Hauser, 2012 De novo assembly of the *Pneumocystis jirovecii* genome from a single bronchoalveolar lavage fluid specimen from a patient. MBio 4: e00428–e00412. https://doi.org/10.1128/mBio.00428-12

Comas, I., M. Coscolla, T. Luo, S. Borrell, K. E. Holt *et al.*, 2013 Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat. Genet. 45: 1176–1182. https://doi.org/10.1038/ng.2744

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davies, G., and B. Henrissat, 1995 Structures and mechanisms of glycosyl hydrolases. Structure 3: 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9

de Hoog, G. S., K. Dukik, M. Monod, A. Packeu, D. Stubbe *et al.*, 2017 Toward a novel multilocus phylogenetic taxonomy for the dermatophytes. Mycopathologia 182: 5–31. https://doi.org/10.1007/s11046-016-0073-9

de Hoog, S., M. Monod, T. Dawson, T. Boekhout, P. Mayser *et al.*, 2017 Skin fungi from colonization to Infection. Microbiol. Spectr. DOI: 10.1128/microbiolspec.FUNK-0049-2016.

Desjardins, C. A., C. Giamberardino, S. M. Sykes, C.-H. Yu, J. L. Tenor *et al.*, 2017 Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res. 27: 1207–1219. https://doi.org/10.1101/gr.218727.116

Eddy, S. R., 2011 Accelerated profile HMM searches. PLoS Comput. Biol. 7: e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797. https://doi.org/10.1093/nar/gkh340

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt *et al.*, 2014 Pfam: the protein families database. Nucleic Acids Res. 42: D222–D230. https://doi.org/10.1093/nar/gkt1223

Fisher, S., A. Barry, J. Abreu, B. Minie, J. Nolan *et al.*, 2011 A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol. 12: R1. https://doi.org/10.1186/gb-2011-12-1-r1

Gnerre, S., I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton *et al.*, 2011 High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA 108: 1513–1518. https://doi.org/10.1073/pnas.1017351108

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652. https://doi.org/10.1038/nbt.1883

Gräser, Y., A. F. Kuijpers, W. Presber, and G. S. de Hoog, 2000 Molecular taxonomy of the *Trichophyton rubrum* complex. J. Clin. Microbiol. 38: 3329–3336.

Gräser, Y., J. Fröhlich, W. Presber, and S. de Hoog, 2007 Microsatellite markers reveal geographic population differentiation in *Trichophyton rubrum*. J. Med. Microbiol. 56: 1058–1065. https://doi.org/10.1099/jmm.0.47138-0

Gräser, Y., J. Scott, and R. Summerbell, 2008 The new species concept in dermatophytes-a polyphasic approach. Mycopathologia 166: 239–256. https://doi.org/10.1007/s11046-008-9099-y

Haas, B. J., A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith *et al.*, 2003 Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31: 5654–5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen *et al.*, 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9: R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haas, B. J., Q. Zeng, M. D. Pearson, C. A. Cuomo, and J. R. Wortman, 2011 Approaches to fungal genome annotation. Mycology 2: 118–141.

Hall, T., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41: 95–98.

Heidemann, S., M. Monod, and Y. Gräser, 2010 Signature polymorphisms in the internal transcribed spacer region relevant for the differentiation of zoophilic and anthropophilic strains of *Trichophyton interdigitale* and other species of *T. mentagrophytes sensu lato*. Br. J. Dermatol. 162: 282–295. https://doi.org/10.1111/j.1365-2133.2009.09494.x

Heitman, J., 2010 Evolution of eukaryotic microbial pathogens via covert sexual reproduction. Cell Host Microbe 8: 86–99. https://doi.org/10.1016/j.chom.2010.06.011

Henk, D. A., R. Shahar-Golan, K. R. Devi, K. J. Boyce, N. Zhan *et al.*, 2012 Clonality despite sex: the evolution of host-associated sexual neighborhoods in the pathogenic fungus *Penicillium marneffei*. PLoS Pathog. 8: e1002851. https://doi.org/10.1371/journal.ppat.1002851

Hubka, V., C. V. Nissen, R. H. Jensen, M. C. Arendrup, A. Cmokova *et al.*, 2015 Discovery of a sexual stage in *Trichophyton onychocola*, a presumed geophilic dermatophyte isolated from toenails of patients with a history of *T. rubrum* onychomycosis. Med. Mycol. 53: 798–809. https://doi.org/10.1093/mmy/myv044

de Jonge, R., H. P. van Esse, A. Kombrink, T. Shinya, Y. Desaki *et al.*, 2010 Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science 329: 953–955. https://doi.org/10.1126/science.1190859

Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany *et al.*, 2005 Repbase update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110: 462–467. https://doi.org/10.1159/000084979

Kane, J., I. F. Salkin, I. Weitzman, and C. Smitka, 1981 *Trichophyton raubitschekii*, sp. nov. Mycotaxon 13: 259–266.

Kano, R., M. Isizuka, M. Hiruma, T. Mochizuki, H. Kamata *et al.*, 2013 Mating type gene (*MAT1–1*) in Japanese isolates of *Trichophyton rubrum*. Mycopathologia 175: 171–173. https://doi.org/10.1007/s11046-012-9603-2

Kent, W. J., 2002 BLAT–the BLAST-like alignment tool. Genome Res. 12: 656–664. https://doi.org/10.1101/gr.229202

Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59.https://doi.org/10.1186/1471-2105-5-59

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway *et al.*, 2004 Versatile and open software for comparing large genomes. Genome Biol. 5: R12. https://doi.org/10.1186/gb-2004-5-2-r12

Lagesen, K., P. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes *et al.*, 2007 RNAmmer: consistent and rapid annotation of

ribosomal RNA genes. Nucleic Acids Res. 35: 3100–3108. https://doi.org/10.1093/nar/gkm160

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [q-bio.GN].

Li, L., C. J. Stoeckert, and D. S. Roos, 2003 OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13: 2178–2189. https://doi.org/10.1101/gr.1224503

Li, W., B. Metin, T. C. White, and J. Heitman, 2010 Organization and evolutionary trajectory of the mating type (*MAT*) locus in dermatophyte and dimorphic fungal pathogens. Eukaryot. Cell 9: 46–58. https://doi.org/10.1128/EC.00259-09

Lin, X., C. M. Hull, and J. Heitman, 2005 Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature 434: 1017–1021. https://doi.org/10.1038/nature03448

Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts *et al.*, 2009 Population genomics of domestic and wild yeasts. Nature 458: 337–341. https://doi.org/10.1038/nature07743

Lowe, T. M., and S. R. Eddy, 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964. https://doi.org/10.1093/nar/25.5.0955

Ma, L., Z. Chen, D. W. Huang, G. Kutty, M. Ishihara *et al.*, 2016 Genome analysis of three *Pneumocystis* species reveals adaptation mechanisms to life exclusively in mammalian hosts. Nat. Commun. 7: 10740. https://doi.org/10.1038/ncomms10740

Majoros, W. H., M. Pertea, and S. L. Salzberg, 2004 TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20: 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Martinez, D. A., B. G. Oliver, Y. Gräser, J. M. Goldberg, W. Li *et al.*, 2012 Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. mBio 3: e00259–e00212. https://doi.org/10.1128/mBio.00259-12

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303. https://doi.org/10.1101/gr.107524.110

Metin, B., and J. Heitman, 2017 Sexual reproduction in dermatophytes. Mycopathologia 182: 45–55. https://doi.org/10.1007/s11046-016-0072-x

Monot, M., N. Honoré, T. Garnier, R. Araoz, J.-Y. Coppée *et al.*, 2005 On the origin of leprosy. Science 308: 1040–1042. https://doi.org/10.1126/science/1109759

O'Gorman, C. M., H. T. Fuller, and P. S. Dyer, 2009 Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. Nature 457: 471–474. https://doi.org/10.1038/nature07528

Ren, X., T. Liu, J. Dong, L. Sun, J. Yang *et al.*, 2012 Evaluating de Bruijn graph assemblers on 454 transcriptomic data. PLoS One 7: e51188. https://doi.org/10.1371/journal.pone.0051188

Rhodes, J., C. A. Desjardins, S. M. Sykes, M. A. Beale, M. Vanhove *et al.*, 2017 Tracing genetic exchange and biogeography of *Cryptococcus neoformans* var. *grubii* at the global population level. Genetics 207: 327–346. https://doi.org/10.1534/genetics.117.203836

Rocha, E. P. C., J. M. Smith, L. D. Hurst, M. T. G. Holden, J. E. Cooper *et al.*, 2006 Comparisons of dN/dS are time dependent for closely related bacterial genomes. J. Theor. Biol. 239: 226–235. https://doi.org/10.1016/j.jtbi.2005.08.037

Seidl-Seiboth, V., S. Zach, A. Frischmann, O. Spadiut, C. Dietzsch *et al.*, 2013 Spore germination of *Trichoderma atroviride* is inhibited by its LysM protein TAL6. FEBS J. 280: 1226–1236. https://doi.org/10.1111/febs.12113

Sequeira, H., J. Cabrita, C. De Vroey, and C. Wuytack-Raes, 1991 Contribution to our knowledge of *Trichophyton megninii*. J. Med. Vet. Mycol. 29: 417–418.

Stamatakis, A., 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Stanke, M., R. Steinkamp, S. Waack, and B. Morgenstern, 2004 AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32: W309–W312. https://doi.org/10.1093/nar/gkh379

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100: 9440–9445. https://doi.org/10.1073/pnas.1530509100

Symoens, F., O. Jousson, A. Packeu, M. Fratti, P. Staib *et al.*, 2013 The dermatophyte species *Arthroderma benhamiae*: intraspecies variability and mating behaviour. J. Med. Microbiol. 62: 377–385. https://doi.org/10.1099/jmm.0.053223-0

Turin, L., F. Riva, G. Galbiati, and T. Cainelli, 2000 Fast, simple and highly sensitive double-rounded polymerase chain reaction assay to detect medically relevant fungi in dermatological specimens. Eur. J. Clin. Invest. 30: 511–518. https://doi.org/10.1046/j.1365-2362.2000.00659.x

Weitzman, I., and M. Silva-Hutner, 1967 Non-keratinous agar media as substrates for the ascigerous state in certain members of the *Gymnoascaceae* pathogenic for man and animals. Sabouraudia 5: 335–340. https://doi.org/10.1080/00362176785190611

Whiston, E., and J. W. Taylor, 2015 Comparative phylogenomics of pathogenic and nonpathogenic species. G3 (Bethesda) 6: 235–244. https://doi.org/10.1534/g3.115.022806

White, T. C., B. G. Oliver, Y. Gräser, and M. R. Henn, 2008 Generating and testing molecular hypotheses in the dermatophytes. Eukaryot. Cell 7: 1238–1245. https://doi.org/10.1128/EC.00100-08

White, T. C., K. Findley, T. L. Dawson, A. Scheynius, T. Boekhout *et al.*, 2014 Fungi on the skin: dermatophytes and *Malassezia*. Cold Spring Harb. Perspect. Med. 4: a019802.

Wu, C. H., R. Apweiler, A. Bairoch, D. A. Natale, W. C. Barker *et al.*, 2006 The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res. 34: D187–D191. https://doi.org/10.1093/nar/gkj161

Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586–1591. https://doi.org/10.1093/molbev/msm088

Young, C. N., 1968 Pseudo-Cleistothecia in *Trichophyton rubrum*. Sabouraudia J. Med. Vet. Mycol. 6: 160–162. https://doi.org/10.1080/00362176885190281

Communicating editor: A. Mitchell