Phylogeny of dermatophytes with genomic character evaluation of clinically distinct \textit{Trichophyton rubrum} and \textit{T. violaceum}

P. Zhan$^{1,2,3,4}$, K. Dukik$^{3,4}$, D. Li$^{1,5}$, J. Sun$^6$, J.B. Stielow$^{3,8,9}$, B. Gerrits van den Ende$^3$, B. Brankovics$^{3,4}$, S.B.J. Menken$^4$, H. Mei$^7$, W. Bao$^7$, G. Lv$^7$, W. Liu$^7$, and G.S. de Hoog$^{1,2,3,4}$

$^1$Institute of Dermatology, Chinese Academy of Medical Sciences & Peking Union Medical College, Jiangsu Key Laboratory of Molecular Biology for Skin Diseases and STIs, Nanjing, China; $^2$Dermatology Hospital of Jiangsu Province, Nanjing Dermatology Institute, Nanchang, China; $^3$Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands; $^4$Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands; $^5$Georgetown University Medical Center, Department of Microbiology and Immunology, Washington, DC, USA; $^6$Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China; $^7$Nanjing General Hospital of Nanjing Command, Nanjing, China; $^8$Thermo Fisher Scientific, Landsmeer, The Netherlands; $^9$Center of Expertise in Mycology of Radboudum/c/Canisius Wilhelmina Hospital, Nijmegen, The Netherlands

*Correspondence: W. Liu, liumyc0@hotmail.com; G.S. de Hoog, s.hoog@westerdijk institute.nl

Abstract: \textit{Trichophyton rubrum} and \textit{T. violaceum} are prevalent agents of human dermatophyte infections, the former being found on glabrous skin and nail, while the latter is confined to the scalp. The two species are phenotypically different but are highly similar phylogenetically. The taxonomy of dermatophytes is currently being reconsidered on the basis of molecular phylogeny. Molecular species definitions do not always coincide with existing concepts which are guided by ecological and clinical principles. In this article, we aim to bring phylogenetic and ecological data together in an attempt to develop new species concepts for anthropophilic dermatophytes. Focus is on the \textit{T. rubrum} complex with analysis of \textit{rDNA ITS} supplemented with \textit{LSU}, TUB2, TEF3 and ribosomal protein L10 gene sequences. In order to explore genomic differences between \textit{T. rubrum} and \textit{T. violaceum}, one representative for both species was whole genome sequenced. Draft sequences were compared with currently available dermatophyte genomes. Potential virulence factors of adhesions and secreted proteases were predicted and compared phylogenetically. General phylogeny showed clear gaps between geophilic species of \textit{Arthroderma}, but multilocus distances between species were often very small in the derived anthropophilic and zoophilic genus \textit{Trichophyton}. Significant genome conservation between \textit{T. rubrum} and \textit{T. violaceum} was observed, with a high similarity at the nucleic acid level of 99.38\% identity. \textit{Trichophyton violaceum} contains more paralogs than \textit{T. rubrum}. About 30 adhesion genes were predicted among dermatophytes. Seventeen adhesions were common between \textit{T. rubrum} and \textit{T. violaceum}, while four were specific for the former and eight for the latter. Phylogenetic analysis of secreted proteases reveals considerable expansion and conservation among the analyzed species. Multilocus phylogeny and genome comparison of \textit{T. rubrum} and \textit{T. violaceum} underlined their close affinity. The possibility that they represent a single species exhibiting different phenotypes due to different localizations on the human body is discussed.

Key words: Adhesion, Arthrodermataceae, Character analysis, Dermatophytes, Genome, Phylogeny, Protease, \textit{Trichophyton rubrum}, \textit{Trichophyton violaceum}.

Available online 21 February 2018; https://doi.org/10.1016/j.simyco.2018.02.004.

INTRODUCTION

Dermatophytes (Onygenales: Arthrodermataceae) are filamentous fungi that invade and grow in keratin-rich substrates. Many species of this family reside as saprobes in the environment or as commensals in animal fur, but particularly among the anthropophiles there are species that are able to invade hairless human skin and nails and cause infection. About 10 dermatophyte species commonly occur on the human host, and it is estimated that about 20–25\% of the world’s population carries a dermatophyte infection (Ates et al. 2008; Kim et al. 2015). Nearly 80\% of these are caused by \textit{Trichophyton rubrum} and its close relatives (Hlavickova et al. 2008). The genus \textit{Trichophyton} in the modern sense contains 16 species, of which seven are anthropophilic (de Hoog et al. 2017), among which are \textit{T. rubrum} and \textit{T. violaceum}. \textit{Trichophyton rubrum} and \textit{T. violaceum} share some significant ecological traits. Both are anthropophilic, i.e., restricted to the human host, causing chronic, non- or mild-inflammatory infections. Their high degree of molecular similarity is expressed in widely used barcoding genes, e.g. \textit{rDNA ITS} Transcribed Spacer (ITS), translation elongation factor 1 and partial \(\beta\)-tubulin (Gräser et al. 2000; Rezaei-Matehkolaee et al. 2014; Mirhendi et al. 2015). However, the species show significant phenotypic and clinical differences. \textit{Trichophyton rubrum} typically presents as fluffy to woolly, pinkish colonies with moderate growth speed, while \textit{T. violaceum} appears as wrinkled, deep purple colonies with slow growth. Microscopically \textit{T. rubrum} has a profusely branched and richly sporulating conidial system with tear-shaped micro- and cigar-shaped macroconidia. In contrast, hyphae of \textit{T. violaceum} are broad, tortuous and distorted, without sporulation and with chlamydospore-like conidia in older cultures. These morphological differences coincide with marked clinical differences. \textit{Trichophyton rubrum} usually causes infections of glabrous skin leading to tinea corporis, tinea pedis, tinea manuum or onychomycosis (Ates et al. 2008). \textit{Trichophyton violaceum} usually infects hair and adjacent skin of the scalp, leading to black dot tinea capitis (Farina et al. 2015, Gräser et al. 2000; Fig. 1).

The molecular basis of the pathogenicity-associated traits among dermatophytes is currently insufficiently understood to explain the striking differences between the clinical predilections of \textit{Trichophyton rubrum} and \textit{T. violaceum} (Gräser et al. 2000). In 2011, the first genomes of dermatophyte species became available: 97\% of the 22.5 Mb genome of \textit{Trichophyton brevicaule} and \textit{T. verrucosum} were completed and aligned (Burmester et al. 2011). Shortly afterwards five further genomes of important dermatophytes, \textit{T. rubrum}, \textit{T. tonsurans}, \textit{T. equinum},
Microsporum canis and Nannizzia gypsea were added (Martinez et al. 2012). This enabled a comparative analysis of gene families that might be responsible for specific types of pathogenesis, such as proteases, kinases, secondary metabolites and proteins with LysM binding domains. The above species account for the majority of tinea infections; however, the main agent of tinea capitis, T. violaceum was not included. To further understand the genetics of T. rubrum siblings and their divergent pathomechanisms, we sequenced the genomes of two clinical strains T. rubrum CMCC(F)T1i (=CBS 139224) and T. violaceum CMCC(F)T3i (=CBS 141829) from South China using Illumina HiSeq®2000 platform. To obtain optimal quality of the reference sequence, PacBio RS single molecule real-time (SMRT) sequencing was also applied to the strain of T. rubrum, besides the Illumina methods. Very high genome quality was yielded for both isolates. The draft sequences from our strains were compared with each other and with seven dermatophyte genomes available in the public domain, with a focus on proteases and adhesins. The study allowed us to explore genomic polymorphism in dermatophytes and its implications for pathogenesis and adaptation, aiming to lead to better understanding of genome organization and evolution of specific pathogenic traits.

MATERIALS AND METHODS

Strains and culture conditions

Strains preserved in the reference collection of Centraalbureau voor Schimmelcultures (housed at Westerdijk Fungal Biodiversity Institute) were used to construct a multilocus phylogeny of the family Arthrodermataceae. In total, 264 strains were included, of which 261 were from CBS and 3 from BCCM/IHEM Biomedical Fungi and Yeasts Collection, Brussels, Belgium (Table 1). Strains were cultured on Sabouraud’s Glucose Agar (SGA) plates inoculated by lyophilized, cryo-preserved or fresh mycelial material. Most of the cultures were grown for 7–14 d at 24 °C. Two strains used for whole-genome sequencing were grown in Sabouraud’s glucose broth (SGB); see below.

DNA extraction, PCR, and sequencing

DNA extraction for phylogeny was performed using MasterPure™ Yeast DNA Purification Kit from Epicentre, using preserved material or material harvested from living cultures using methods and PCR protocols of Stielow et al. (2015) and de Hoog et al. (2017). In total eight gene regions were amplified: ITS and LSU loci of the rDNA operon, partial β-tubulin II (TUB2), γ-actin (ACT), translation elongation factor 1-α (TEF1), RNA polymerase II (RPB2), 60S ribosomal protein L10 (RP 60S L1) and two primer sets for the fungal-specific translation elongation factor 3 (TEF3). The chosen loci, corresponding primer sets, primer sequences, PCR volumes and PCR reactions are given by Stielow et al. (2015). After visualization of amplicons on 1 % agarose, positive products were sequenced using ABI big dye terminator v. 3.1, with one quarter of its suggested volume (modified manufacturer’s protocol). Bidirectional sequencing was performed using a capillary electrophoresis system (Life Technologies 3730XL DNA analyser). Obtained sequences of CBS strains were manually edited and stored in a BIOLOMICS database at Westerdijk Institute (Vu et al. 2012). Consensus sequences of IHEM strains were edited using SeqMan in the LASERGENE 219 software (DNASTAR, WI, U.S.A.).

Alignment, phylogeny and locus assessment

A subset of 123 strains was firstly tested with nine sets of primers for eight DNA loci. In the second analysis, 141 strains were added
| CBS number | Current taxon name | New taxon name | Status | Source |
|------------|-------------------|----------------|--------|--------|
| CBS 221.75 | A. borellii       | A. borellii    |        | Rat    |
| CBS 967.68 | A. borellii       | A. borellii    | ST     | ST Nannizzia borellii |
| CBS 272.66 | C. georgiae       | A. ciferrii    | T      | Arthroderma ciferrii |
| CBS 492.71 | A. cuniculi       | A. cuniculi    | ST     | Rabbit burrow |
| CBS 272.66 | A. cuniculi       | A. cuniculi    | ST     | Arthroderma ciferrii |
| CBS 353.66 | A. curreyi        | A. ciferrii    | ET     | Soil |
| CBS 117155 | T. eboreum        | A. ciferrii    | T      | Trichophyton eboreum |
| CBS 117155 | T. sp.            | A. ciferrii    | T      | Skin |
| CBS 473.78 | A. flavescent     | A. flavescent  | ST     | Arthroderma flavescent |
| CBS 473.78 | A. flavescent     | A. flavescent  | ST     | Arthroderma flavescent |
| CBS 598.66 | A. gertleri       | A. gertleri    | ST     | Soil |
| CBS 666.77 | A. gertleri       | A. gertleri    | ST     | Kingfisher |
| CBS 228.79 | A. gloriae        | A. gloriae     | T      | Arthroderma gloriae |
| CBS 667.77 | A. gloriae        | A. gloriae     | T      | Soil |
| CBS 667.77 | A. gloriae        | A. gloriae     | T      | Soil |
| CBS 521.71 | A. insingulare    | A. insingulare | T      | Soil |
| CBS 522.71 | A. insingulare    | A. insingulare | T      | Soil |
| CBS 307.65 | A. lenticulare    | A. lenticulare | T      | Arthroderma lenticulare |
| CBS 308.65 | A. lenticulare    | A. lenticulare | T      | Gopher burrow |
| CBS 120.30 | T. tonsurans      | A. melis       | T      | Human |
| CBS 308.65 | A. melis          | A. melis       | T      | Badger burrow |
| CBS 117.61 | A. quadrifidum    | A. quadrifidum | AUT    | Arthroderma quadrifidum |
| CBS 138.26 | A. quadrifidum    | A. quadrifidum | AUT    | Arthroderma quadrifidum |
| CBS 310.65 | A. quadrifidum    | A. quadrifidum | AUT    | Arthroderma quadrifidum |
| CBS 311.65 | A. quadrifidum    | A. quadrifidum | AUT    | Arthroderma quadrifidum |
| CBS 134551 | T. redellii       | A. redellii    | T      | Trichophyton redellii |
| CBS 132920 | T. thuringiense   | A. sp.         | T      | Nail |
| CBS 473.77 | T. thuringiense   | A. thuringiense| T      | Mouse |
| CBS 101515 | A. uncinatum      | A. uncinatum   | T      | Keratinomyces uncinatum |
| CBS 119779 | T. ajelloi var. a. | A. uncinatum   | T      | Nail |
| CBS 128.75 | A. uncinatum      | A. uncinatum   | ST     | E. stockdaleae |
| CBS 179.57 | A. uncinatum      | A. uncinatum   | ST     | Soil |
| CBS 180.57 | A. uncinatum      | A. uncinatum   | ST     | Soil |
| CBS 180.64 | A. uncinatum      | A. uncinatum   | ST     | Keratinomyces ajelloi var. nanum |
| CBS 315.65 | A. uncinatum      | A. uncinatum   | ST     | Arthroderma uncinatum |
| CBS 316.65 | A. uncinatum      | A. uncinatum   | ST     | Arthroderma uncinatum |
| CBS 355.93 | C. vespertilii    | A. vespertilii | T      | Chrysosporium vespertilii |
| CBS 187.61 | Ctenomyces serratus| Ctenomyces serratus | NT    | Ctenomyces serratus |
| CBS 544.63 | Ctenomyces serratus| Ctenomyces serratus | NT    | Ctenomyces serratus |
| CBS 100148 | A. uncinatum      | E. floccosum   | NT     | Epidermophyton floccosum |
| CBS 108.67 | E. floccosum var. | E. floccosum   | NT     | Human |
| CBS 230.76 | E. floccosum var. | E. floccosum   | NT     | Human |

(continued on next page)
| CBS number | Current taxon name | New taxon name | Status | Source |
|------------|-------------------|----------------|--------|--------|
| CBS 240.67 | *E. floccosum* var. *floccosum* | *E. floccosum* | Skin | |
| CBS 457.65 | *E. floccosum* var. *nigricans* | *E. floccosum* | Human | |
| CBS 553.84 | *E. floccosum* var. *floccosum* | *E. floccosum* | Human | |
| CBS 269.89 | *Keratinomyces ceretanicus* | *Guarromyces ceretanicus* | Human | |
| CBS 100083 | A. gruby | L. gallinae | Dog | |
| CBS 243.66 | A. gruby | L. gallinae | T Nannizzia grubyi | |
| CBS 244.66 | A. gruby | L. gallinae | NT Lophopyton gallinae | |
| CBS 300.52 | M. audouinii | M. audouinii | NT Microsporum audouinii | Scalp |
| CBS 495.86 | A. otae | M. audouinii | T Nannizzia otae | |
| CBS 102894 | M. audouinii | M. audouinii | Scalp | |
| CBS 108932 | M. audouinii | M. audouinii | Human | |
| CBS 108933 | M. audouinii | M. audouinii | Human | |
| CBS 108934 | M. audouinii | M. audouinii | Scalp | |
| CBS 119449 | M. audouinii | M. audouinii | Human | |
| CBS 404.61 | M. audouinii | M. audouinii | Human | |
| CBS 101514 | A. otae | M. canis | T Microsporum distortum | Scalp |
| CBS 114329 | M. canis | M. canis | Skin | |
| CBS 130922 | M. canis | M. canis | Skin | |
| CBS 130931 | M. canis | M. canis | Skin | |
| CBS 130932 | M. canis | M. canis | Skin | |
| CBS 130949 | M. canis | M. canis | Scalp | |
| CBS 156.69 | A. otae | M. canis | Human | |
| CBS 191.57 | A. otae | M. canis | Dog | |
| CBS 214.79 | A. otae | M. canis | Rabbit | |
| CBS 217.69 | A. otae | M. canis | Nail | |
| CBS 238.67 | A. otae | M. canis | Human | |
| CBS 274.62 | A. otae | M. canis | Monkey | |
| CBS 281.63 | A. otae | M. canis | Scalp | |
| CBS 283.63 | A. otae | M. canis | | |
| CBS 284.63 | A. otae | M. canis | Gibbon | |
| CBS 445.51 | M. ferrugineum | M. canis | Skin | |
| CBS 482.76 | A. otae | M. canis | Contaminant | |
| CBS 496.86 | A. otae | M. canis | ST Nannizzia otae, NT Microsporum canis | |
| CBS 109478 | M. audouinii | M. canis | Scalp | |
| CBS 317.31 | M. ferrugineum | M. ferrugineum | Human | |
| CBS 373.71 | M. ferrugineum | M. ferrugineum | Human | |
| CBS 449.61 | M. ferrugineum | M. ferrugineum | Human | |
| CBS 452.59 | T. concentricum | M. ferrugineum | Skin | |
| CBS 497.48 | M. ferrugineum | M. ferrugineum | Scalp | |
| CBS 366.81 | A. corniculatum | N. corniculata | ST Nannizzia corniculata | Soil |
| CBS 384.81 | A. corniculatum | N. corniculata | ST Nannizzia corniculata | Soil |
| CBS 349.49 | M. duboisii | N. duboisii | T Sabouraudites duboisii | Skin |
| CBS 599.66 | A. fulvum | N. fulva | T Microsporum boullardi | Soil |
| CBS 130934 | M. fulvum | N. fulva | Human | |
| CBS 130942 | M. fulvum | N. fulva | Human | |
| CBS 146.66 | M. gypseum | N. fulva | AUT Favomicrosporon pinetti | Contaminant |
| CBS number | Current taxon name | New taxon name | Status | Source |
|------------|-------------------|----------------|--------|--------|
| CBS 147.66 | M. gypseum        | N. fulva       | AUT    | Favomicrosporon pinetii | Contaminant |
| CBS 243.64 | A. fulvum         | N. fulva       | T      | Keratinomyces longifusus | Scalp |
| CBS 287.55 | A. fulvum         | N. fulva       | T      | Microsporum fulvum       | Human |
| CBS 385.64 | A. otae           | N. fulva       | Human  |         |         |
| CBS 529.71 | A. fulvum         | N. fulva       | T      | Microsporum ripariae      | Birdnest |
| CBS 120675 | M. gypseum        | N. gypseum     | ST     | Microsporum gypseum var. vinosum | Skin |
| CBS 118893 | M. gypseum        | N. gypseum     | Skin   |         |         |
| CBS 130936 | M. gypseum        | N. gypseum     | Skin   |         |         |
| CBS 130939 | M. gypseum        | N. gypseum     | Skin   |         |         |
| CBS 171.64 | A. gypseum        | N. gypseum     | T      | Microsporum gypseum      | Birdnest |
| CBS 258.61 | A. gypseum        | N. gypseum     | NT     | Gymnoascus gypseus       | Soil |
| CBS 130948 | M. gypseum        | N. incurvata   | Skin   |         |         |
| CBS 173.64 | A. incurvatum     | N. incurvata   | Skin   |         |         |
| CBS 174.64 | A. incurvatum     | N. incurvata   | T      | Nannizzia incurvata      | Skin |
| CBS 314.54 | A. obtusum        | N. nana        | T      | Microsporum gypseum var. nanum | Scalp |
| CBS 321.61 | A. obtusum        | N. nana        | ST     | Nannizzia obtusa         | Human |
| CBS 322.61 | A. obtusum        | N. nana        | ST     | Nannizzia obtusa         | Human |
| CBS 632.82 | A. obtusum        | N. nana        | Human  |         |         |
| CBS 421.74 | A. persicolor     | N. persicolor  | ST     | Nannizzia quinckeana     | Skin |
| CBS 871.70 | A. persicolor     | N. persicolor  | ST     | Nannizzia quinckeana     | Skin |
| CBS 288.55 | M. praecox        | N. praecox     | AUT    | Microsporum praecox      | Human |
| CBS 128066 | M. praecox        | N. praecox     | Skin   | from horse               |         |
| CBS 128067 | M. praecox        | N. praecox     | Skin   | from horse               |         |
| CBS 121947 | M. amazonicum     | N. sp.         | Skin   |         |         |
| CBS 450.65 | A. racemosum      | N. sp.         | T      | Microsporum racemosum    | Rat |
| CBS 130935 | M. racemosum      | P. cookei      | Soil   |         |         |
| CBS 227.58 | A. cajetanum      | P. cookei      | Soil   |         |         |
| CBS 228.58 | A. cajetanum      | P. cookei      | T      | Microsporum cookei       | Soil |
| CBS 337.74 | A. cajetanum      | P. cookei      | Soil   |         |         |
| CBS 423.74 | A. racemosum      | P. cookei      | ST     | Nannizzia racemosa       | Soil |
| CBS 424.74 | A. racemosum      | P. cookei      | ST     | Nannizzia racemosa       | Soil |
| CBS 101.83 | A. cookielium     | P. cookielium  | ST     | Nannizzia cookielium     | Soil |
| CBS 102.83 | A. cookielium     | P. cookielium  | ST     | Nannizzia cookielium     | Soil |
| CBS 124422 | M. mirabile       | P. mirabile    | ST     | Microsporum mirabile     | Chamois |
| CBS 129179 | M. mirabile       | P. mirabile    | ST     | Microsporum mirabile     | Nail |
| CBS 646.73 | A. vanbreuseghemii| T. vanbreuseghemii | T   | Arthroderma vanbreuseghemii |         |
| CBS 809.72 | A. benhamiae      | T. benhamiae   | Skin   | from guinea pig          |         |
| CBS 112368 | A. benhamiae      | T. benhamiae   | Skin   | from guinea pig          |         |
| CBS 112369 | A. benhamiae      | T. benhamiae   | Skin   | from guinea pig          |         |
| CBS 112370 | A. benhamiae      | T. benhamiae   | Skin   | from guinea pig          |         |
| CBS 112371 | A. benhamiae      | T. benhamiae   | Skin   | from guinea pig          |         |
| CBS 112857 | A. benhamiae      | T. benhamiae   | Skin   | from rabbit              |         |
| CBS 120669 | T. mentagrophytes | T. benhamiae   | Guinea pig |             |         |
| CBS 280.83 | A. benhamiae      | T. benhamiae   | Skin   |         |         |
| CBS 623.66 | A. benhamiae      | T. benhamiae   | ST     | Arthroderma benhamiae    | Human |

(continued on next page)
| CBS number | Current taxon name | New taxon name | Status | Source |
|------------|-------------------|----------------|--------|--------|
| CBS 624.66 | A. benhamiae      | T. benhamiae   | ST     | Arthroderma benhamiae |
| CBS 806.72 | A. benhamiae      | T. benhamiae   |        | Guinea pig |
| CBS 934.73 | A. benhamiae      | T. benhamiae   |        |        |
| CBS 131645 | T. bullosum       | T. bullosum    |        | Skin   |
| CBS 363.35 | T. bullosum       | T. bullosum    |        | Horse  |
| CBS 557.50 | T. bullosum       | T. bullosum    |        |        |
| CBS 196.26 | T. concentricum   | T. concentricum| NT     | Skin   |
| CBS 448.61 | T. concentricum   | T. concentricum|        |        |
| CBS 563.83 | T. concentricum   | T. concentricum|        | Skin   |
| CBS 109036 | T. equinum        | T. equinum     |        | Skin   |
| CBS 100080 | T. equinum        | T. equinum     |        | Horse  |
| CBS 270.66 | T. equinum        | T. equinum     | NT     | Horse  |
| CBS 285.30 | T. tonsurans      | T. equinum     |        | Horse  |
| CBS 634.82 | T. equinum        | T. equinum     |        | Skin   |
| CBS 344.79 | T. erinacei       | T. erinacei    |        | Skin   |
| CBS 474.76 | T. erinacei       | T. erinacei    |        | Skin   |
| CBS 511.73 | T. erinacei       | T. erinacei    |        | Hedghog|
| CBS 124411 | T. sp.            | T. erinacei    |        | Dog    |
| CBS 220.25 | T. eriotrephon    | T. eriotrephon |        | Skin   |
| CBS 108.91 | T. erinacei       | T. interdigitale|      | Skin   |
| CBS 110.65 | T. mentagrophytes | T. interdigitale|      | Groin  |
| CBS 113880 | T. mentagrophytes | T. interdigitale|      | Nail   |
| CBS 117723 | A. vanbreuseghemii| T. interdigitale|      | Skin   |
| CBS 119447 | T. violaceum      | T. interdigitale|      | Scalp  |
| CBS 232.76 | A. vanbreuseghemii| T. interdigitale|      | Skin   |
| CBS 425.63 | A. vanbreuseghemii| T. interdigitale|      | T. Trichophyton batonrougei |
| CBS 428.63 | A. vanbreuseghemii| T. interdigitale|      | NT T. Trichophyton interdigitale |
| CBS 449.74 | A. vanbreuseghemii| T. interdigitale|      | Skin   |
| CBS 475.93 | A. vanbreuseghemii| T. interdigitale|      | T. Trichophyton krajdenii |
| CBS 559.66 | A. vanbreuseghemii| T. interdigitale|      | Skin   |
| CBS 647.73 | A. vanbreuseghemii| T. interdigitale|      | T. Trichophyton candelabrum |
| CBS 124426 | T. interdigitale  | T. mentagrophytes|  | Dog    |
| CBS 124410 | T. interdigitale  | T. mentagrophytes|  | Dog    |
| CBS 124419 | T. interdigitale  | T. mentagrophytes|  | Chamois|
| CBS 124424 | T. interdigitale  | T. mentagrophytes|  | Cat    |
| CBS 124425 | T. interdigitale  | T. mentagrophytes|  | Cat    |
| CBS 304.38 | T. radicosum      | T. mentagrophytes|  | T. Trichophyton radicosum |
| CBS 428.63 | A. vanbreuseghemii| T. interdigitale|      | NT T. Trichophyton interdigitale |
| CBS 126.34 | T. interdigitale  | T. mentagrophytes|  | T. Bodinia abyssinica |
| CBS 120324 | T. interdigitale  | T. mentagrophytes|  | Skin   |
| CBS 120356 | T. mentagrophytes | T. mentagrophytes|  | Scalp  |
| CBS 124401 | A. benhamiae      | T. mentagrophytes|  | Guinea pig |
| CBS 124404 | T. interdigitale  | T. mentagrophytes|  | Rabbit |
| CBS 124408 | T. interdigitale  | T. mentagrophytes|  | Dog    |
| CBS 124415 | T. interdigitale  | T. mentagrophytes|  | Cat    |
| CBS 124421 | T. interdigitale  | T. mentagrophytes|  | Rabbit |
| CBS 124420 | T. interdigitale  | T. mentagrophytes|  | Rabbit |
| CBS number | CBS number | Current taxon name | New taxon name | Status | Source |
|------------|------------|-------------------|---------------|--------|--------|
| CBS 120357 | CBS 120357 | *T. mentagrophytes* | *T. mentagrophytes* | Scalp | |
| CBS 158.66 | CBS 158.66 | *T. mentagrophytes* | *T. quinckeanum* | Skin | |
| CBS 318.56 | CBS 318.56 | *T. mentagrophytes* | *T. quinckeanum* | NOT NT Microsporum mentagrophytes | |
| IHEM 13697 | IHEM 13697 | *T. quinckeanum* | *T. quinckeanum* | NT Trichophyton quinckeanum | Mouse |
| CBS 100081 | CBS 100081 | *T. rubrum* | *T. rubrum* | T Trichophyton fischeri | Contaminant |
| CBS 100084 | CBS 100084 | *T. rubrum* | *T. rubrum* | T Trichophyton raubitschekii | Skin |
| CBS 100238 | CBS 100238 | *T. rubrum* | *T. rubrum* | Nail | |
| CBS 102856 | CBS 102856 | *T. rubrum* | *T. rubrum* | Nail | |
| CBS 110399 | CBS 110399 | *T. rubrum* | *T. rubrum* | Skin | |
| CBS 115314 | CBS 115314 | *T. rubrum* | *T. rubrum* | Nail | |
| CBS 115315 | CBS 115315 | *T. rubrum* | *T. rubrum* | Skin | |
| CBS 115316 | CBS 115316 | *T. rubrum* | *T. rubrum* | Human | |
| CBS 115317 | CBS 115317 | *T. rubrum* | *T. rubrum* | Skin | |
| CBS 115318 | CBS 115318 | *T. rubrum* | *T. rubrum* | Nail | |
| CBS 117539 | CBS 117539 | *T. rubrum var. flavum* | *T. rubrum* | Nail | |
| CBS 118892 | CBS 118892 | *T. rubrum* | *T. rubrum* | Nail | |
| CBS 289.86 | CBS 289.86 | *T. rubrum* | *T. rubrum* | T Trichophyton kanei | Skin |
| CBS 376.49 | CBS 376.49 | *T. rubrum* | *T. rubrum* | T Trichophyton rodhainii | Skin |
| CBS 392.58 | CBS 392.58 | *T. rubrum* | *T. rubrum* | NT Epidermophyton rubrum | Skin |
| CBS 592.68 | CBS 592.68 | *T. rubrum* | *T. rubrum* | T Trichophyton fluviomuniense | Skin |
| CBS 120425 | CBS 120425 | *T. rubrum* | *T. rubrum* | Nail | |
| CBS 202.88 | CBS 202.88 | *T. rubrum* | *T. rubrum* | Skin | |
| CBS 288.86 | CBS 288.86 | *T. rubrum* | *T. rubrum* | Contaminant | |
| CBS 118537 | CBS 118537 | *T. schoenleinii* | *T. schoenleinii* | Scalp | |
| CBS 118538 | CBS 118538 | *T. schoenleinii* | *T. schoenleinii* | Scalp | |
| CBS 118539 | CBS 118539 | *T. schoenleinii* | *T. schoenleinii* | Scalp | |
| CBS 433.63 | CBS 433.63 | *T. schoenleinii* | *T. schoenleinii* | Scalp | |
| CBS 458.59 | CBS 458.59 | *T. schoenleinii* | *T. schoenleinii* | NT Trichophyton schoenleinii | Human |
| CBS 417.65 | CBS 417.65 | *A. simii* | *T. simii* | ST Arthroderma simii | Poultry |
| CBS 448.65 | CBS 448.65 | *A. simii* | *T. simii* | ST Arthroderma simii | Poultry |
| CBS 449.65 | CBS 449.65 | *A. simii* | *T. simii* | ST Arthroderma simii | Poultry |
| CBS 520.75 | CBS 520.75 | *A. simii* | *T. simii* | ST Arthroderma simii | Poultry |
| CBS 109033 | CBS 109033 | *T. tonsurans* | *T. tonsurans* | Horse | |
| CBS 109034 | CBS 109034 | *T. tonsurans* | *T. tonsurans* | Horse | |
| CBS 112186 | CBS 112186 | *T. tonsurans* | *T. tonsurans* | Human | |
| CBS 112187 | CBS 112187 | *T. tonsurans* | *T. tonsurans* | Human | |
| CBS 112188 | CBS 112188 | *T. equinum* | *T. tonsurans* | Horse | |
| CBS 112189 | CBS 112189 | *T. tonsurans* | *T. tonsurans* | Horse | |
| CBS 112190 | CBS 112190 | *T. tonsurans* | *T. tonsurans* | Horse | |
| CBS 112191 | CBS 112191 | *T. tonsurans* | *T. tonsurans* | Human | |
| CBS 112192 | CBS 112192 | *T. tonsurans* | *T. tonsurans* | Horse | |
| CBS 112193 | CBS 112193 | *T. equinum* | *T. tonsurans* | Horse | |
| CBS 112194 | CBS 112194 | *T. tonsurans* | *T. tonsurans* | Horse | |
| CBS 112195 | CBS 112195 | *T. tonsurans* | *T. tonsurans* | Horse | |
| CBS 112198 | CBS 112198 | *T. equinum* | *T. tonsurans* | Human | |
| CBS 112856 | CBS 112856 | *A. benhamiae* | *T. tonsurans* | Scalp, zoo transmission | |

(continued on next page)
for which five loci were analyzed. A phylogenetic tree was constructed using tools provided at http://www.phylogeny.fr/ with and without GBLOCKS to eliminate poorly aligned positions and divergent regions because of the large phylogenetic distance and high divergence of studied taxa. In the *T. rubrum* complex, all known ex-type/neotype strains of synonymized species were included. In total, 134 strains were successfully amplified. Obtained sequences were aligned with MAFFT v. 8.850b using default settings, except for the 'genafpair' option (Katoh et al. 2009). The nine alignments obtained from the initial dataset of 123 strains were subjected to phylogenetic analysis using maximum likelihood (ML) in MEGA v. 8.0 software. In the second analysis of the total dataset of 264 strains, RAxML analysis v. 8.0.0 employing GTRCAT model and 1 000 bootstrap replicates and Markov chain Monte Carlo (MCMC) algorithm with MRBAYES v. 3.2.6 in the CIPRES portal (https://www.phylo.org/) were performed.

Multilocus analysis of *T. rubrum* complex

A total of 44 strains were amplified for five genes: ITS with primers ITS4 and ITS5, the D1-D2 region of LSU with primers LR0R and LR5, partial b-tubulin (*TUB2*) with primers TUB2Fd and TUB4Fd, 60S ribosomal protein L10 with 60S-908R and 60S-506F and translation elongation factor 3 (*TEF3*) with primers EF3-3188F and EF3-3984R (Stielow et al. 2015). Sequences were aligned with web alignment tool MAFFT. ITS phylogeny tree was reconstructed by MEGA v. 6.06 using Maximum likelihood with Tamura-Nei model and 500 bootstrap replications with *Trichophyton benhamiae* as outgroup. The remaining genes were explored with SNP analysis because of their high degrees of conservation.

**Table 1. (Continued).**

| CBS number | Current taxon name | New taxon name | Status | Source |
|------------|--------------------|----------------|--------|--------|
| CBS 182.76 | *T. tonsurans*     | *T. tonsurans* | H | Horse |
| CBS 318.31 | *T. tonsurans*     | *T. tonsurans* | T | Trichophyton toriforme |
| CBS 338.37 | *T. immergens*     | *T. tonsurans* | S | Skin |
| CBS 496.48 | *T. tonsurans*     | *T. tonsurans* | NT | Trichophyton tonsurans |
| CBS 130944 | *T. verrucosum*   | *T. verrucosum* | Sk | Scalp |
| CBS 130945 | *T. verrucosum*   | *T. verrucosum* | Sk | Scalp |
| CBS 130946 | *T. verrucosum*   | *T. verrucosum* | Sk | Scalp |
| CBS 130947 | *T. verrucosum*   | *T. verrucosum* | Sk | Scalp |
| CBS 134.66 | *T. verrucosum*   | *T. verrucosum* | Sk | Scalp |
| CBS 161.66 | *T. verrucosum*   | *T. verrucosum* | Sk | Scalp |
| CBS 282.82 | *T. verrucosum*   | *T. verrucosum* | C | Cow |
| CBS 326.82 | *T. verrucosum*   | *T. verrucosum* | C | Cow |
| CBS 365.53 | *T. verrucosum*   | *T. verrucosum* | NT | Trichophyton verrucosum |
| CBS 517.63 | *T. rubrum*        | *T. violaceum* | T | Trichophyton kuryangei |
| CBS 452.61 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 118535 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 119446 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 120316 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 120317 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 120318 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 120319 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 120320 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 178.91 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 201.88 | *T. violaceum*     | *T. violaceum* | S | Scalp |
| CBS 118548 | *M. ferrugineum*   | *T. violaceum* | S | Scalp |
| CBS 359.60 | *T. violaceum*     | *T. violaceum* | T | Trichophyton yaoundei |
| CBS 397.52 | *T. balcanicum*    | *T. violaceum* | T | Trichophyton balcanicum |
| CBS 374.92 | *T. violaceum*     | *T. violaceum* | S | Skin |
| IHEM 19751 | *T. soudanense*    | *T. violaceum* | NT | Trichophyton soudanense |

A = *Arthroderma*; C = *Chrysosporium*; E = *Epidermophyton*; M = *Microsporum*; N = *Nannizzia*; L = *Lophophyton*; P = *Paraphyton*; AUT = authentic; ET = epitype; NT = neotype; ST = syntype; T = (ex-)holotype.
during late log phase by centrifugation at 6 000 g for 30 min. Genomic DNA was extracted using the EZNA Fungal DNA kit (Omega Bio-tek, Doraville, GA, USA) according to the manufacturer’s instructions. Purified DNA was quantified by TBS-380 fluorometer (Turner BioSystems, Sunnyvale, CA, USA). High quality DNA (OD260/280 = 1.8–2.0 > 10 μg) was used for further research.

Library construction and sequencing

For Illumina sequencing, at least 2 μg genomic DNA was used for each strain for library construction. Paired-end libraries with insert sizes of ~300 bp were constructed according to the manufacturer’s instructions (AIR™ Paired-End DNA Sequencing Kit, BioScientific, Beijing, China). Sequencing was done using Illumina HiSeq2000 technology with PE100 or PE125 mode. Raw sequencing data was generated by Illumina base calling software CASAVA v. 1.8.2 (http://support.illumina.com/sequencing/sequencing_software/casava.html). Contamination reads, such as ones containing adaptors or primers were identified by SeqPrep (https://github.com/jstjohn/SeqPrep) with parameters: “--q 20 -L 25 -B AGATCGGAAGAGCGTCGTGT-A AGATCGGAA-
GAGCACACGTCT’. Sickle (https://github.com/najoshi/sickle) was applied to conduct reads data trimming with default parameters to get clean data in this study. Clean data obtained by above quality control processes were used to do further analysis.

For PacBio RS (Pacific Biosciences, Menlo Park, CA, USA) sequencing, 8–10 k insert whole genome shotgun libraries were generated and sequenced using standard protocols. An aliquot of 10 μg DNA was spun in a Covaris g-TUBE (Covaris, Woburn, MA, USA) at 3 500 g genomic DNA was used. DNA fragments were purified, end-repaired and ligated with SMRTbell™ sequencing adapters following manufacturer’s recommendations (Pacific Biosciences). Resulting sequencing libraries were purified three times using 0.45 × vols of Agencourt AMPure XP beads (Beckman Coulter Genomics, Danvers, MA, USA) following the manufacturer’s recommendations. PacBio RS sequencing work was performed with C2 reagents.

Genome assembly and annotation

The T. rubrum CBS 139224 genome was sequenced using a combination of PacBio RS and Illumina sequencing platforms. The T. violaceum CBS 141829 was sequenced by Illumina only. Data were used to evaluate the complexity of the genome and assembled using Velvet (v. 1.2.09) with a k-mer length of 17. Contigs with lengths of less than 200 bp were discarded to increase reliability. The assembly was first produced using a hybrid de novo assembly solution modified by Koren et al. (2012), in which a de-Bruijn based assembly algorithm and a CLR reads correction algorithm were integrated in a PacBio To CA with Celera assembler pipeline. The final assembled genome was verified manually.

The protein coding genes (CDS) were predicted with a combination of three gene prediction methods, GeneMark-ES v. 2.3a (Ter-Hovhannisyan et al. 2008), AUGUSTUS v. 2.5.5 (Stanke & Morgenstern 2015) and SNAP (http://korlab.ucdavis.edu/software.html) prediction. The resulting predictions were integrated using Clean (Elisik et al. 2007).

Open-reading frames with less than 300 base pairs were discarded. The remaining ORFs were queried against non-redundant database (nr in NCBI), SWISS-PROT (http://uniprot.org), KEGG (http://www.genome.jp/kegg/), and EGGNOG (http://eggnogdb.embl.de/#/app/home) in view of functional annotation (BLASTX cutoff: e-value < 1e^-5).

Repetitive elements and microsatellites

The assembled genomes of T. rubrum CBS 139224 and T. violaceum CBS 141829 were searched for repetitive elements by RepBase (http://www.girinst.org/server/RepBase/index.php) (Bao et al. 2015), and RepeatMasker (http://www.repeatmasker.org/) using the eukaryotic repeat database. A Perl-based script MATFINDER v. 2.0.9 was used for microsatellite identification from assembled scaffolds. The mononucleotide repeats were ignored by modifying the configuration file. The repeat thresholds for di-, tri-, tetra-, penta-, and hexa-nucleotide motifs were set as 8, 5, 5 and 5 respectively. Microsatellite sequences with flanking sequences longer than 50 bp at both sides were collected for future marker development.

Adhesin prediction

Gene families were determined using OrthoMCL v. 2.0.3 (Inflation value = 1.5, E-value < 1e^-5) and domains were annotated for each orthologous cluster using InterProScan software (http://www.ebi.ac.uk) with default parameters. On the basis of previous reports of fungal adhesion prediction, six software products were tested, viz. FungalRV (Chaudhuri et al. 2011), SignalP 4.0 (de Groot et al. 2013; Teixeira et al. 2014), PredGPI (http://gpcr2.biocomp.unibo.it/predgpi/pred.htm), FAPred (Ramana & Gupta 2010), Bio-PiPred (Teixeira et al. 2014) and TMHMM (de Groot et al. 2013; Teixeira et al. 2014). Overall results were evaluated to provide best combinations.

Mitochondrial genome assembly and annotation

The mitochondrial genomes of T. rubrum CBS 139224 and T. violaceum CBS 141829 were assembled from the Illumina reads using GRAbB (https://github.com/b-brankovics/grabb; Brankovics et al. 2016). The mitochondrial genome sequence of T. rubrum BMU 01672 (GenBank acc. Nr. FJ385026) was used as a reference for the GRAbB assembly. The final assembly was done using SPAdes 3.8.1 (Bankevich et al. 2012), which resulted in the assembly of a single circular contig in both cases.

The initial mitochondrial genome annotations were done using MFAannot (http://megasun.bch.umontreal.ca/) and were manually curated. Annotation of rRNA genes was improved using tRNAscan-SE (Pavese et al. 1994), annotation of intron-exon boundaries was improved by comparing to available reference sequences. Intron encoded proteins were identified using NCBI’s ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) and annotated using InterPro (Mitchell et al. 2015) and CD-Search (Marchler-Bauer & Bryant 2004).

Phylogeny of proteases

Whole genome sequences (nucleotide and amino acid levels) of the nine dermatophytes (T. rubrum CBS 118892, T. tonsurans CBS
112818, Nannizzia gypsea CBS 118893, Microsporum canis CBS 113480, T. equinum CBS 127.97, T. benhamiae CBS 112371, T. verrucosum HKI0517) were set as a local database in CLC genomics bench 7.0. Reference protease sequences within the protease families of subtilisins (S8A family), fungalysins (M36 family) and deuterolysins (M35 family) were loaded from UniProt database (http://www.uniprot.org/) and homologous sequences were searched in a dermatophyte comparative database (https://olive.broadinstitute.org/projects/Dermatophyte%20Comparative). Orthologs were selected when E-values were less than $10^{-10}$ and similarities were higher than 90% at nucleotide level. MEGA v. 6.0 was used to construct the phylogenetic tree. The robustness of the phylogenetic trees calculated by NJ, and the ML approach was estimated by bootstrap analyses with 500 replications using

ZHAN ET AL.
maximum likelihood and Bayesian techniques. The trees were modified in Adobe Photoshop CS5.

Secondary metabolite biosynthesis

Analysis with the antibiotic and Secondary Metabolite Analysis Shell (ANTISMASH; Weber et al. 2015) predicted several potential secondary metabolite biosynthesis gene clusters.

RESULTS

ITS phylogeny of dermatophytes

Obtained rDNA ITS sequences were used to construct a taxonomic overview of Arthrodermataceae allowing comparison of relative distances between species when ITS is used as a barcode marker. Guarromyces ceratanicus was used as outgroup in the tree constructed with Maximum likelihood using RAxML v. 8.0.0 under GTRCAT model and 1000 bootstrap replications (Fig. 2). Bootstrap support above 80 % is shown above branches. In addition to Guarromyces and Ctenomyces, seven bootstrap-supported clades can be observed which were interpreted by de Hoog et al. (2017; Clades A–G) as genera: Arthroderma, Epidermophyton, Lophophyton, Microsporum, Nannizzia, Paraphyton, and Trichophyton. In this review, Arthroderma contains 21 currently accepted species and the ITS tree shows 16 bootstrap-supported branches (76 %); similarly, the multi-species genera Microsporum (3 species), Nannizzia (9 species), Paraphyton (3 species) and Trichophyton (16 species) had a ratio of bootstrap-support of 0 %, 56 %, 100 % and 56 %, respectively.

For an assessment of primer sets and resulting phylogenetic trees, we employed four parameters to select loci with superior identification and phylogenetic power and to exclude the ones with poor performance: (1) PCR robustness, (2) number of obtained amplicons, (3) total number of supported clades (BS > 80 %) in trees, and (4) monophyly of the genera. Based on these parameters, TUB2, RP 60S L1, and TEF3 were chosen for
further analysis along with the standard ITS and LSU loci. From the total set of 264 strains five alignments were created of successfully amplified sequences, containing \( n = 238 \) (ITS), \( n = 219 \) (LSU), \( n = 198 \) (TUB2), \( n = 211 \) (TEF3) and \( n = 222 \) (RP 60S L1) sequences respectively (Table 2). The alignments were analyzed using RAxML v. 8.0.0 and MrBayes v. 3.2.6 analysis with additional RAxML v. 8.0.0 (R147) on the five alignments containing 147 sequences from successfully amplified loci (Table 3). The highest number of supported clades (\( n = 40 \)) was found in ITS three, followed by RP 60S L1 (\( n = 30 \)), TUB2 (\( n = 29 \)), and LSU (\( n = 17 \)) and TEF3 (\( n = 17 \)) (Table 2). Monophyly of seven ITS-defined genera (Arthroderma, Epidermophyton, Lophophyton, Microsporum, Nannizzia, Paraphyton, and Trichophyton) in Arthrodermataceae was taken as lead for attributing clades in the other genes using the three analyses mentioned above are displayed in (Fig. 3). ITS trees were stable, independent from size of data sets or used algorithm. In the largest comparisons, the more remote Guarramycetes ceteranacis was selected as outgroup because of an unclear position of Ctenomyces serratus in these trees, but ITS topology was identical. Nannizzia was a supported clade in all three analyses, with lower bootstrap and posterior probability values. In none of the remaining trees, the genus Nannizzia was recognized as a monophyletic group. Poor resolution was achieved with TUB2 and RP 60S L1, while LSU and TEF3 had very poor performance as phylogenetic markers. Since the ITS produced by far the largest number of supported clades, taking another gene as reference would not yield trustworthy results. Therefore, datasets of ITS sequences were concluded to be superior compared to the other four loci, decomposing the entire tree in clades with BS > 80 %. In a final analysis, the TEF3 dataset was excluded. The tree of concatenated sequences of ITS, TUB2, RP 60S L1 and partial LSU had similar topology as ITS alone (de Hoog et al. 2017).

### Table 2. Robustness of phylogenetic trees for ITS LSU, TUB2, TEF3, RP 60S L1.

| Locus | ITS | LSU | TUB2 | TEF3 | RP 60S L1 |
|-------|-----|-----|------|------|-----------|
| \( n = \) amplicons \(^1\) | 238 | 219 | 198 | 211 | 222 |
| \( n = \) clades \(^2\) | 40 | 17 | 29 | 17 | 30 |

\(^1\) Number of amplicons per locus for the data set of 264 strains.

\(^2\) Number of clades with bootstrap value BS > 70 %, in phylogenetic trees obtained from 147 strains possessing all 5 amplicons.

Polymorphism of \( T. rubrum \) complex

Polymorphism of the \( T. rubrum \) complex was explored in 47 strains was analyzed in the \( T. rubrum \) (\( n = 30 \)) and \( T. violaceum \) (\( n = 17 \)) (Table S1). LSU, TUB2, RP 60S L1 and TEF3 were nearly invariant with only a few SNPs found (data not shown). TUB2 has two genotypes due to a single difference at position 106 (Fig. S1). In TEF3 three haplotypes were found, among which 38 (81 %) belonged to a single genotype 1. Five strains (\( T. violaceum \) CBS 730.88, \( T. rubrum \) CBS 110399, \( T. violaceum \) CBS 120319, \( T. violaceum \) CBS 120318) presented as genotype 2, with six SNPs compared to genotype 1. Five strains (\( T. violaceum \) CBS 305.60, CBS 374.92, CBS 118535, CBS 120320, CBS 120317) presented as genotype 3 with a single SNP (Fig. S1). In LSU three genotypes were found (Fig. S1). RP 60S L1 comprised two genotypes, with nine strains of \( T. violaceum \) deviating from the prevalent type (Fig. S1). ITS showed limited variability (Fig. S2). A group including the neotype CBS 392.58 was nearly monomorphic, with \( T. violaceum \) separated by four SNPs. A group of strains from various African countries and with \( T. rubrum \) phenotypes shared characteristics with both main clusters. Strains of the \( T. rubrum \) cluster had 7 AT repeats at the end of ITS2 while \( T. violaceum \) mostly had 8–14 and the African \( T. rubrum \) were variable in this character.

### Table 3. Assessment of phylogenetic trees for ITS, LSU, TUB2, TEF3, and RP 60S L1 obtained by Maximum Likelihood in RAxML v8.0.0 and MrBayes v3.2.6. Numbers represent bootstrap supports and posterior probabilities of the clades higher than 80 % and 0.9 (BS > 80 %, PP > 0.9) respectively, representing the genera.

| Locus | Software | ITS | LSU | TUB2 | TEF3 | RP 60S L1 |
|-------|----------|-----|-----|------|------|-----------|
| \( T \) | B238 | R147 | R238 | B219 | R147 | R219 | B198 | R147 | R198 | B211 | R147 | R211 | B222 | R147 | R222 |
| E | 1 | 100 | 100 | X | X | X | 1 | 99 | 100 | 0 | 0 | 0 | 1 | 98 | 100 |
| N | 0.98 | 85 | 84 | X | X | X | X | X | X | X | X | X | X | X | X |
| P | 1 | 100 | 100 | 0.99 | X | 86 | 0.99 | 0 | 86 | X | X | 99 | 1 | 100 | X |
| M | 1 | 100 | 100 | 1 | 99 | 99 | 1 | 100 | 99 | 1 | 100 | 100 | 1 | 100 | 100 |
| A | 1 | 99 | 95 | 0 | X | 83 | 0.99 | 95 | 0 | X | X | X | 0 | 0 | 0 |
| L | 1 | ND | 100 | 0.97 | ND | 0 | ND | ND | ND | ND | ND | ND | ND | 1 | ND | 100 |

Abbreviations used: \( R = \) RAxML v8.0.0 software; \( B = \) MrBayes v3.2.6 software; \( T = \) Trichophyton; \( E = \) Epidermophyton; \( N = \) Nannizzia; \( P = \) Paraphyton; \( M = \) Microsporum; \( A = \) Arthroderma; \( L = \) Lophophyton; \( ND = \) no data; \( X = \) no clade.
Fig. 3. Comparison of five gene-trees based on maximum datasets of strains analyzed (ITS \( n = 238 \), LSU \( n = 219 \), TUB2 \( n = 198 \), TEF3 \( n = 211 \), RP-60S L1 \( n = 222 \)), compared with a set of strains for which all genes were sequenced \( (n = 147) \). Phylogenetic analysis was done with RaxML, MrBayes, using Guarromyces ceretanicus or Ctenomyces serratus as outgroup. Bootstrap values > 80 % are shown with the branches.
Quently 278 scaffolds (N50 = 1,335,347 bp) were obtained with a above 90 % and Q30 of clean data was above 93 %. Subse-

Table 4. Raw genome data of T. rubrum CBS 139224 and CBS 118892, T. violaceum CBS 141829.

| Index                  | CBS 139224 | CBS 118892 | CBS 141829 |
|------------------------|------------|------------|------------|
| Isolated sites         | China, nail| Germany, nail China, Scalp |           |
| Isolated time          | 2012       | 2004       | 2013       |
| Matting type           | MAT1-1     | MAT1-1     | MAT1-1     |
| Scaffold number        | 19         | 36         | 278        |
| Length of all scaffolds| 22,301,977 | 22,530,013 | 23,378,626 |
| G + C (%)              | 48.344 %   | 48.31 %    | 47.22 %    |
| Scaffolds N50          | 2,198,313bp| /          | 1,335,347bp|
| No. of genes           | 7,170      | 8,804      | 7,415      |
| N %                    | 0.055 %    | /          | 0.476 %    |
| GC content in gene region (%) | 51.2 %   | /          | 50.0 %    |
| Gene/Genome            | 53.9 %     | /          | 50.6 %    |
| Gene average length    | 1,677 bp   | 1,393 bp   | 1,595 bp   |
| Intergenetic region length | 10,277,739bp | /      | 11,549,290 bp |
| GC content in intergenic region (%) | 44.9 % | /        | 43.2 %    |
| Intergenetic length/genome (%) | 46.1 % | 51.69 % | 49.4 % |

Table 5. Specific domains for T. rubrum CBS 139224 and T. violaceum CBS 141829 by interpro analysis. "Null" refers to absence of records in Interpro database.

| E-value | ACESSION | IPR-ID | Functional domain | Annotation               |
|---------|----------|--------|-------------------|-------------------------|
| T. rubrum CBS 139224 | 0 | A7D00_4483 | NULL | NULL | transferase activity |
| T. rubrum CBS 139224 | 0 | A7D00_5801 | NULL | NULL | catalytic domain |
| T. violaceum CBS 141829 | 5.30E-11 | A7C99_7410 | IPR022414 | ATP:guanidophosphotransferase, catalytic domain |
| T. violaceum CBS 141829 | 3.70E-18 | A7C99_7399 | NULL | NULL | catalytic domain |
| T. violaceum CBS 141829 | 1.20E-21 | A7C99_7396 | IPR010009 | Apolipophorin-III | lipid transport |
| T. violaceum CBS 141829 | 2.20E-11 | A7C99_7390 | IPR011992 | EF-hand domain pair | calcium ion binding |
| T. violaceum CBS 141829 | 1.50E-06 | A7C99_7408 | NULL | NULL | catalytic domain |
| T. violaceum CBS 141829 | 4.70E-39 | A7C99_7391 | IPR022414 | ATP:guanidophosphotransferase, catalytic domain |
| T. violaceum CBS 141829 | 3.10E-20 | A7C99_7375 | IPR005204 | Hemocyanin, N-terminal |
| T. violaceum CBS 141829 | 1.10E-19 | A7C99_7415 | IPR022413 | ATP:guanidophosphotransferase, N-terminal |
| T. violaceum CBS 141829 | 5.40E-39 | A7C99_7385 | IPR000896 | Hemocyanin/hexamerin middle domain |
| T. violaceum CBS 141829 | 2.80E-12 | A7C99_7394 | IPR011992 | EF-hand domain pair | calcium ion binding |
| T. violaceum CBS 141829 | 1.10E-23 | A7C99_7376 | IPR005204 | Hemocyanin, N-terminal | transferase activity |
| T. violaceum CBS 141829 | 2.90E-102 | A7C99_7381 | IPR022414 | ATP:guanidophosphotransferase, catalytic domain |
| T. violaceum CBS 141829 | 1.50E-65 | A7C99_7374 | IPR005203 | Hemocyanin, C-terminal | transferase activity |
| T. violaceum CBS 141829 | 8.20E-11 | A7C99_7395 | IPR011992 | EF-hand domain pair | calcium ion binding |
| T. violaceum CBS 141829 | 4.30E-36 | A7C99_7404 | IPR022413 | ATP:guanidophosphotransferase, N-terminal |
Table 6. Notable different paralogs of *T. rubrum* CBS 139224 and *T. violaceum* CBS 141829.

| Orthomcl-22 | TRCMCC | Probable function | TVCMCC | Probable function |
|-------------|---------|-------------------|--------|-------------------|
| Orthomcl-33 | A7D00_1963 | Ribosome biogenesis protein BRX1 | A7C99_6542 | hypothetical protein |
| Orthomcl-37 | A7D00_2627 | phosphatidylinositol-4-phosphate 5-kinase, PIP5K | A7C99_6653 | hypothetical protein |
| Orthomcl-39 | A7D00_2721 | Cell division protein Sep4a | A7C99_523 | Mitochondrial carrier protein |
| Orthomcl-43 | A7D00_2958 | dipeptidylaminopeptidase | A7C99_206 | dipeptidylaminopeptidase |
| Orthomcl-55 | A7D00_4766 | Cutinase transcription factor 1 alpha | A7C99_2923 | Killer toxin subunits alpha/beta |
| Orthomcl-59 | A7D00_7080 | phospholipase | A7C99_3575 | Phospholipase |
| Orthomcl-60 | A7D00_814 | wd and tetratricopeptide repeat protein | A7C99_4699 | hypothetical protein |
| Orthomcl-4 | A7D00_5168 | beta-glucosidase | A7C99_1137 | Probable beta-glucosidase E |
| Orthomcl-10 | A7D00_1652 | Vacuolar protein sorting-associated protein 54 | A7C99_1678 | Vacular protein |

**Repetitive sequences and transposable elements**

Repetitive sequences (RS) have been shown to serve as vehicle to maintain genomic variability and serves evolutionary change.
(Chibana et al. 2005). Compared to Aspergillus and Candida, T. rubrum CBS 139224 and T. violaceum CBS 141829 have extremely few repeats (Chibana et al. 2005, Nierman et al. 2005). We identified a total of 77 LRFs (long repeat fragments) in T. rubrum occupying 0.04 % of the global genome, and 166 LRFs in T. violaceum with a percentage of 0.33 %. The number of transposable elements (TEs) was also low, i.e. 46 in T. rubrum (9 209 bp) and 51 in T. violaceum (11 354 bp). A total of 92 retrotransposons were predicted within T. rubrum with a total length of 10 765 bp and 213 in T. violaceum measuring 83 153 bp.

Mitochondria

Mitochondrial genomes were successfully assembled from Illumina reads using GnuB and SPAdes. The mitochondrial genomes assembled into single circular contigs for both species (Fig. 4). The lengths of the sequences were 26 985 bp and 24 305 bp for T. rubrum CBS 139224 and T. violaceum CBS 141829, respectively. The mitochondrial genomes encoded for 13 proteins typical for filamentous fungi, the rRNAs of the small and large subunit of the ribosome (rns and rnl, respectively) and 25 tRNAs. The rnl gene contained a group I intron that codes for rps3. Both mitochondrial genomes contain a ribozyme gene, rnpB, and an ORF with unknown function between the cob and nad3 genes. The genes in both genomes are in the same orientation and show complete conservation.

The most striking difference between the two mitochondrial genomes is that there is a 2.6 kb insertion in the mitogenome of T. rubrum between two tRNA genes that are found downstream the rnl gene. The insert is located between trnV(tav) and trnM(cat); in T. violaceum there is only a single nucleotide separating the two genes, while in T. rubrum there is a more than 2.6 kb region separating them. This region contains 3 ORFs, the first of them encodes a putative GIY-YIG endonuclease, the other two have no functional prediction. GIY-YIG endonucleases belong to the homing endonucleases that are frequently found in group I introns.
Functional classification of EGGNOG

The gene sequences of T. rubrum CBS 139224 and T. violaceum CBS 141829 were compared against NCBI, SWISS-Prot, KEGG and EGGNOG databases. The prediction of gene function from EGGNOG revealed 3,491 orthologous genes which accounted for 48.8% of entire genome in T. rubrum CBS 139224, while T. violaceum CBS 141829, 6,081 orthologs were generated by EGGNOG annotation and took a percentage of 82.01% of the genome. A comparison of EGGNOG classification of T. rubrum CBS 139224 and T. violaceum CBS 141829 is provided in Table S2.

Although the total numbers of genes annotated by the data-base tools are similar, the annotation power seems to be different in each category. In main traits T. violaceum and T. rubrum histograms were similar, but the former was consistently somewhat higher, which might due to more orthologous genes annotated in T. violaceum. Genes involved in cell wall biogenesis [M], cell motility [N] and extracellular structures [W] take small percentages in COG classification, related to the lack of reference databases.

Genes related to metabolism

A total of 2,585 genes involved in 277 metabolic pathways were annotated in T. rubrum CBS 139224 according to KEGG functional analysis, while 2,888 genes were predicted for T. violaceum CBS 141829 involving in 174 pathways. Table S3 displays the top-14 pathways containing more than 50 genes. Genes of carbohydrate metabolism corresponding to glycosis/gluconeogenesis, TCG cycle, and degradation of pentose phosphate, fructose, mannose, sucrose, and ketone are all present in T. rubrum and T. violaceum. Key genes responsible for lipid biosynthesis and catabolism were also annotated, including metabolism of triglycerides, glycerophospholipids, and sphingolipids. However, metabolism of arachidonic acid and linoleic acid, and the alpha-linolenic acid pathway were not complete in both fungi analyzed. Trichophyton rubrum and T. violaceum possess all pathways for biosynthesis and metabolism of the 20 basic amino acids. In addition, arginine succinate lyase, ornithine carbamoyl transferase are also identified and so does a complete urea cycle. Critical genes in porphyrin metabolic pathways are enriched; however, chlorophyll synthesis pathway is interrupted in these two dermatophytes.

Most vitamins, such as thiamine, riboflavin, vitamin B6, nicotinate and icotinamide, pantothenate, and CoA biosynthesis, folate and biotin metabolism were all found in T. rubrum and T. violaceum, but ascorbate and aldarate metabolism were missing. Both dermatophytes encode the genes that transfer nitrogen residues to L-asparagine, glycine and L-glutamate and break down the latter into ammonia. The sulfur reduction and fixation pathway is also complete in these two strains. Trichophyton violaceum has been reported to be vitamin B dependent, growing better and sporulating abundantly in vitamin B-rich media (Gräser et al. 2000). Surprisingly, however, the fungus shares this vitamin pathway with T. rubrum and shows no deficiency. Overall, the above analysis demonstrates that T. rubrum and its molecular siblings possess basic metabolic abilities as most eukaryotic organisms.

Mating type locus

While a sexual life cycle with ascigerous gymnothecia has been described for some zoophilic Trichophyton species, mating
within *T. rubrum* and *T. violaceum* has not been observed, and it is unclear whether sexuality plays an important role in their natural ecology. Sexual reproduction in heterothallic ascomycetes is governed by a single mating type locus (MAT) with two idiomorphs of highly divergent sequences: either alpha (MAT1-1) or high mobility group, HMG (MAT1-2) (Li et al. 2010). *Trichophyton rubrum* CBS 118892 and *T. violaceum* CBS 141829 are identified as of MAT1-1 type, with upstream SLA2, COX3 and APN2 and downstream the 40S rDNA encoding gene as the flanking regions (Fig. 5). With the exception of *T. equinum* CBS 127.97, which is MAT1-2, all remaining strains are MAT1-1 type.

The MAT region, compared in nine dermatophyte species by local blast in the CLC genomics bench, proved to be highly conserved (Fig. 5). Some variation is noted in the number of amino acids of MAT flanking regions of *T. rubrum* and *T. violaceum* (Fig. 5C and D).

### Adhesion

Following previous reports, six web servers/platforms with fungal adhesin predictors were consulted to search putative adhesin-like proteins (see Methods). However, unexpected results of these applications emerged. Firstly, the online database of FAAPRED can only receive 25 sequences each time, which is not applicable for genome data. Secondly, Fungal RV is suitable for prediction for some medically important yeasts and *Aspergillus* strains (Chaudhuri et al. 2011), but results were inappropriate for dermatophytes: e.g. 5 probable adhesins were yielded for *T. rubrum* CBS 118892 and 25 for *T. rubrum* CBS 139224. Thirdly, GPI-anchor Predictor has a similar problem as Fungal RV, with unstable results for dermatophytes. For these reasons three software products were selected for our prediction, i.e. SIGNALP, GPI-modification and TMHMM Server, which produced consistent results among nine genomes. Putative adhesins were chosen with the following parameters: SIGNALP 4.0 positive; TMHMM 2.0 < 1 helices and number of AA to exclude as 45 from N-terminus and 35 for C-terminus; BIG-PI positive. Table S4 lists the probable adhesins for nine dermatophytes. *Nannizzia gypsea* carried only five adhesins, while *Microsporum canis* had 17 adhesins. In *Trichophyton*, the number varied from 20 to 26, most being hypothetical proteins but some were known adhesins and GPI-anchor proteins, e.g. EGE03127.1 of *T. equinum* and EEQ28337.1 of *M. canis*. These sequences are ecm33, a gene encoding adhesins in *Saccharomyces cerevisiae*, *Candida albicans* and *Aspergillus fumigatus*. EFQ97364.1 and EFQ97072.1 of *N. gypsea* are gel4 and gel12, important genes facilitating adherence of *A. fumigatus* (Free 2013). Among the 21 genes predicted as adhesins for *T. rubrum*, 17 sequences are very similar to those of *T. violaceum*, which implies that *T. rubrum* has four specific adhesins and *T. violaceum* has eight.
Secreted proteases

Three kinds of endoproteases belonging to three families were analyzed in the nine dermatophytes under study, including metalloproteases (M36 family), deuterolysins (M35 family) and subtilisins (S8A family) (Li & Zhang 2014, Tran et al. 2016, Martinez et al. 2012). Five duplicated genes belonging to metalloprotease and five genes belonging to deuterolysin persist in the genomes. M35 family members have high sequence similarity with neutral protease 2 and clustered in the M35-A clade (Li & Zhang 2014), which is specific for Arthrodermataceae. The different protease families showed comparable patterns of similarity, with four clusters: (1) T. tonsurans and T. equinum, (2) T. verrucosum and T. benhamiae, (3) T. rubrum and T. violaceum, and (4) M. canis and N. gypseum. This corresponds to known phylogenetic distances where sequences of T. rubrum and T. violaceum are highly similar, rather closely related to other T. tonsurans/ equinum, while Microsporum and Nannizzia are remote (Rezaei-Matehkolaei et al. 2014; de Hoog et al. 2017).

Previous studies suggested that exoprotease genes have expanded independently (Li & Zhang 2014). The non-rooted trees of M35 and M36 showed similar topology suggesting that the families evolve with comparable speed, and the clustering was largely consistent with known phylogenetic relationships among the studied fungi. For this reason, we denominated the M35 genes as NP-2a to NP-2e in the order of MEP1–5 genes, for easy comparison (Fig. 6).

A total of 106 sequences were identified in the S8A family, comprising 12 sequences types which designated as Sub1 to Sub12 in the tree, except for Sub12 which was lost in M. canis and Sub8 lost in T. equinum (Fig. 7, Table S5). Two other subtilisin-like proteases (A7D00_1654 and A7D00_4929) which also contained an S8 domain were found in the genome of T. rubrum CBS 139224, but they had low identity with the classical S8A family. No similar sequences were found in T. violaceum CBS 141829. The S8A family showed much more diversity and a correspondence with phylogenetic relations was noted, as Sub7 is much closer to Sub6 and Sub8 and Sub2 may have evolved from this ancestor. The sequences could be grouped in two main clusters, with the upper one including Sub7, Sub6, Sub8, Sub2, Sub1 and Sub5, while the lower contained Sub9, Sub3, Sub4, Sub11, Sub12 and Sub10. Comparing the dendrogram of dermatophyte proteinase trees to the ribosomal tree, a similar topology became apparent. Members within the above families are listed in Table S5.

In addition to endoproteases, exoproteases play an important role in the degradation of hard keratin; these include dipeptidyl peptidase IV (DPPIV), dipeptidyl peptidase V (DPPV), leucine aminopeptidases (LAPs), carboxypeptidase A (MCPA), carboxypeptidase B (MCPB), carboxypeptidase S1 homolog A (SCPA), and carboxypeptidase S1 homolog B (SCPB) (Monod 2008; Tran et al. 2016). SED1 and SED2 are genes belonging to tripeptidyl-peptidases which degrade proteins at acidic pH and are known to be involved in virulence of Aspergillus fumigatus (Reichard et al. 2006). Genes A7D00_5713 and A7C99_545 have high identity (over 90 %) with SED2 of A. fumigatus, but for SED1 no homologs were found. All proteinases annotated were displayed in Table S6.
Secondary metabolism

Table S7 lists the results of secondary metabolism of *T. rubrum* CBS 138224 and *T. violaceum* CBS 141829. Totally nine metabolite clusters are present in both fungi, seven of which are conserved. Additionally, there is an ochratoxin A biosynthetic gene cluster specific for *T. rubrum* with high identity (66 %) to *Penicillium nordicum* (AY557343), which is absent in *T. violaceum*. Two specific clusters, i.e. patulin biosynthetic and fusaridione A biosynthetic gene cluster, seem to be present in *T. violaceum*, but the results are uncertain, with only 13 % and 12 % identity with reference sequences.

DISCUSSION

The family *Arthrodermataceae* has recently been revised on the basis of molecular phylogeny (de Hoog et al. 2017). Ancestral, mostly geophilic dermatophytes with thick-walled macroconidia were placed in *Arthroderma*, while *Trichophyton* was restricted to a clade which covers all anthropophiles in addition to some zoophilic species. Ribosomal genes ITS and partial LSU are sequenced as standard, but in order to obtain better resolution, TUB2, RP 60S L1, TEF3 were included in addition. Surprisingly and in conflict with most other groups of filamentous fungi, best resolution (highest number of supported clades) was obtained with ITS (Table 3). Monophyly was confirmed for the currently distinguished seven genera (de Hoog et al. 2017). All seven genera were represented as clades by ITS with BS > 80 %, and for *Nannizzia* ITS was even the only marker with sufficient support. Average performance was achieved with TUB2 and RP 60S L1, while LSU and TEF3 had very poor performance.

The first phylogenetic overview of *Arthrodermataceae* was reconstructed on the basis of ITS is used as barcoding marker, with *Guarromyces ceretanicus* as outgroup in the tree constructed with Maximum likelihood using RAxML v. 8.0.0 under GTRCAT model and 1000 bootstrap replications (Fig. 2). The obtained phylogeny has a topology which in main traits confirmed early phylogenies published by Gräser et al. (2000), with seven bootstrap-supported clades now recognized as genera (de Hoog et al. 2017). Since *Homo sapiens* is the phylogenetically most recent mammal host of dermatophytes, the strictly anthropophilic species should appear in derived position in the tree. *Arthroderma*, comprising prevalently geophilic species, is found as an ancestral group. *Arthroderma* contains 21 currently accepted, mostly well-resolved species associated with burrows and fur of wild animals. Species occurring on domesticated animals are found near the anthropophiles (Fig. 2). Accordingly, the tree shows an evolutionary trend of increasing association with mammal hosts. We may assume that this reflects a true phylogenetic history and therefore it can be expected that the tree is robust, providing stable nomenclature.

While *Arthroderma* species are easily distinguished with large barcoding gaps, and frequently produce elaborate gymnothelial sexual states, the anthropophilic *Trichophyton* species are difficult to distinguish molecularly, and no sexual fruit bodies are known. Some species are phenotypically reduced, in culture just producing hyphal elements or chlamydomsore-like structures. Significant adaptations are needed to colonize the hairless human skin, which may explain the loss of sexuality and reduced conidiation. Transmission takes place by skin flakes rather than by conidia or ascospores.

The close affinity of clinically different entities poses a diagnostic problem. The species *T. violaceum* is a highly specialized sibling of *T. rubrum*, having a predilection for the human scalp, while *T. rubrum* is found on skin and nails. As an alternative, it may be hypothesized that *T. violaceum* is just a phenotypically different strain of *T. rubrum* that has emerged because of differences in physiological stress exerted in different habitats. For correct affiliation of species and determination of species boundaries, understanding of virulence and adaptation are essential.

Epidemiological analyses suggested that the *Trichophyton rubrum* originated with humans in Africa and subsisted on this continent for a long time, often without causing significant disease (Gräser et al. 1999). This changed around the end of the nineteenth century when the fungus was transported on a worldwide scale via human travels and social activities. The fungus was also introduced to the Western world and became epidemic due to the preference of wearing closed leather shoes and sneakers as a part of modern life style (Gräser et al. 2000, 2007; Dismukes et al. 2003). In contrast, its close relative *T. violaceum* has remained in some endemic pockets in Africa and Asia as an agent of tinea capitis (Zhan et al. 2015) and was brought to Europe mainly by scattered, recent immigration events; it shows no tendency to epidemic expansion (Nenoff et al. 2014). In contrast to the globally widespread *T. rubrum*, *T. violaceum* as an endemic fungus is restricted to semiarid climate zones of the Mediterranean, Northern Africa, Iran and Northwestern and Southern China (Ayanbimpe et al. 2008; Patel & Schwartz 2011; Zhan et al. 2015). Using microsatellites, Ohst et al. (2004) demonstrated that *T. violaceum* showed more variation, while *T. rubrum* was nearly monomorphic suggesting a founder effect after adaptation to a new ecological niche. As a zoophilic species, *Trichophyton benhamiae* causes inflammatory disease when transmitted to humans (Drouot et al. 2009), while obligatorily anthropophilic species such as *T. rubrum* show low inflammation. In addition, the zoophile *T. benhamiae* is sexually competent. Li et al. (2010) reported that common features of the MAT locus are shared among five common dermatophytes (*N. gypseae*, *M. canis*, *T. equinum*, *T. rubrum* and *T. tonsurans*). Evolution of the MAT locus apparently is not synchronous with that of ecologically relevant parameters, and occasional successful mating between classical species can be observed (Kawasaki 2011). Our data in nine dermatophytes are consistent with these reports showing that the MAT locus is shared among species. Both *T. rubrum* and *T. violaceum* presented as MAT1-1 mating type in our study, suggesting drift of unbalanced mating types leads to loss of mating in anthropophilic species (Gräser et al. 2008).

Genome sequencing has become one of the conventional means to study the biology and ecological abilities of microbes. Pacific Biosciences developed the single-molecule real-time sequencing technique (SMRT) which enables long reads (up to 23 000 bp) and has high efficiency of 1 080 Mb each run. However, the raw data generated from the *PacBio* RS platform is inherently error-prone, with up to 17.9 % errors having been reported (Chin et al. 2011). Most of these concerns *inde* events caused by incorporation events, or intervals that are too short to be reliably detected (Eid et al. 2009). The *PacBio* platform has been widely used in viruses, bacteria and small genome size organisms, but thus far rarely for fungi of medical interest. In our study we applied two genome sequencing platforms to obtain maximum genome quality, i.e. an Illumina platform with high
accuracy, as well as a PacBio RS platform which allows long reads. Considering the high consistence of the compared species, and given the high cost of PacBio, there is no need for additional SMRT sequencing of *T. violaceum*. Trichophyton rubrum CBS 139224 yielded sequences with less scaffolds and with an N % as low as 0.055 %, which is much better than the presently available genome of *T. rubrum* CBS 118892. In addition we reconstructed the complete mitochondrial genome of *T. rubrum* CBS 139224 and *T. violaceum* CBS 141829, which showed high similarity to the previous report on this species (Wu et al. 2009).

The oriental strain *T. rubrum* CBS 139224 had a genome size of 22.3 Mb, CBS 118892 was published to be 22.5 Mb, while *T. violaceum* (CBS 141829) had 23.4 Mb. The genomes were compared with each other and with seven dermatophyte species available in the public domain. Four *Trichophyton* species ranged in size from 22.6 to 24.1 Mb. Draft genomes of dermatophytes show very high degrees of conservation, both at nucleotide and at amino acid levels (Martínez et al. 2012).

Most genomes of Eukaryotes, including fungi contain significant amounts of repetitive DNA, which usually occur in multiple copies and are with or without coding domains. In Candida albicans, major repeat sequences (MRS) have been identified in all but one chromosome (Lephart & Magee 2006). The *Aspergillus fumigatus* genome harbors rich duplication events, the majority clustering in 13 chromosomal islands, which are related to pathogenesis of clinical strains (Fedorova et al. 2008). Copy numbers and location may differ between populations and are stably inherited; the elements have widely been used in epidemiological profiling. Contrary to these fungi, dermatophytes show an extraordinarily high coherence at nucleotide and gene level, with very few repeat elements in the genome. These data indicate that dermatophytes are consistent pathogens with a short divergence time; very few genetic events occurred in the evolutionary history of dermatophytes.

Secreted proteases are key virulence factors for dermatophytes. Two types of endo-proteases are prevalent in dermatophytes, i.e. subtilisins belonging to the S8A family, and metalloproteinases which comprise two different subfamilies, the deuterolysins (M35) and the fungalysins (M36) (Monod 2008; Li & Zhang 2014). These proteases share high identity among the species and have low degrees of identity among different families (Li & Zhang 2014). The protease families M35 (deuterolysins) and M36 (fungalysin) are among the most important metalloproteinases, in which zinc is an essential metal ion required for catalytic activity. These genes are found in numerous pathogenic fungi, but show expansion in *Arthrodermataceae* (dermatophytes) and *Onygenaceae* (the family containing *Coccidioides*; Li & Zhang 2014). Members of the M35 and M36 families seem to be highly conserved among dermatophytes with signatures that are specific for *Arthrodermataceae* and have low identity with other *Onygenales* (Li & Zhang 2014). In our study, M35 and M36 have the same five copies in all nine dermatophytes. As suggested by Li & Zhang (2014), M35 duplicated and M36 was lost in *Coccidioides* compared to dermatophytes, probably due to different life styles of these fungi as systemic and cutaneous pathogens, respectively.

Thus far twelve subtilisin-encoding genes within the S8A family have been reported (Martínez et al. 2012) and in our study, they were successfully annotated in nine genomes (see Methods), with the exception of M. canis (Sub12 lost) and *T. equinum* (Sub8 lost), which both have their natural niche in animal fur (Martínez et al. 2012). Phylogenetically, *T. rubrum* (human skin) is close to *T. violaceum* (human scalp), *T. equinum* (horse) is close to *T. tonsurans* (human skin), and *T. verrucosum* (cattle) close to *T. benhamiae* (guinea pigs), while all these species are remote from *Microsporum canis* (dog) and *Nannizzia gypsea* (soil). These data matched well with known phylogenetic relationships (de Hoog et al. 2017). A high degree of similarity was found between *T. rubrum* and *T. violaceum*; no unambiguous protease difference explaining the clinical difference between the two species was found, suggesting that the divergence between the entities concerns a very recent evolutionary event.

The two ecological niches of skin and scalp differ in the constitution of keratins, lipids, and immunity (Bologna et al. 2008; Wolff et al. 2008). K1/K10 and K5/K14 are the most common keratins expressed in skin (designated as ‘soft keratin’), while Hakb1-8 and Haka1-6 are only expressed in hair-follicles (‘hard keratin’; Schweizer et al. 2007). Distribution of lipids and melanocytes are also quite different between skin and scalp. Our genome data indicate that the two species possess a high molecular conservation with 99.9 % identity at the amino acid level. Only 17 species-specific genes were discovered by protein-domain analysis. However, the genome of *T. violaceum* (23.4 Mb) is larger than that of *T. rubrum* (22.3–22.5 Mb), with more unigenes (7 415 vs.7 170). This is consistent with the recent global expansion and adaptation in *T. rubrum*, assuming that the species reduced superfluous antigens while adapting to superficial invasion of the skin.

Adherence to host tissue is a prerequisite for microorganisms for invasion and infection (de Groot et al. 2013). Fungal adhesins are cell wall proteins with a highly conserved signal peptide at the N-terminus and a GPI-anchor at the C-terminus. This enables prediction of putative adhesins by bioinformatics. At least three families are known: anagglutinin-like sequences (ALS) family, a hyphal wall protein (HWP) family, and an IFF/HYP family with over 20 genes (Mayer et al. 2013). RodA, CalA and CspA enhance adherence to bronchial epithelial cells in *Aspergillus* (Jahn et al. 2000; Levdansky et al. 2010), but as yet no adhesin genes have been described for dermatophytes because of insufficiency of genetic tools. *Fungalrv* (Chaudhuri et al. 2011) and *faafred* have been applied (Chaudhuri et al. 2011; de Groot et al. 2013; Teixeira et al. 2014), but in our study *signalp*, GPI-modification and the TMHHM server proved to be more successful. We found about 20 adhesins in each dermatophyte studied here except *N. gypsea*, which matches with its life style as a geophilic fungus. Some genes even align well with known adhesins from non-dermatophyte species. *Trichophyton rubrum* and *T. violaceum* possess 17 common adhesins, four being specific for the former and 8 for the latter. It seems plausible that the clinical difference between these species might be at least partly ascribed to these adhesins.

An alternative hypothesis for the clinical divergence of *T. rubrum*/ *T. violaceum* might be their interpretation as a single species which shows epigenetic change when growing on either the scalp or on naked skin. Possibly some pathways such as the polyketide pathway are blocked with growth on the scalp, leading to reduced sporulation and growth and to formation of pigmented secondary metabolites seen in *T. violaceum* cultures, which can be restored by repeated transfer on artificial media; sporulation is vitamin B dependent (Gräser et al. 2000). Phenotypic segregates *T. soudanense*, *T. gournvillii* and *T. yaoundei* are as yet not reliably distinguishable molecularly from *T. violaceum* and might concern mutated forms or strains in stress response. Given the probably
haploid condition of these fungi, genomic polymorphism is relatively easy and might provide some advantages in natural selection. However, this hypothesis does not match with microsatellite (Gräser et al. 2007) and Maldi-tof (A. Packeu, pers. comm.) data, which suggest a recent divergence of clonal T. rubrum from a more variable African gene pool.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (31600021), CAMS Innovation Fund for Medical Science (CAMS-I2M, 2016-I2M-3-02), the Ministry of Science and Technology Foundation special projects (2013FY113700), and Jiangsu provincial Special Program of Medical Science (BL201221003). We acknowledge Ye Tao for library construction, NGS sequencing and data analysis work at Shanghai Biozeron Biotechnology Co. Sarah Abdallah Ahmed is thanked for providing the photoplate of NGS sequencing and data analysis work at Shanghai Biozeron Biotechnology Co. We acknowledge Ye Tao for library construction, projects (2013FY113700), and Jiangsu provincial Special Program of Medical

REFERENCES

Ales AN, Occan K, Ilkil MT (2008). Diagnostic value of morphological, physiological and biochemical tests in distinguishing Trichophyton rubrum from Trichophyton mentagrophytes complex. Medical Mycology 46: 811–822.

Ayambime GM, Taghir H, Diya A, et al. (2008). Tinea capitis among primary school children in some parts of central Nigeria. Mycoses 51: 336–340.

Bankevich A, Nurk S, Antipov D, et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455–477.

Bao W, Kojima KK, Kohany O (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6: 11.

Bolognia JL, Jorizzo JL, Rapinil RP (2008). Dermatology. London, UK.

Branikovic B, Zhang H, van Diepening AD, et al. (2016). GRABB: Selective Assembly of Genomic Regions, a new niche for genomic research. PLoS Computational Biology 12: e1004753.

Burmeister A, Shellet E, Glööker G, et al. (2011). Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biology 12: R7.

Chaudhuri R, Ansari FA, Rahgunandan MV, et al. (2011). FungalRV: adhesion prediction and immunoinformatics portal for human fungal pathogens. BMC Genomics 12: 192.

Chibana H, Oka N, Nakayama H, et al. (2005). Sequence finishing and gene mapping for Candida albicans chromosome 7 and syntenic analysis against the Saccharomyces cerevisiae gene. Genetics 170: 1525–1537.

Chin CS, Sorensen J, Harris JB, et al. (2011). The origin of the Haitian cholera outbreak strain. The New England Journal of Medicine 364: 33–42.

d’Groot PW, Bader O, de Boer AD, et al. (2013). Adhesins in human fungal pathogens: glue with plenty of stick. Eukaryotic Cell 12: 470–481.

de Hoog GS, Dukké K, Monod K, et al. (2017). Toward a novel multilocus phylogenetic taxonomy for the dermatophytes. Mycopathologia 182: 5–31.

Dismukes WG, Peter G, et al. (2003). Clinical Mycology. Oxford University Press.

Drouot S, Mignon B, Fratti M, et al. (2009). Pals as the main source of two zoonotic species of the Trichophyton mentagrophytes complex in Switzerland. Arthroderma vanbreuseghemii and Arthroderma benhamiae. Veterinary Dermatology 20: 13–18.

Eid J, Fehr A, Gray J, et al. (2009). Real-time DNA sequencing from single polymerase molecules. Science 323: 133–138.

Elsk CG, Mackey AJ, Reese JT, et al. (2007). Creating a honey bee consensus gene set. Genome Biology 8: R13.

Farina C, Fazli P, Imbert G, et al. (2015). Trichophyton violaceum and T. soudanense: re-emerging pathogens in Italy. 2005-2013. New Microbiology 38: 409–415.

Fedorova ND, Khali N, Joarda VS, et al. (2008). Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. PLoS Genetics 4 e1000046.

Frie SJ (2013). Fungal cell wall organization and biosynthesis. Advances in Genetics 81: 33–82.

Gräser Y, Kühner J, Presber W (1999). Molecular markers reveal exclusively clonal reproduction in Trichophyton rubrum. Journal of Clinical Microbiology 37: 3713–3717.

Gräser Y, Kuipers AF, Presber W, et al. (2000). Molecular taxonomy of the Trichophyton rubrum complex. Journal of Clinical Microbiology 38: 3329–3336.

Gräser Y, Frohlich J, Presber W, et al. (2007). Microsatellite markers reveal geographic population differentiation in Trichophyton rubrum. Journal of Medical Microbiology 56: 1058–1065.

Gräser Y, Scott J, Summerbell R (2008). The new species concept in dermatophytes—a polyphasic approach. Mycopathologia 166: 239–256.

Havlickova B, Czaika VA, Friedrich M (2008). Epidemiological trends in skin mycoses worldwide. Mycoses 51(Suppl 4): 2–15.

Jahn B, Boukhallouk F, Litz J, et al. (2000). Interaction of human phagocytes with pigmentless Aspergillus conidia. Infection and Immunity 68: 3735–3739.

Katoh K, Asimenos G, Toh H (2009). Multiple alignment of DNA sequences with MAFFT. Methods in Molecular Biology 537: 39–64.

Kawasaki M (2011). Verification of a taxonomy of dermatophytes based on mating results and phylogenetic analyses. Medical Mycology Journal 52: 291–295.

Kim MS, Kim JK, Lee MW, et al. (2015). Epidemiology of deep cutaneous fungal infections in Korea (2006–2010). Journal of Dermatology 42: 962–966.

Koren S, Schatz MC, Walenz BP, et al. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nature Biotechnology 30: 693–700.

Leshart PR, Magee PT (2006). Effect of the major repeat sequence on mitotic recombination in Candida albicans. Genetics 174: 1737–1744.

Levdansky E, Kashi O, Sharon H, et al. (2010). The Aspergillus fumigatus cspA gene encoding a repeat-rich cell wall protein is important for normal conidial cell wall architecture and interaction with host cells. Eukaryotic Cell 9: 1403–1415.

Li J, Zhang KQ (2014). Independent expansion of zinc in metalloproteinases in Onygenales fungus may be associated with their pathogenicity. PLoS One 9: e90225.

Li W, Metin B, White TC, et al. (2010). Organization and evolutionary trajectory of the mating type (MAT) locus in dermatophyte and dimorphic fungal pathogens. Eukaryotic Cell 9: 46–58.

Marclher-Bauer A, Bryant SH (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Research 32: 327–331.

Martinez DA, Oliver BG, Gräser Y, et al. (2012). Comparative genome analysis of Trichophyton rubrum and related dermatophytes reveals candidate genes involved in infection. MBio 3: e00259–12.

Mayer FL, Wilson D, Hube B (2013). Candida albicans pathogenicity mechanisms. Virulence 4: 119–128.

Mirhendi H, Makimura K, de Hoog GS, et al. (2015). Translation elongation factor 1-gene as a potential taxonomic and identification marker in dermatophytes. Medical Mycology 53: 215–224.

Mitchell A, Chang HY, Daugherty L, et al. (2015). The InterPro protein families database: the classification resources after 15 years. Nucleic Acids Research 43: 213–221.

Monod M (2008). Secreted proteases from dermatophytes. Mycopathologia 166: 285–294.

Nenoff P, Krüger C, Ginter-Hanselmayr G, et al. (2014). Mycology—an update. Part 1: Dermatomycoses: Caustive agents, epidemiology and pathogenesis. Journal of Deutschen Dermatologischen Gesellschaft 12: 188–210.

Nieman WC, Pain A, Anderson MJ, et al. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature 438: 1151–1156.

Onist T, de Hoog GS, Presber W, et al. (2004). Origins of microsatellite diversity in the Trichophyton rubrum-T. violaceum clade (dermatophytes). Journal of Clinical Microbiology 42: 4444–4448.

Patek GA, Schwartz RA (2011). Tinea capitis: still an unsolved problem? Mycoses 54: 183–188.

Pavesi A, Contiero F, Bolchi A, et al. (1994). Indentification of new eukaryotic RNA genes in genomic DNA databases by a multistep weight matrix
analysis of transcriptional control regions. Nucleic Acids Research 22: 1247–1256.

Ramana J, Gupta D (2010). FaaPred: a SVM-based prediction method for fungal adhesins and adhesin-like proteins. PLoS One 5 e9695.

Reichard U, Léchêne B, Asif AR, et al. (2006). Sedolins, a new class of secreted proteases from Aspergillus fumigatus with endoprotease or tripeptidyl-peptidase activity at acidic pHs. Applied and Environmental Microbiology 72: 1739–1748.

Rezaei-Matehkolaei A, Mirhendi H, Makimura K, et al. (2014). Nucleotide sequence analysis of beta tubulin gene in a wide range of dermatophytes. Medical Mycology 52: 674–688.

Schweizer J, Langbein L, Rogers MA, et al. (2007). Hair follicle-specific keratins and their diseases. Experimental Cell Research 313: 2010–2020.

Stanké M, Morgenstern B (2015). Augustus: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research 33: W465–W467.

Stielow JB, Lévesque CA, Seifert KA, et al. (2015). One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. Persoonia 35: 242–263.

Teixeira MM, de Almeida LG, Kubitschek-Barreira P, et al. (2014). Comparative genomics of the major fungal agents of human and animal sporotrichosis: Sporothrix schenckii and Sporothrix brasiliensis. BMC Genomics 5: 943.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, et al. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Research 18: 1979–1990.

Tran VD, De Coi N, Feuermann M, et al. (2016). RNA sequencing-based genome reannotation of the dermatophyte Arthroderma benhamiae and characterization of its secretome and whole gene expression profile during infection. mSystems 2 e0036-16.

Vu TD, Eberhardt U, Szöke S, et al. (2012). A laboratory information management system for DNA barcoding workflows. Integrative Biology 4: 744–755.

Weber T, Blin K, Duddela S, et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43: 237–243.

Wolff K, Goldsmith L, Katz S, et al. (2008). Fitzpatrick’s dermatology in general medicine, 7th edn. McGraw-Hill Companies, Inc., USA.

Wu Y, Yang J, Yang F, et al. (2009). Recent dermatophyte divergence revealed by comparative and phylogenetic analysis of mitochondrial genomes. BMC Genomics 10: 238.

Zhan P, Li D, Wang C, et al. (2015). Epidemiological changes in tinea capitis over the sixty years of economic growth in China. Medical Mycology 53: 691–698.