Genetics: Early Online, published on February 21, 2018 as 10.1534/genetics.117.300573

Whole genome analysis illustrates global clonal population structure of the ubiquitous dermatophyte pathogen *Trichophyton rubrum*

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

[*]Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Brazil

[†]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA

[‡]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, USA

[§]Department of Pharmaceutical Microbiology, Faculty of Pharmacy, University of Mersin, Mersin, Turkey

[**]Center for Infectious Disease Research, Seattle, Washington, USA

[††]University of Leeds, Leeds, United Kingdom

[‡‡]Department of Food Engineering, Faculty of Engineering and Natural Sciences, Istanbul Sabahattin Zaim University, Istanbul, Turkey

[§§]Department of Microbiology, Faculty of Medicine, University of Ege, Izmir, Turkey

[***]Division of Mycology, Department of Microbiology, Faculty of Medicine University of Çukurova, Adana, Turkey

[†††]Institute of Microbiology and Hygiene, University Medicine Berlin - Charité, Berlin, Germany

[‡‡‡]School of Biological Sciences, University of Missouri-Kansas City, Kansas City, Missouri, USA

[§§§§]equal contribution as first authors. [****]corresponding authors.

Current addresses: [1]Veritas Genetics, Danvers, Massachusetts 01923, USA; [2]National Center for Biotechnology Information (NCBI), 8600 Rockville Pike, Bethesda, MD 20894, USA; [3]Stowers Institute for Medical Research, Kansas City, Missouri, USA; [4]Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA; [5]LabCorp, Westborough, MA 01581, USA.

Data access: Genome sequence data is available in NCBI under the Umbrella BioProject PRJNA186851.

Running title: Global clonal population structure of *Trichophyton rubrum*

Keywords: *Trichophyton rubrum*, *Trichophyton interdigitale*, dermatophyte, genome sequence, MLST, mating, recombination, LysM

Corresponding authors:
Christina A. Cuomo, Broad Institute, 415 Main Street, Cambridge, MA 02142. Phone: (617) 714-7904. Email: cuomo@broadinstitute.org

Joseph Heitman, 322 CARL Building, Box 3546, Duke University Medical Center, Durham, N.C. 27710. Phone: (919) 684-2824. Email: heitm001@duke.edu

Abstract

Dermatophytes include fungal species that infect humans, as well as those which 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, the population structure and ability to undergo sexual reproduction are not well understood. In this study, we analyze a large set of *T. rubrum* and *Trichophyton 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.

Introduction

Dermatophyte species are the most common fungal species causing skin infections. Of the more than 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 *Trichophyton tonsurans* and *Microsporum canis*. Some dermatophyte species only cause human infections, including *T. rubrum*, *T. tonsurans*, and *T. interdigitale*. Other species, including *Trichophyton benhamiae, Trichophyton equinum*, *Trichophyton verrucosum*, and *M. canis*, infect mainly animals and occasionally humans, while others such as *Microsporum gypseum* (*Nannizzia gypsea* (de Hoog *et al.* 2017a)) 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 (de Hoog *et al.* 2017b). The species within these genera are closely related phylogenetically and

are within the Ascomycete order Onygenales, family Arthrodermatacaea (White *et al.* 2008; de Hoog *et al.* 2017a).

The *Trichophyton 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; de Hoog *et al.* 2017a). Some morphotypes display phenotypic variation, though these differences can be modest. For example, *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.*). *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 ITS rDNA locus; additionally, some of the morphotypes do not appear to be monophyletic (Gräser *et al.* 2000, 2007; de Hoog *et al.* 2017a), complicating any simple designation of all types as separate species. Combining morphological and multilocus sequence typing (MLST) data has helped 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 alpha box domain or 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 *Trichophyton 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*, *megninii* and *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 **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 PDA medium (Difco) for

10 days at 25°C. 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 65ºC for 30 minutes, followed by addition of 150 µl Epicentre MPC Protein Precipitation Solution. After vortexing, the mixture was centrifuged for 10 minutes, 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 417.52 were also used in mating assays. DNA extraction was performed according to the protocol described by Turin et al. (Turin *et al.* 2000). These isolates were typed by ITS sequence analysis. rDNA sequences spanning the internal transcribed spacer (ITS) 1 region were PCR-amplified using the universal fungal primers ITS1 (5'-TCCGTAGGTGAACCTGCGG3') 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 in GenBank.

Multilocus sequence typing (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 alpha-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°C, followed by 35 cycles of denaturation for 10 sec at 98°C, an annealing time of 15 sec at 54°C, and an extension cycle for 1 min at 72°C. The amplification was completed with an extension period of 5 min at 72°C. 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 region was amplified using the ITS5 (5'-GAAGTAAAAGTCGTAACAAGG-3') and Mas266 (5'-GCATTCCCAAACAACTCGACTC-3') primers with initial denaturation at 94°C for 4 minutes, 35 cycles of denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds, extension at 72°C for 1 minute, and final extension at 72°C for 10 minutes. The reactions were carried out using a BioRad C1000 Touch thermocycler. 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 30000 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 alpha 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°C for 4 minutes, 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, extension at 72°C for 1 minute, with a final extension at 72°C for 7 minutes. For isolates from Turkey, PCR amplifications were performed with the same primers using a Biorad C1000 Touch™ Thermal Cycler, and slightly modified conditions were

used for amplification: initial denaturation at 94°C for 5 minutes, 35 cycles of denaturation at 95°C for 45 seconds, annealing at 55°C for 1.5 minutes, and extension at 72°C for 1 minute, and final extension at 72°C for 10 minutes. The presence of the alpha 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. *Trichophyton 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/L oatmeal agar (Difco), 1 g/L $MgSO_4.7H_2O$, 1 g/L $NaOH_3$, 1g/L KH_2PO_4, and 16 g/L agar (Weitzman and Silva-Hutner 1967)) and Takashio medium (1/10 Sabouraud containing 0.1% neopeptone, 0.2% dextrose, 0.1% $MgSO_4.7H_2O$, and 0.1%. KH_2PO_4). *MAT1-1* and *MAT1-2* isolates grown on Sabouraud Dextrose Agar (SDA) for one 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 each on the Illumina HiSeq 2000 platform. Roughly 100X of 101 base-paired Illumina reads were assembled using ALLPATHS-LG (Gnerre *et al.* 2011) run with 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-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, 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 UTR.

Gene model predictions in repetitive elements were identified and removed from gene sets if they overlapped TPSI 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 at least 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 or predicted transfer or ribosomal RNAs, contained exons spanning sequence gaps, had incomplete codons, or with 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 in NCBI (**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 118829 reference assembly using BWA-MEM (Li 2013); reads from the 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 p. 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 re-called 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 *C. 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 synonymous or non-synonymous change in coding region. The total number of synonymous and non-synonymous sites across the *T. rubrum* CBS 118829 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 non-synonymous 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 1kb. 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 1,000 bootstrap replicates and model GTRCAT. RAxML version 7.7.8 was used for phylogenetic analysis of SNP variant 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 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* dataset 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 1kb 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 is available in NCBI 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 intergenic region (IR), and an alpha-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 (**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, sequence at a subset of the loci was 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 6-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 differred 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 ten *T. rubrum* and two *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 8,616 and 9,064 predicted genes in the ten *T. rubrum* isolates, and 7,993 and 8,116 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 5,236 single-copy orthologs present in all species and estimated a phylogeny with RAxML (Stamatakis 2006) (**Figure 2A**). This phylogeny more precisely delineates the species 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 *fishcheri* 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 118829 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 re-establishment of *soudanense* isolates as a separate species (de Hoog *et al.* 2017a).

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; de Hoog *et al.* 2017a). For the two genomes of these species that we sequenced, MR816 was identical to *T. interdigitale* at the ITS1 locus, wheras 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 alpha 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 3**, **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 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 dataset (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 if 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 (approximately 21 to 22°C) without Parafilm in the dark. Although the assay plates were incubated for longer than five months, ascomata or ascomatal initials 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 8,092 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 *raubitscheckii*) have an average of 3,930 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 *Cryptococcus 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 approximately 24-fold higher than that in *T. rubrum* (**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 intergenic regions for both species, representing 76% and 81% of total variants respectively (**Table 1**, **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 over-represented 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 mating type locus was found in the diverged *soudandense* 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 *Trichophyton 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 (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 8 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 *raubitscheckii*.

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 7,618 ortholog groups, yet there are also substantial differences in gene content. A total of 1,253 orthologs 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 recent dynamics of the LysM family, which binds 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 to 18 copies of 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; de Hoog *et al.* 2017a). 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 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*). Only the *T. rubrum* type *megninii* isolate, Consistent with prior reports (Gräser *et al.* 2008; de Hoog *et al.* 2017a p. 201), only the *T. rubrum* morphotype *megninii* isolates contains the opposite mating type (*MAT1-2*) while all other *T. rubrum* isolates that 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 *Trichophyton 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 *Cryptococcus 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 8 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 3,930 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* (de Hoog *et al.* 2017a), 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

633 separate species, this has not yet been proposed when considering additional
634 phenotypic data in addition to molecular data, however further study would help clarify
635 species assignments. The low level of variation is remarkable in comparison to other
636 fungal pathogens; for example, while *T. rubrum* isolates are identical at 99.97% of
637 positions on average, isolates of *Cryptococcus neoformans* var. *grubii* isolates are
638 99.36% identical on average (Desjardins *et al.* 2017; Rhodes *et al.* 2017). Global
639 populations of *Saccharomyces* have even higher reported diversity (Liti *et al.* 2009). The
640 low diversity and the dependency on the human host for growth suggests that *T. rubrum*
641 may have a low effective population size impacted by the reduction of intra-species
642 variation by genetic drift. In addition, direct tests for recombination found a low level of
643 candidate reassortments that was not in excess of the estimated number of
644 homoplasmic mutations; further, as there was no apparent decay of linkage
645 disequilibrium over genetic distance, our analyses support the overall clonal nature of
646 this species. The high clonality observed in *T. rubrum* is also supported by MLST
647 analysis of eight microsatellite markers in approximately 230 *T. rubrum* isolates,
648 including morphotypes from diverse geographic origins (Gräser *et al.* 2007). With
649 additional genome sequencing geographic substructure may become more apparent;
650 the fungal pathogen *Talaromyces marneffei* also displays high clonality yet isolates from
651 the same country or region are more closely related (Henk *et al.* 2012). While low
652 levels of diversity seems surprising in a common pathogen, this is similar to findings in
653 some bacterial pathogens including *Mycobacterium tubercul*osis and *Mycobacterium*
654 *leprae* (Monot *et al.* 2005; Comas *et al.* 2013), which also display high clonality despite
655 phenotypic variation.
656
657 LysM domain proteins are involved in dampening host recognition of fungal chitin (de
658 Jonge *et al.* 2010) and can also regulate fungal growth and development (Seidl-Seiboth
659 *et al.* 2013), yet their specific function in dermatophytes and closely related fungi is not
660 well understood. We also observed variation in genes containing the LysM domain
661 across the sequenced isolates, both in the gene number and domain organization.
662 LysM genes are present in higher copy number in dermatophytes than related fungi in
663 the Ascomycete order Onygenales (Martinez *et al.* 2012). Recent sequencing of

additional non-pathogenic 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, highlights 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 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).

Acknowledgements

We thank the Broad Institute Genomics Platform for generating the DNA sequence described here. We thank Yonathan Lewit for technical assistance and Cecelia Wall for providing helpful comments on the manuscript. Financial support was provided by the National Human Genome Research Institute grant number U54HG003067 to the Broad Institute and by NIH/NIAID R37 MERIT Award AI39115-20 and RO1 Award AI50113-13 to JH. This study was supported by The Scientific and Technological Research Council of Turkey-2219 Research Fellowship Programme for International Researchers Project

No. [1059B191501539] to AD and Brazilian funding agency FAPESP Fundação de Amparo à Pesquisa do Estado de São Paulo, Postdoctoral Fellowship 12/22232-8 and 13/19195-6 to GFP.

Author contributions

C.A.C., D.A.M., T.C.W. and J.H. conceived and designed the project. A.D., B.M, S.H., M. I., R.B., B.O., Y.G, N.M.M, and T.W. provided the isolates. W.L. and A.D. performed the laboratory experiments. G.F.P, D.A.M, W.L, A.D., R.B.B, A.A, J.M., G,T.S., S.Y, Q.Z, and C.A.C analyzed the data. C.A.C. and J.H. wrote the paper with input from all authors. C.A.C. and J.H. supervised and coordinated the project.

References

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

Achterman R. R., White T. C., 2013 Dermatophytes. Curr. Biol. 23: R551-552.

Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–402.

Anzawa K., Kawasaki M., Mochizuki T., Ishizaki H., 2010 Successful mating of *Trichophyton rubrum* with *Arthroderma simii*. Med. Mycol. Off. Publ. Int. Soc. Hum. Anim. Mycol. 48: 629–34.

Birney E., Clamp M., Durbin R., 2004 GeneWise and Genomewise. Genome Res 14: 988–95.

Borodovsky M., Lomsadze A., Ivanov N., Mills R., 2003 Eukaryotic gene prediction using GeneMark.hmm. Curr Protoc Bioinforma. Chapter 4: Unit4 6.

Buist G., Steen A., Kok J., Kuipers O. P., 2008 LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol. Microbiol. 68: 838–847.

Burmester A., Shelest E., Glockner G., Heddergott C., Schindler S., *et al.*, 2011 Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biol. 12: R7.

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

Cissé O. H., Pagni M., Hauser P. M., 2012 De novo assembly of the *Pneumocystis jirovecii* genome from a single bronchoalveolar lavage fluid specimen from a patient. mBio 4: e00428-00412.

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

Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., *et al.*, 2011 The variant call format and VCFtools. Bioinforma. Oxf. Engl. 27: 2156–2158.

Davies G., Henrissat B., 1995 Structures and mechanisms of glycosyl hydrolases. Struct. Lond. Engl. 1993 3: 853–859.

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

Eddy S. R., 2011 Accelerated Profile HMM Searches. PLoS Comput. Biol. 7: e1002195.

Edgar R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–7.

747 Finn R. D., Bateman A., Clements J., Coggill P., Eberhardt R. Y., *et al.*, 2014 Pfam: the
748 protein families database. Nucleic Acids Res. 42: D222-230.

749 Fisher S., Barry A., Abreu J., Minie B., Nolan J., *et al.*, 2011 A scalable, fully automated
750 process for construction of sequence-ready human exome targeted capture
751 libraries. Genome Biol. 12: R1.

752 Gnerre S., Maccallum I., Przybylski D., Ribeiro F. J., Burton J. N., *et al.*, 2011 High-
753 quality draft assemblies of mammalian genomes from massively parallel
754 sequence data. Proc Natl Acad Sci U A 108: 1513–8.

755 Grabherr M. G., Haas B. J., Yassour M., Levin J. Z., Thompson D. A., *et al.*, 2011 Full-
756 length transcriptome assembly from RNA-Seq data without a reference genome.
757 Nat. Biotechnol. 29: 644–652.

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

760 Gräser Y., Fröhlich J., Presber W., Hoog S. de, 2007 Microsatellite markers reveal
761 geographic population differentiation in *Trichophyton rubrum*. J. Med. Microbiol.
762 56: 1058–1065.

763 Gräser Y., Scott J., Summerbell R., 2008 The new species concept in dermatophytes-a
764 polyphasic approach. Mycopathologia 166: 239–56.

765 Haas B. J., Delcher A. L., Mount S. M., Wortman J. R., Smith R. K., *et al.*, 2003
766 Improving the Arabidopsis genome annotation using maximal transcript
767 alignment assemblies. Nucleic Acids Res 31: 5654–66.

768 Haas B. J., Salzberg S. L., Zhu W., Pertea M., Allen J. E., *et al.*, 2008 Automated
769 eukaryotic gene structure annotation using EVidenceModeler and the Program to
770 Assemble Spliced Alignments. Genome Biol. 9: R7.

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

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

775 Heidemann S., Monod M., Gräser Y., 2010 Signature polymorphisms in the internal
776 transcribed spacer region relevant for the differentiation of zoophilic and
777 anthropophilic strains of *Trichophyton interdigitale* and other species of
778 *T. mentagrophytes sensu lato*. Br. J. Dermatol. 162: 282–295.

779 Heitman J., 2010 Evolution of eukaryotic microbial pathogens via covert sexual
780 reproduction. Cell Host Microbe 8: 86–99.

781 Henk D. A., Shahar-Golan R., Devi K. R., Boyce K. J., Zhan N., *et al.*, 2012 Clonality
782 despite sex: the evolution of host-associated sexual neighborhoods in the
783 pathogenic fungus Penicillium marneffei. PLoS Pathog. 8: e1002851.

784 Hoog G. S. de, Dukik K., Monod M., Packeu A., Stubbe D., *et al.*, 2017a Toward a novel
785 multilocus phylogenetic taxonomy for the dermatophytes. Mycopathologia 182:
786 5–31.

787 Hoog S. de, Monod M., Dawson T., Boekhout T., Mayser P., *et al.*, 2017b Skin Fungi
788 from Colonization to Infection. Microbiol. Spectr. 5.

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

793 Jonge R. de, Esse H. P. van, Kombrink A., Shinya T., Desaki Y., *et al.*, 2010 Conserved
794 fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science
795 329: 953–5.

796 Jurka J., Kapitonov V. V., Pavlicek A., Klonowski P., Kohany O., *et al.*, 2005 Repbase
797 Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res
798 110: 462–7.

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

801 Kano R., Isizuka M., Hiruma M., Mochizuki T., Kamata H., *et al.*, 2013 Mating type gene
802 (*MAT1-1*) in Japanese isolates of *Trichophyton rubrum*. Mycopathologia 175:
803 171–173.

804 Kent W. J., 2002 BLAT--the BLAST-like alignment tool. Genome Res 12: 656–64.

805 Korf I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59.

806 Kurtz S., Phillippy A., Delcher A. L., Smoot M., Shumway M., *et al.*, 2004 Versatile and
807 open software for comparing large genomes. Genome Biol 5: R12.

808 Lagesen K., Hallin P., Rodland E. A., Staerfeldt H. H., Rognes T., *et al.*, 2007
809 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic
810 Acids Res 35: 3100–8.

811 Li L., Stoeckert C. J., Roos D. S., 2003 OrthoMCL: identification of ortholog groups for
812 eukaryotic genomes. Genome Res 13: 2178–89.

813 Li W., Metin B., White T. C., Heitman J., 2010 Organization and evolutionary trajectory
814 of the mating type (*MAT*) locus in dermatophyte and dimorphic fungal pathogens.
815 Eukaryot Cell 9: 46–58.

816 Li H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-
817 MEM. ArXiv13033997 Q-Bio.

818 Lin X., Hull C. M., Heitman J., 2005 Sexual reproduction between partners of the same
819 mating type in *Cryptococcus neoformans*. Nature 434: 1017–1021.

820 Liti G., Carter D. M., Moses A. M., Warringer J., Parts L., *et al.*, 2009 Population
821 genomics of domestic and wild yeasts. Nature 458: 337–41.

822 Lowe T. M., Eddy S. R., 1997 tRNAscan-SE: a program for improved detection of
823 transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–64.

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

827 Majoros W. H., Pertea M., Salzberg S. L., 2004 TigrScan and GlimmerHMM: two open
828 source ab initio eukaryotic gene-finders. Bioinforma. Oxf. Engl. 20: 2878–2879.

829 Martinez D. A., Oliver B. G., Gräser Y., Goldberg J. M., Li W., *et al.*, 2012 Comparative
830 genome analysis of *Trichophyton rubrum* and related dermatophytes reveals
831 candidate genes involved in infection. mBio 3: e00259-00212.

832 McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., *et al.*, 2010 The
833 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation
834 DNA sequencing data. Genome Res 20: 1297–303.

835 Metin B., Heitman J., 2017 Sexual reproduction in dermatophytes. Mycopathologia 182:
836 45–55.

837 Monot M., Honoré N., Garnier T., Araoz R., Coppée J.-Y., *et al.*, 2005 On the origin of
838 leprosy. Science 308: 1040–1042.

839 O'Gorman C. M., Fuller H. T., Dyer P. S., 2009 Discovery of a sexual cycle in the
840 opportunistic fungal pathogen *Aspergillus fumigatus*. Nature 457: 471–4.

841 Ren X., Liu T., Dong J., Sun L., Yang J., *et al.*, 2012 Evaluating de Bruijn graph
842 assemblers on 454 transcriptomic data. PLoS ONE 7: e51188.

843 Rhodes J., Desjardins C. A., Sykes S. M., Beale M. A., Vanhove M., *et al.*, 2017 Tracing
844 Genetic Exchange and Biogeography of *Cryptococcus neoformans* var. *grubii* at
845 the Global Population Level. Genetics 207: 327–346.

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

849 Seidl-Seiboth V., Zach S., Frischmann A., Spadiut O., Dietzsch C., *et al.*, 2013 Spore
850 germination of *Trichoderma atroviride* is inhibited by its LysM protein TAL6.
851 FEBS J. 280: 1226–1236.

852 Sequeira H., Cabrita J., De Vroey C., Wuytack-Raes C., 1991 Contribution to our
853 knowledge of *Trichophyton megninii*. J. Med. Vet. Mycol. Bi-Mon. Publ. Int. Soc.
854 Hum. Anim. Mycol. 29: 417–418.

855 Stamatakis A., 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
856 with thousands of taxa and mixed models. Bioinformatics 22: 2688–90.

857 Stanke M., Steinkamp R., Waack S., Morgenstern B., 2004 AUGUSTUS: a web server
858 for gene finding in eukaryotes. Nucleic Acids Res 32: W309-12.

859 Storey J. D., Tibshirani R., 2003 Statistical significance for genomewide studies. Proc
860 Natl Acad Sci U A 100: 9440–5.

861 Symoens F., Jousson O., Packeu A., Fratti M., Staib P., *et al.*, 2013 The dermatophyte
862 species *Arthroderma benhamiae*: intraspecies variability and mating behaviour.
863 J. Med. Microbiol. 62: 377–385.

864 Turin L., Riva F., Galbiati G., Cainelli T., 2000 Fast, simple and highly sensitive double-
865 rounded polymerase chain reaction assay to detect medically relevant fungi in
866 dermatological specimens. Eur. J. Clin. Invest. 30: 511–518.

867 Weitzman I., Silva-Hutner M., 1967 Non-keratinous agar media as substrates for the
868 ascigerous state in certain members of the Gymnoascaceae pathogenic for man
869 and animals. Sabouraudia 5: 335–340.

870 Whiston E., Taylor J. W., 2015 Comparative phylogenomics of pathogenic and
871 nonpathogenic species. G3 Bethesda Md 6: 235–244.

872 White T. C., Oliver B. G., Gräser Y., Henn M. R., 2008 Generating and testing molecular
873 hypotheses in the dermatophytes. Eukaryot. Cell 7: 1238–45.

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

876 Wu C. H., Apweiler R., Bairoch A., Natale D. A., Barker W. C., *et al.*, 2006 The
877 Universal Protein Resource (UniProt): an expanding universe of protein
878 information. Nucleic Acids Res. 34: D187–D191.

879 Yang Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol.
880 24: 1586–1591.

881 Young C. N., 1968 Pseudo-Cleistothecia in *Trichophyton rubrum*. Sabouraudia J. Med.
882 Vet. Mycol. 6: 160–162.

883

884

Table

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, non-synonymous SNP sites; pN/pS, (NSY/total NSY sites)/(SYN/total SYN sites).

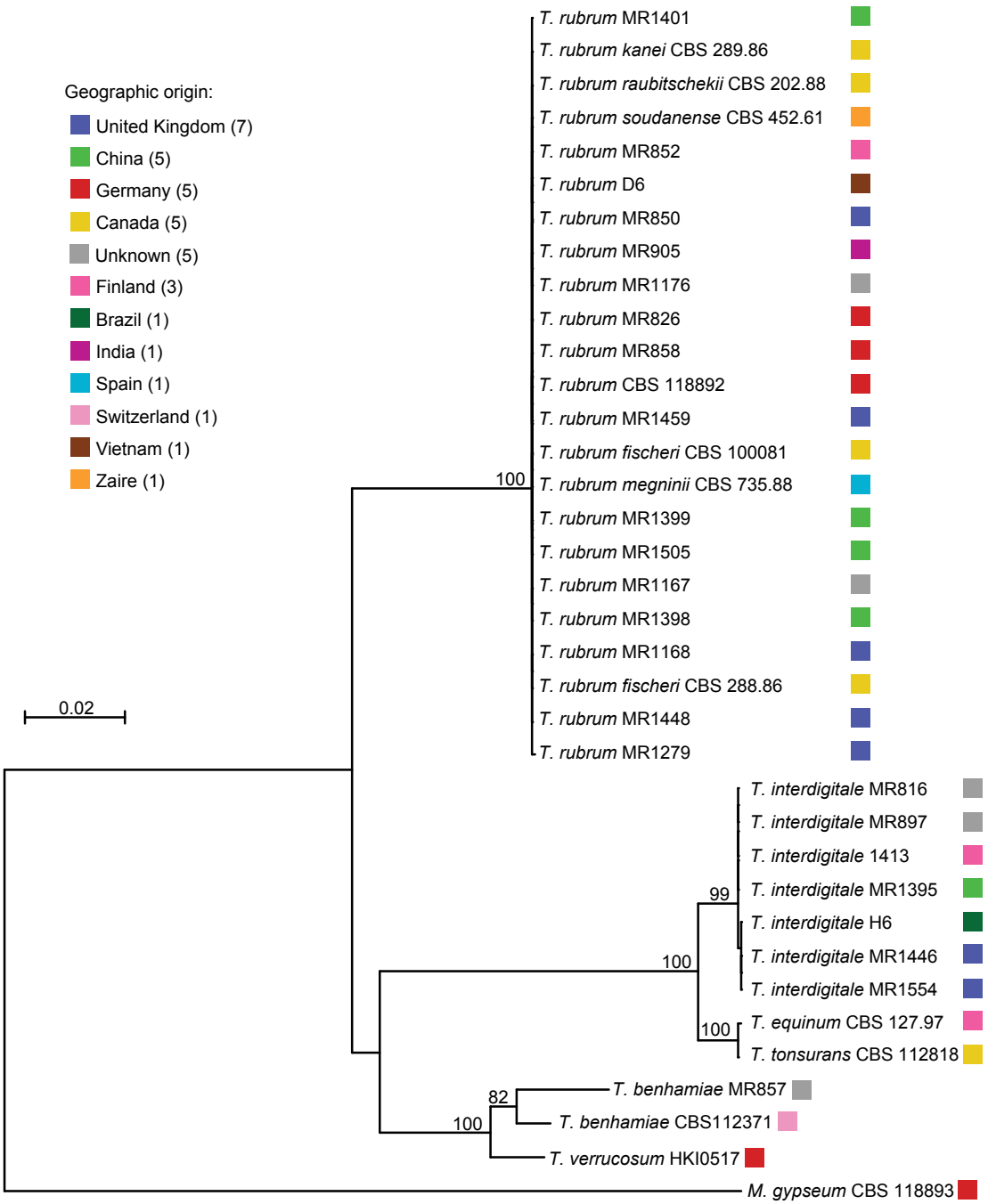
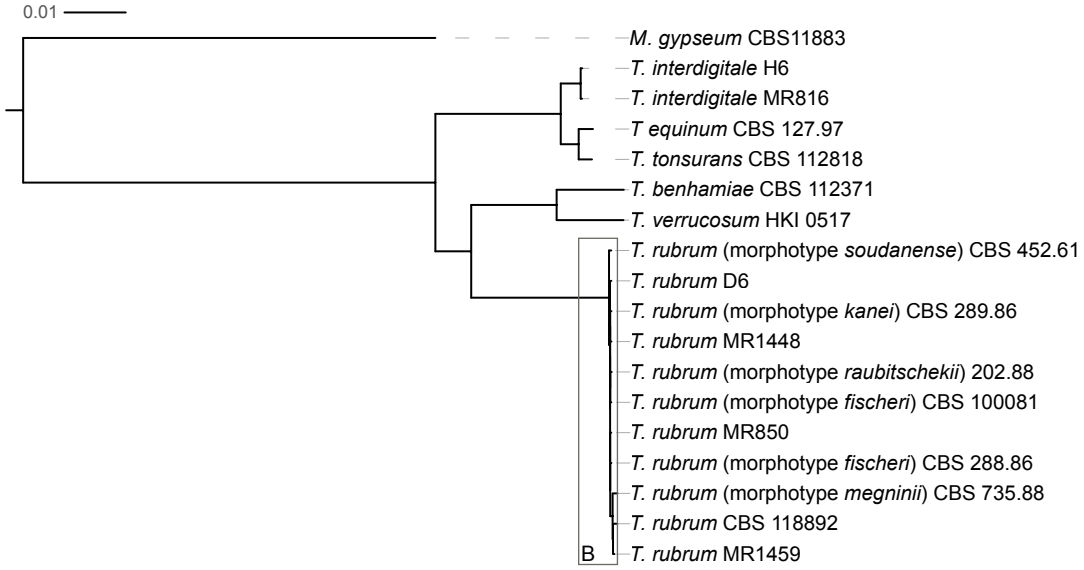


Figure 1. Phylogeny inferred from concatenated MLST sequences. Three MLST loci (ABC transporter, outer membrane protein, and CAP59 protein) were amplified and sequenced from 79 isolates and sequences were identified in an additional 19 assemblies. The concatenated sequence for each isolate was used to build a maximum likelihood tree using MEGA 5.2. Isolate MR1168 is representative of 73 *T. rubrum* isolates that have identical MLST sequences.

A.



B.

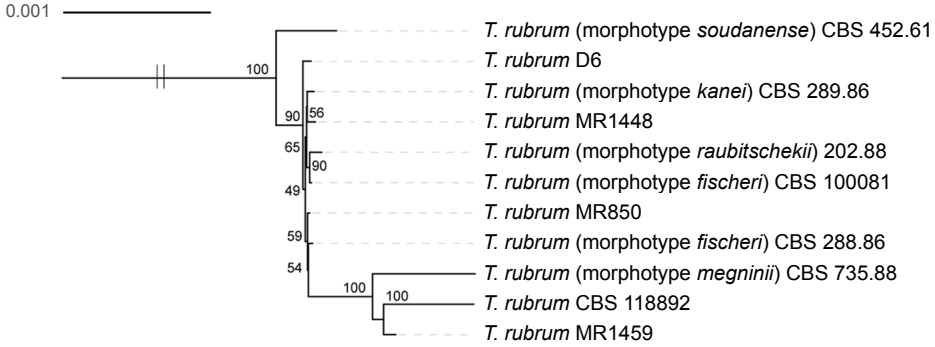


Figure 2. Phylogenetic relationship of *Trichophyton* isolates. A total of 5,236 single copy genes were each aligned with MUSCLE; the concatenated alignment was used to infer a species phylogeny with RAxML (GTRCAT model) with 1,000 bootstrap replicates using either A. all species including the outgroup *M. gypseum* or B. only *Trichophyton rubrum* isolates.

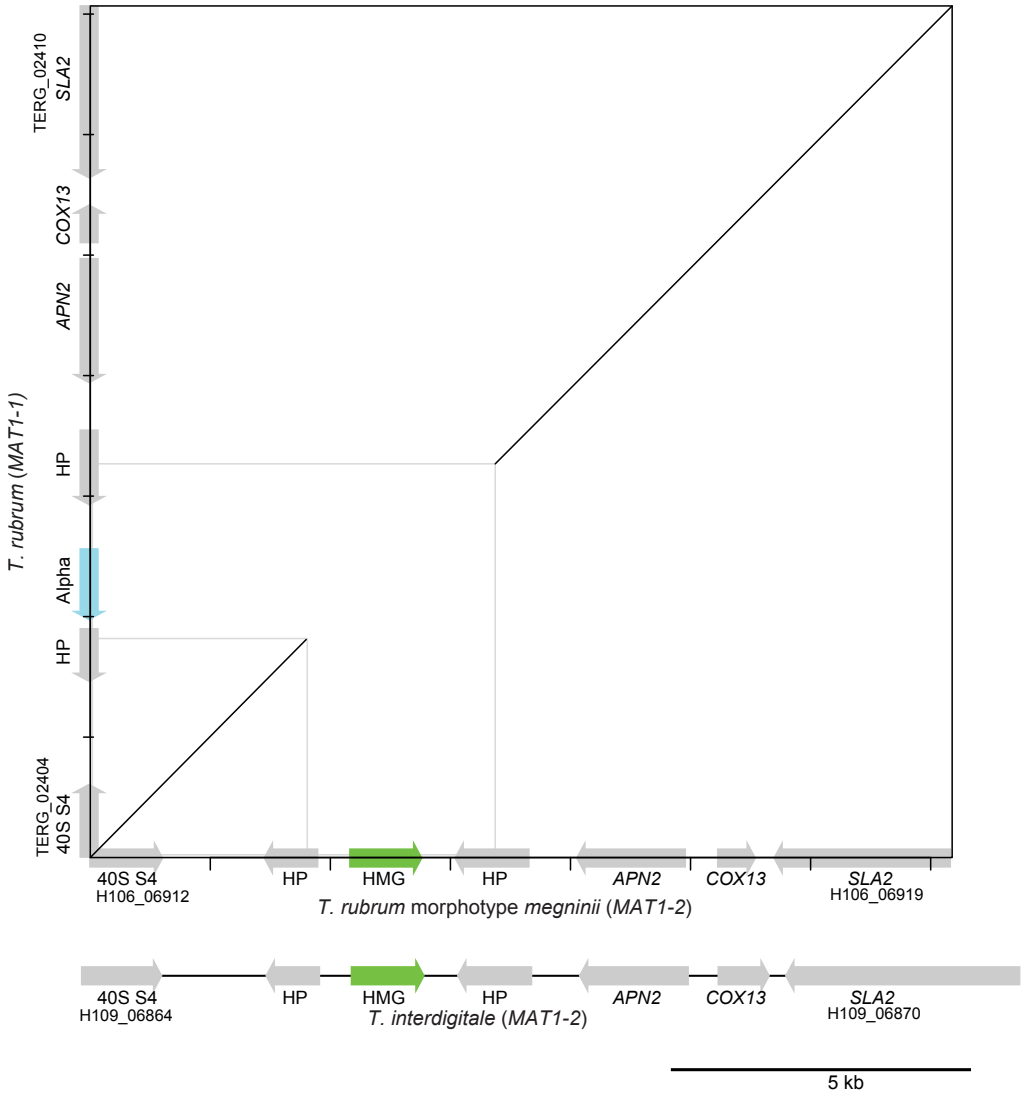


Figure 3. Alignment of the mating type locus of selected isolates. Mating type genes of *T. rubrum* morphotype *megninii* (CBS 735.88) and *T. rubrum* (CBS 188992) are shown along the x- and y- axes, respectively, with regions aligning by NUCmer show in the dotplot. The alignment extends into two hypothetical proteins (HP) immediately flanking the alpha or HMG domain gene that specifies mating type. Most *T. rubrum* (*MAT1-1*) isolates contain an alpha 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).

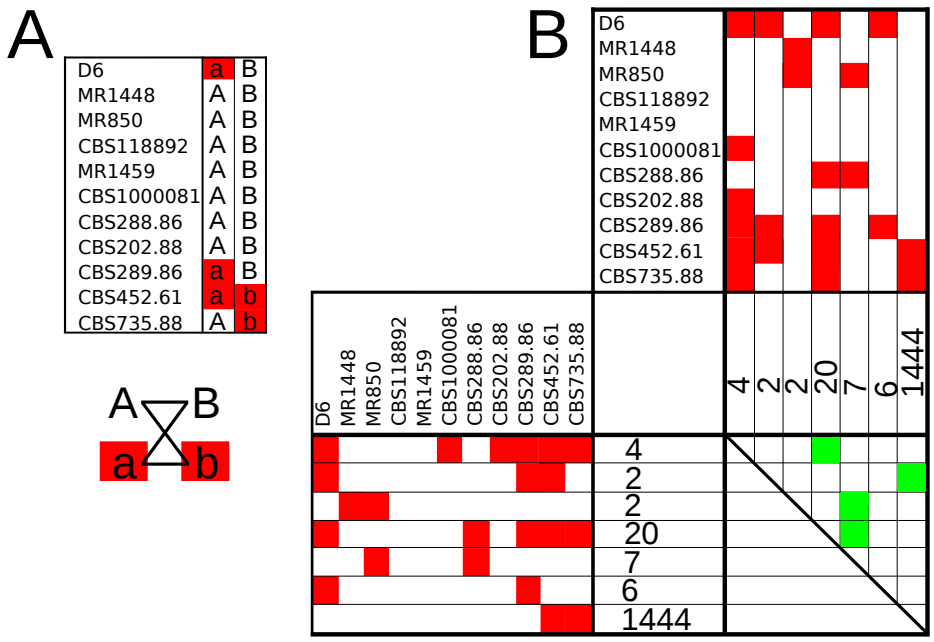


Figure 4. Paired allele compatibility test suggests limited evidence for sexual reproduction. A. A single example of a positive paired allele compatibility test from the *T. rubrum* population. In this test, two loci are examined and typed across the population. To perform a meaningful test, at least two individuals in the population must share a variant allele at each site. Here alternative SNPs are depicted in red and 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 variant 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.

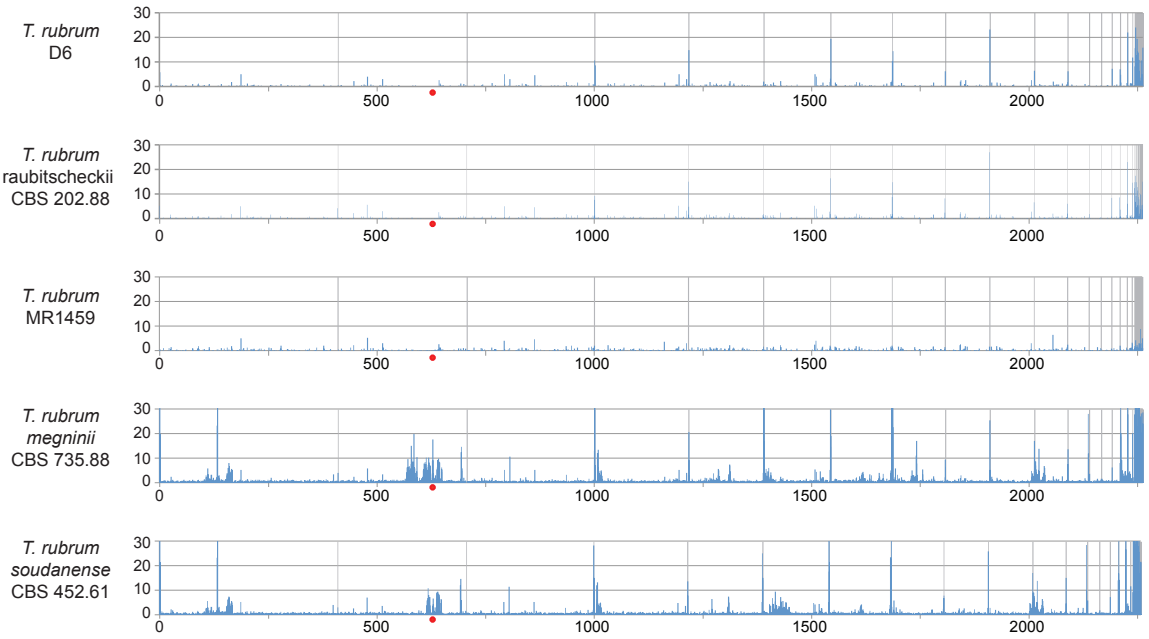


Figure 5. Genome-wide SNP frequency highlights hotspots. For each panel, the frequency of SNPs in 5-kb windows is shown across the genome. The genome assembly of isolate CBS 11892 was used for all comparisons, and scaffolds are ordered along the x-axis with grey lines representing scaffold boundaries. Red dots indicate the position of the mating type locus.

947 **Supplemental Figure and Table Legends**

948

949 **Figure S1. ITS sequence variation in *T. interdigitale*.** Aligned ITS sequence is shown
950 for four isolates, including the two for which whole genomes were sequenced (H6 and
951 MR816) and two previously characterized isolates (AF168124 and AY062119). Isolate
952 AY062119 has been re-classified as *T. mentagrophytes*. Variant sites are highlighted.

953

954 **Figure S2. Detection of *MAT1-1* in *T. rubrum* and *MAT1-2* in *T. interdigitale* by**
955 **PCR.** A. PCR-based determination of the *MAT1-1* alpha domain of *T. rubrum* isolates:
956 *T. rubrum* MR 851, *T. megninii* CBS 389.58, *T. megninii* CBS 384.64, and *T. megninii*
957 CBS 417.52. The alpha domain *MAT1-1* gene was identified from samples = 1-12 and
958 MR 851 *T. rubrum*; the alpha domain *MAT1-1* gene was not identified from *T. megninii*
959 CBS 389.58, *T. megninii* CBS 384.64, or *T. megninii* CBS 417.52. M: DNA ladder. B.
960 PCR-based determination of *MAT 1-2* HMG domain of *T. rubrum*: *T. rubrum* MR851, *T.*
961 *megninii* CBS 389.58, *T. megninii* CBS 384.64, and *T. megninii* CBS 417.52. The HMG
962 domain *MAT1-2* gene was identified from *T. megninii* CBS 389.58, *T. megninii* CBS
963 384.64, and *T. megninii* CBS 417.52*,* and was not identified from *T. rubrum* isolate = 1-
964 12 and *T. rubrum* MR 851, M: DNA ladder*.* C. PCR-based determination of *MAT1-2*
965 HMG domain of *T. interdigitale. T. interdigitale* isolates = 1-9, *T. interdigitale* MR8801
966 (positive control), and *T. rubrum* MR851 (negative control), M; DNA ladder. D. PCR
967 based determination of *MAT 1-1* alpha domain of *T. interdigitale* isolates. *T. interdigitale*
968 isolates = 1-9, *T. interdigitale* MR8801 (negative control), and *T. rubrum* MR851
969 (positive control). M: DNA ladder.

970

971 **Figure S3. Phylogenetic relationship and sharing of variant sites of sequenced *T.***
972 ***rubrum* isolates.** A. Phylogenetic relationship of *T. rubrum* isolates inferred using
973 RAxML (Methods). B. Classification of SNP sites based on conservation across the
974 sequenced isolates; unique: only in one isolate; shared: in two to seven isolates; fixed:
975 in all eight isolates.

976 **Figure S4. Lack of decay of linkage disequilibrium (LD) in *T. rubrum*.** LD (r^2) was
977 calculated for all pairs of SNPs separated by 0 to 300 kb and then averaged for every

1kb. LD values for each window were then calculated by averaging over all pairwise calculations in the window.

Figure S5. Mating assays. A. Mating assay plate; *T. rubrum* and *T. interdigitale* on E medium. B. Mating assay plate*; T. rubrum* and *A. simii* (a) E medium, (b) Takashio medium eight weeks. C*. T. rubrum* and *T. megninii* on E medium for eight weeks.

Figure S6. Read depth of sequenced isolates. Reads from each isolate were aligned to the *T. rubrum* reference genome and normalized read depth was computed for 5kb windows. Read depth is even across the reference genome for most isolates, with small regions of higher depth detected in some isolates.

Figure S7. Variation in LysM-Hce gene cluster across sequenced dermatophytes. In *T. rubrum*, the LysM-Hce gene is closely linked to two other LysM genes; this organization is most similar to that found in *M. canis*, although these genes are located on two different scaffolds.

Table S1. Properties of sequenced isolates.
Table S2. Accessions for sequenced genomes.
Table S3. MLST sequence, genotypes, and GenBank accession numbers.
Table S4. Primers for MLST gene amplification.
Table S5. *Trichophyton* genome assembly statistics.
Table S6. Primer pairs used for mating type determination of *T. rubrum*
Table S7. Mating assays and results.
Table S8. Frequency of SNPs in *T. rubrum* and *T. interdigitale* isolates **by mutation type.**
Table S9. Duplicated regions in sequence isolates.
Table S10. List of genes found in duplicated regions.
Table S11. Deleted regions in sequenced isolates.
Table S12. List of genes in deleted regions in sequenced isolate
Table S13. Genes containing the LysM binding domain in dermatophytes