The taxonomy of the *Trichophyton rubrum* complex: a phylogenomic approach

Luc Cornet¹, Elizabet D’hooge¹, Nicolas Magain², Dirk Stubbe¹, Ann Packeu¹, Denis Baurain³,* and Pierre Becker¹,*

**Abstract**

The medically relevant *Trichophyton rubrum* species complex has a variety of phenotypic presentations but shows relatively little genetic differences. Conventional barcodes, such as the internal transcribed spacer (ITS) region or the beta-tubulin gene, are not able to completely resolve the relationships between these closely related taxa. *T. rubrum*, *T. soudanense* and *T. violaceum* are currently accepted as separate species. However, the status of certain variants, including the *T. rubrum* morphotypes *meginini* and *kuryangei* and the *T. violaceum* morphotype *yaoundei*, remains to be deciphered. We conducted the first phylogenomic analysis of the *T. rubrum* species complex by studying 3105 core genes of 18 new strains from the BCCM/IHEM culture collection and nine publicly available genomes. Our analyses revealed a highly resolved phylogenomic tree with six separate clades. *Trichophyton rubrum*, *T. violaceum* and *T. soudanense* were confirmed in their status of species. The morphotypes *T. meginini*, *T. kuryangei* and *T. yaoundei* all grouped in their own respective clade with high support, suggesting that these morphotypes should be reinstituted to the species-level. Robinson-Foulds distance analyses showed that a combination of two markers (a ubiquitin-protein transferase and a MYB DNA-binding domain-containing protein) can mirror the phylogeny obtained using genomic data, and thus represent potential new markers to accurately distinguish the species belonging to the *T. rubrum* complex.

**INTRODUCTION**

Dermatophytes (*Onygenales, Arthrodermataceae*) are a group of closely related, pathogenic, fungi that cause superficial skin infections in both humans and animals. Species belonging to the *Trichophyton rubrum* complex are strictly anthropophilic dermatophytes and can infect the glabrous skin (tinea corporis and tinea pedis), the scalp (tinea capitis) and the nails (onychomycosis) in immunocompetent patients. Three species are currently accepted in the *T. rubrum* complex: *T. rubrum* (Castell.) Sabour. 1911, *T. violaceum* Sabour. ex E. Bodin 1902 and *T. soudanense* Joyeux 1912 [1–3]. *T. rubrum* has a worldwide distribution. It has become an increasingly prevalent dermatophyte in North America, Europe, Australia and East Asia since the 1950s, following a change in habits, such as the use of occlusive footwear. It causes tinea pedis, tinea corporis and onychomycosis, and is characterized morphologically by a fast growth and high sporulation in culture. *Trichophyton violaceum* is predominant in Middle Eastern countries, East Africa and South China. *In vitro*...
The taxonomic relationships of the dermatophyte fungi were revised in 2017 by de Hoog et al. [2], regrouping the species in seven molecularly supported genera. The genus *Trichophyton* is placed in a derived position on the evolutionary tree and contains both zoophilic and anthropophilic species. The rRNA intergenic transcribed spacer region (i.e. ITS1, 5.8S rRNA and ITS2) is currently the most informative marker available for this genus [1, 2, 4]. Species within the genus *Trichophyton* are closely related, with only small genetic distances between them. Interestingly, the low genetic variation of the ITS region in the *T. rubrum* complex contrasts with a high phenotypic variability. This resulted in the description of taxa that were later synonymized with the current species. The close genetic distances suggest that the species in *Trichophyton* are genetically separate species. We also provide two potential new gene markers that resolve the *T. rubrum* complex phylogeny better than conventional barcodes, which will facilitate future identifications of isolates, and studies on the relationships between species belonging to this complex.

The *T. rubrum* complex affects human health on a global scale. Groups with distinct morphology, geographical distribution and clinical aspects are well-known within the complex, despite low genetic variation. Therefore, our study aims to provide additional support for a stable taxonomy through phylogenomic analyses. Hence, we test the confirmation of *T. rubrum*, *T. violaceum* and *T. soudanense* and investigate the possible reinstatement of *T. megninii*, *T. kuryangei* and *T. yaoundei*. Finally, we mine core genes in order to identify new candidate markers for future phylogenetic analyses and species identification.

**METHODS**

**Isolates**

Twenty-seven strains belonging to the *Trichophyton rubrum* complex were selected. Nineteen strains were obtained from the BCCM/IHEM fungi collection, of which 18 were newly sequenced in this study. In addition, eight strains and their assembled genomes were downloaded from GenBank. Two strains of *Trichophyton interdigitale* were selected as the outgroup of the *T. rubrum* complex. The strains, along with their geographic origin, source of isolation and ENA accession numbers, are listed in Table 1. Isolates were chosen to maximize the diversity within the *T. rubrum* complex, including representatives of a wide variety of morphotypes (*megeninii*, *kuryangei*, *yaoundei*, *kanei*, *fischeri* and *raubitschekii*).

**DNA extraction**

The strains were cultivated for 7 to 21 days at 25 °C on Sabouraud agar with a cellophane filter for easier harvesting of fungal material. Genomic DNA was extracted using the QiaGen Genomic-tip 20 G−1 kit (QiaGen, Valencia, CA, USA), according to the manufacturer's instructions and following the protocol for yeast lysis with several adaptations: (i) before lysis, the collected fungal material...
Table 1. Details of the IHEM strains and public assemblies

| Genome accession | Strain n° | Species | Former species | Geographic origin | Source | Isolation date |
|------------------|-----------|---------|----------------|-------------------|--------|----------------|
| GCA_910591655.1  | IHEM 13979| *Trichophyton kuryangei* | *Trichophyton rubrum* (morphotype kuryangei) | Burundi | tinea capitis | 1966 |
| GCA_910591595.1  | IHEM 4712 | *Trichophyton kuryangei* | *Trichophyton rubrum* (morphotype kuryangei) | Burundi | tinea capitis | 1968 |
| GCA_012184535.1  | IHEM 26527 | *Trichophyton kuryangei* | *Trichophyton rubrum* (morphotype kuryangei) | Burundi | scalp | 1956 |
| GCA_910591615.1  | IHEM 13968 | *Trichophyton megninii* | *Trichophyton rubrum* (morphotype megninii) | Portugal | onychomycosis | 1987 |
| GCA_910591905.1  | IHEM 13976 | *Trichophyton megninii* | *Trichophyton rubrum* (morphotype megninii) | Portugal | tinea corporis | 1989 |
| GCA_000616965.1  | CBS 735.88 | *Trichophyton megninii* | *Trichophyton rubrum* (morphotype megninii) | Spain | chin | 1988 |
| GCA_910591955.1  | IHEM 23536 | *Trichophyton rubrum* | *Trichophyton rubrum* | Belgium | pachyonychia | 2012 |
| GCA_910592265.1  | IHEM 26523 | *Trichophyton rubrum* | *Trichophyton rubrum* | Netherlands | tinea pedis | 1958 |
| GCA_910592115.1  | IHEM 26721 | *Trichophyton rubrum* | *Trichophyton rubrum* | Belgium | tinea pedis | 2015 |
| GCA_910591845.1  | IHEM 4915 | *Trichophyton rubrum* | *Trichophyton rubrum* | Belgium | onychomycosis | 1989 |
| GCF_000151425.1  | CBS 118892 | *Trichophyton rubrum* | *Trichophyton rubrum* | Germany | onychomycosis | unknown |
| GCA_000616805.1  | CBS 100081 | *Trichophyton rubrum* (morphotype fischeri) | *Trichophyton rubrum* (morphotype fischeri) | Canada | contaminant | 1997 |
| GCA_000616825.1  | CBS 288.86 | *Trichophyton rubrum* (morphotype fischeri) | *Trichophyton rubrum* (morphotype fischeri) | Canada | contaminant | 1986 |
| GCA_000616845.1  | CBS 289.86 | *Trichophyton rubrum* (morphotype kanei) | *Trichophyton rubrum* (morphotype kanei) | Canada | buttock | 1986 |
| GCA_910592315.1  | IHEM 26520 | *Trichophyton rubrum* (morphotype raubitscheckii) | *Trichophyton rubrum* (morphotype raubitscheckii) | Canada | skin | 1997 |
| GCA_000616985.1  | CBS 202.88 | *Trichophyton rubrum* (morphotype raubitscheckii) | *Trichophyton rubrum* (morphotype raubitscheckii) | Canada | foot | ≤1988 |
| GCA_910591815.1  | IHEM 13459 | *Trichophyton soudanense* | *Trichophyton soudanense* | Somalia | tinea corporis | 1966 |
| GCA_910592065.1  | IHEM 19743 | *Trichophyton soudanense* | *Trichophyton soudanense* | Senegal | tinea capitis | ≤1970 |
| GCA_910592025.1  | IHEM 19744 | *Trichophyton soudanense* | *Trichophyton soudanense* | Senegal | tinea capitis | ≤1970 |
| GCA_910592235.1  | IHEM 19751 | *Trichophyton soudanense* | *Trichophyton soudanense* | Togo | tinea capitis | 1980 |
| GCA_000616865.1  | CBS 432.61 | *Trichophyton soudanense* | *Trichophyton soudanense* | Congo | unknown | 1939 |
|                 |           |         |                |                  |        |                |

continued
was placed at −80 °C for at least 30 min and subsequently lyophilized overnight, (ii) samples were immersed in liquid nitrogen and ground with a micropestle prior to enzymatic lysis, (iii) during the washing steps, DNA was washed four times instead of three times, (iv) after precipitation of DNA with isopropanol, samples were centrifuged for 60 min at 15 °C instead of 4 °C. The integrity of the extracted genomic DNA was checked with the 4200 TapeStation System (Agilent, Santa Clara, CA, USA). Highly intact DNA (DNA integrity values >8.5) was used for further analyses.

**Library construction and genome sequencing**

Genomic library preparation and whole genome sequencing were performed by Eurofins Genomics Europe Sequencing (Konstanz, Germany). Paired-end libraries were constructed based on the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) and sequenced at a 100× depth using the Illumina HiSeq 2000 platform (Illumina, Inc., San Diego, CA) with PE125 mode.

**Genome assembly and annotation**

For the 18 new strains, raw Illumina paired-end reads were trimmed for adaptors and filtered for low-quality reads using fastp v0.19.6 [10], with default settings. Remaining reads were assembled using metaSPAdes v3.10.1 [11], with default settings. The metagenomes obtained were binned using CONCOCT, with default settings [12]. The contamination levels and the completeness of fungal bins were estimated with EukCC [13], with default settings. Contigs resulting from the assembly of the fungal bins were scaffolded on the representative genome *T. rubrum* CBS 118892 (GCF_000151425.1) [7] using RaGOO v1.1 [14], with default settings. Average Nucleotide Identity (ANI) between genomes was estimated using dRep software [15], with the ‘no quality filtration’ option activated. Proteins were predicted with AMAW v1.0 (Meunier et al., unpublished), a wrapper tool for MAKER [16], with no prior gene model. Annotations are available in the figshare repository (https://doi.org/10.6084/m9.figshare.14762394.v2). However, the proteins of *T. rubrum* CBS 118892 (GCF_000151425.1) and *T. violaceum* CMCC(T)3I (GCA_001651435.1) were provided as supporting evidence, whereas RNA-seq data of *T. rubrum* (SRR9861013), freshly assembled using Trinity v2.4.0 [17], with default settings, were provided as expressed sequence tag (EST) evidence. Proteins from the 11 NCBI genomes (including two outgroups) were also predicted with AMAW.

**Core genes**

The protein sequence files were used for orthology inference using OrthoFinder v2.3.3 [18], with default settings. The script classify-mcl-out.pl [19] (available at https://metacpan.org/dist/Bio-MUST-Tools-Mcl) was used to select single-copy genes present in all genomes, resulting in the selection of 3105 core genes. Protein sequence alignments were then back-translated by capturing and aligning the corresponding DNA sequences with the programme leel ([20]; available at https://metacpan.org/dist/Bio-MUST-Apps-FortyTwo).
Phylogenomic analyses

Phylogenomic analyses were conducted on the 3105 core genes with the same jackknife protocol for both DNA and protein jackknife trees. Sequences from the 3105 files were first aligned using MAFFT v7.453 [21], run with the anysymbol, auto and reorder parameters. Conserved sites were then selected using trimAl v1.12 [22] with the gappyout option. One hundred datasets of ca. 100000 conserved positions (either nucleotides or amino-acids) were constructed by randomly combining alignment files using the script jack-ali-dir.pl from Bio-MUST-Core (available at https://metacpan.org/dist/Bio-MUST-Core). The 100 supermatrices were assembled using SCaFoS v1.30k [23], with default settings. Trees were inferred using RAxML v8.1.17 [24] under the fast experimental tree search method. The PROTGAMMALGF and GTRGAMMA model were used for protein-based and DNA-based analyses, respectively. Two consensus trees were built from the two sets of 100 jackknife trees using consense v3.695 (from the PHYLIP package [25], but modified to handle long sequence names), with default settings.

A large protein phylogenomic analysis was performed on the whole set of 3105 protein core genes. After selection of conserved sites with trimAl [22] v1.12, a supermatrix of 29 organisms × 1688754 unambiguously aligned amino-acid positions (0.87% missing character states) was assembled using SCaFoS v1.30k, with default settings. A large DNA phylogenomic analysis was performed on the same set of 3105 protein core genes. After selection of conserved sites with trimAl, a supermatrix of 29 organisms × 5083094 unambiguously aligned nucleotide positions (1.25% missing character states) was assembled with SCaFoS, as for the protein tree. The supermatrices are available in the figshare repository https://doi.org/10.6084/m9.figshare.14762394.
v2. The trees were inferred using RAxML v8.1.17 with 100 bootstrap replicates under the PROTGAMMALGF model for the protein tree and GTRGAMMA model for the DNA tree.

Ribosomal RNA region

The ribosomal RNA region was predicted and extracted for both the 18 newly sequenced genomes and the 11 NCBI genomes. RNAmmer v1.2 [26] was used in eukaryotic mode to predict small subunit (SSU) rRNA (18S) and large subunit (LSU) rRNA (28S) coordinates. Internal Transcribed Spacers (ITS) 1 and 2 were predicted using ITSx [27], with default settings. The contig corresponding to the ribosomal RNA region was manually added to the main fungal bin when it had been excluded by CONCOCT (Table S1).

Marker gene analyses

The 3105 aligned genes, post-BMGE alignments, were used to compute single-gene trees using RAxML v8.1.17 with 100 bootstrap replicates under the PROTGAMMALGF model for the protein tree and GTRGAMMA for the DNA tree. RAxML analyses were conducted on 3105 core genes in DNA and in proteins. 100×100000 concatenation of genes were produced with SCaFoS and the corresponding trees computed with RAML under the models GTRGAMMA for DNA and PROTGAMMALGF for proteins. *T. rubrum* is in red, *T. kuryangei* in orange, *T. megninii* in purple, *T. soudanense* in brown and *T. violaceum* - *T. yaoundei* is in blue. Jackknife support values are shown at the nodes.

**RESULTS**

Genome assembly and annotation

The 18 newly sequenced IHEM strains (5× *T. rubrum*; 3× *T. violaceum*; 2× *T. yaoundei*; 4× *T. soudanense*; 2× *T. megninii*; 2× *T. kuryangei*) of this study were all highly complete after the assembly process, with completeness >99.5%, as estimated by EukCC (Table S1). Ribosomal RNA subunits (SSU, LSU) and the ITS region were generally lacking in the 11 public genomes while they were mostly present in the genomes assembled in the present study (see Table S1 for details). The representative genome used for scaffolding, *T. rubrum* CBS 118892 (GCF_000151425.1), was composed of 36 scaffolds and featured an N50 of 2.15 Mb. The number of contigs and N50 of the 18 newly assembled genomes are lower, with scaffold numbers ranging from 17 to 19, and N50 ranging from 1.8 to 2.1 Mb. Pairwise ANI values computed between all genomes, including public genomes, indicated that the strains are closely related, with ANI values >99.5% between taxa of the *T. rubrum* complex.

**Fig. 2.** Comparison between consensus jackknife phylogenomic trees. Phylogenomic analyses were conducted on 3105 core genes in DNA and in proteins. 100×100000 concatenation of genes were produced with SCaFoS and the corresponding trees computed with RAML under the models GTRGAMMA for DNA and PROTGAMMALGF for proteins. *T. rubrum* is in red, *T. kuryangei* in orange, *T. megninii* in purple, *T. soudanense* in brown and *T. violaceum* - *T. yaoundei* is in blue. Jackknife support values are shown at the nodes.
Phylogenomic analyses
To accurately reconstruct the phylogeny of the *T. rubrum* complex, we used two independent methods to evaluate statistical support, namely bootstrapping and jackknifing (Fig 1). The three maximum likelihood analyses based on the 3105 core genes, i.e., the DNA jackknife analysis, the protein jackknife analysis and the large protein phylogenomic analysis, all show similar topologies (Figs 2 and 3). In the first bifurcation within the ingroup, the strains are split into two highly supported groups. The first group includes *T. violaceum* and its morphotype *yaoundei*. The *T. violaceum* strains and those from the morphotype *yaoundei* are each placed into their own clade, which both receive full bootstrap support, as well as high jackknife support (*T. violaceum*: 80; *T. yaoundei*: 86). Jackknife support values will always be lower considering that jackknife replicates are much smaller (100000 positions) and thus contain less phylogenetic signal than the longer bootstrap replicates (1688754 positions). On average, the jackknife proportions are lower in the DNA jackknife tree compared to the protein jackknife tree. The second group includes *T. rubrum*, *T. soudanense* and the *T. rubrum* morphotypes *megninii* and *kuryangei*. Within this second group, the *T. rubrum* clade is highly supported by both jackknife and bootstrap proportions. The clade comprising *T. soudanense* and the morphotypes *megninii* and *kuryangei* is well-supported in the bootstrapping analysis, but receives lower support in jackknife (60). The *T. soudanense* clade is well-supported by both jackknife and bootstrap proportions. The clade containing morphotypes *megninii* and *kuryangei* is well supported in both analyses. Each morphotype is placed into its own clade, which both receive full support in bootstrapping and lower support in the jackknife (*megninii*: 61; *kuryangei*: 60). The branching patterns within the different taxa show some variation between the different ML analyses, with the exception of the *megninii* group, which has stable internal relationships for all its strains. The evolutionary distances, as represented by branch lengths, are generally short (Fig. 3, with slightly longer lengths for the *T. violaceum* – *yaoundei* group segregating from the other taxa.

Marker genes
The tree resulting from the large phylogenomic analysis of all core genes was used as a reference to investigate the presence of potential marker genes within the core gene dataset. All sets of R-F distances for single-gene protein trees were lower than R-F distances for single-gene DNA trees. The single-gene protein trees were thus selected for systematic comparison. R-F distances for DNA and protein trees are available in Table S2.

With a cutoff of 0.40 in ‘info R-F polytomies’ values, the 3105 genes can be reduced to 11 genes (Table 2). R-F values were ≥0.38 (median=0.48, IQR=0.08), and ‘R-F polytomies’ values ≥0.33 (median=0.39, IQR=0.025), ‘info R-F’ values
Table 2. Robinson-Foulds values for best core genes. The four R-F values were computed with the ape software package. R-F values are given for the 11 best core genes and for various concatenations of them (see Methods for details). The two best concatenations are shown in red.

| Gene                                             | Individual gene | Ascending | Descending |
|--------------------------------------------------|-----------------|-----------|------------|
|                                                  | R-F  | R-F poly | Info R-F | #gene | R-F  | R-F poly | Info R-F | #gene | R-F  | R-F poly | Info R-F | #gene |
| Ubiquitin-protein transferase                     | 0.54 | 0.45     | 0.41     | 0.28   | 1    | 0.38     | 0.35     | 0.4    | 0.27   | 11    | 0.38     | 0.31     | 0.38   |
| MYB DNA-binding domain-containing protein         | 0.42 | 0.40     | 0.38     | 0.32   | 2    | 0.42     | 0.34     | 0.33   | 0.21   | 10    | 0.35     | 0.30     | 0.35   |
| hybrid PKS-NRPS enzyme                            | 0.38 | 0.33     | 0.44     | 0.32   | 3    | 0.46     | 0.33     | 0.33   | 0.21   | 9     | 0.31     | 0.27     | 0.31   |
| hypothetical protein                              | 0.42 | 0.37     | 0.42     | 0.37   | 4    | 0.46     | 0.35     | 0.46   | 0.35   | 8     | 0.27     | 0.26     | 0.27   |
| hypothetical protein                              | 0.62 | 0.52     | 0.50     | 0.38   | 5    | 0.42     | 0.31     | 0.33   | 0.21   | 7     | 0.31     | 0.28     | 0.31   |
| kynureninase                                      | 0.50 | 0.38     | 0.50     | 0.38   | 6    | 0.42     | 0.32     | 0.42   | 0.32   | 6     | 0.31     | 0.28     | 0.31   |
| hypothetical protein                              | 0.50 | 0.39     | 0.42     | 0.39   | 7    | 0.38     | 0.27     | 0.38   | 0.27   | 5     | 0.38     | 0.32     | 0.38   |
| multidrug resistance protein; ATPase activity     | 0.50 | 0.39     | 0.50     | 0.39   | 8    | 0.38     | 0.28     | 0.38   | 0.28   | 4     | 0.35     | 0.30     | 0.35   |
| hypothetical protein                              | 0.46 | 0.35     | 0.47     | 0.40   | 9    | 0.42     | 0.33     | 0.42   | 0.33   | 3     | 0.42     | 0.35     | 0.42   |
| nonribosomal peptide synthase                     | 0.50 | 0.40     | 0.50     | 0.40   | 10   | 0.42     | 0.33     | 0.42   | 0.33   | 2     | 0.38     | 0.31     | 0.38   |
| hypothetical protein                              | 0.38 | 0.40     | 0.38     | 0.40   | 11   | 0.38     | 0.31     | 0.38   | 0.32   | 1     | 0.50     | 0.49     | 0.50   |
were ≥0.38 (median=0.43, IQR=0.0975), and 'info R-F polytomies' values ≥0.28 (median=0.38, IQR=0.05). None of these 11 genes presented values <0.30, except for the best gene (Ubiquitin-protein transferase) with an 'info R-F' value of 0.28 after branch collapse. The optimal combination of these genes was tested in two ways, ascending or descending (see Methods for details). The best association is the concatenation of eight genes using the descending method (Table 2). This association presents R-F values and 'info R-F values' of 0.27 and 0.26, respectively, after branch collapse. Low R-F values can also be obtained by combining only two genes (a ubiquitin-protein transferase and a MYB DNA-binding domain-containing protein), using the ascending method. Concatenation of these two genes gives R-F values of 0.42, 'R-F polytomies' of 0.33, 'info R-F' of 0.34 and 'info R-F polytomies' of 0.21.

Comparison of the trees produced by these two concatenations to the reference tree shows only differences among strains within species-level clades (i.e. shallow internal nodes), the backbone being the same for all trees (Fig. 4). The '8-gene' concatenation shows differences within the T. soudanense and T. rubrum species while The '2-gene' concatenation shows differences in internal nodes for all taxa, with the exception of T. megninii (Fig. 4). The sequences of these two genes are available in the figshare repository (https://doi.org/10.6084/m9.figshare.14762394.v2).

**DISCUSSION**

The taxonomic units within the T. rubrum complex have been changed and debated many times over the years, and phylogenetic studies on this complex have been unable to fully resolve the relationships between the different species [2, 3, 5]. These species remained largely polyphyletic in phylogenetic trees, resulting in species boundaries drawn with the support of other characteristics such as morphology, physiology, clinical manifestation and geographic distribution. In this study, phylogenomic analysis on 3105 single-copy genes resulted in a highly resolved tree with high bootstrap support for the individual species-clades. The species T. rubrum, T. violaceum and T. soudanense are confirmed in their taxonomic status. The T. rubrum morphotypes kuryangei and megninii and the T. violaceum morphotype yaoundei each group in a distinct clade with high support, suggesting their re-establishment as separate species following the phylogenetic species concept. These conclusions are also supported by both DNA and protein-based jackknife analyses.

*Trichophyton kuryangei* Vanbreuseghem & Rosenthal was first described in 1961 [32]. This is the oldest name and therefore the epithet with priority. Its type specimen is CBS 517.63 T. This species mostly causes tinea capitis (endothrix, sometimes with a sheath of spores at the base of the hair shaft) and is endemic to Central Africa (Burundi). The colonies are white and have a downy to cottony texture with radial grooves. The reverse is rather pale with colours ranging from white and cream to yellow. Microconidia are pyriform and can also be elongated and divided by two to three septa, resembling pseudomacroconidia.

*Trichophyton megninii* Blanchard was first described in 1896 [33]. Cultures were not made at that time, therefore no authentic material exists. Also, the protologue is not detailed enough to assure that this taxon is identical to our clade. In 1902, Bodin described a similar species, named T. roseum, which might be a synonym of T. megninii [34]. However, authentic material has been lost and the detailed description mentions both anthropophilic and zoophilic (chicken) strains. Similarly, in 1909, a species called T. rosaceum Sabouraud was described, also mentioning anthropophilic and zoophilic (bird) strains [35]. It is therefore unsure whether T. roseum and T. rosaceum are synonyms of T. megninii [1]. The neotype
designated as *T. megninii* in this study is IHEM 13976=RV 67086; it originates from tinea corporis (*tinea faciei*), and was isolated in 1989 in Lisbon, Portugal, by de Sequeira. *Trichophyton megninii* mainly causes tinea corporis and tinea barbae (endothrix, sometimes with a sheath of spores at the base of the hair shaft). Geographically, it is restricted to the Mediterranean area (Portugal, Spain, Sardinia), but can occasionally be found in African regions (Burundi, Somalia). The colonies of *T. megninii* resemble those of *T. rubrum*. They are velvety to cottony and have radial grooves. The colony colour is white and the reverse is wine-red to brownish-red. Microconidia are clavate to pyriform. A unique and diagnostic feature of this species is its requirement of L-histidine for growth [1, 36]. In addition, a pink hue can be seen in the aerial mycelium of primary isolates of *T. megninii*. Nevertheless, this feature is quickly lost with subculturing [36].

The original description of *Trichophyton yaoundei* was done by Cochet & Doby-Dubois in 1957 [37]. The type strain for this species is CBS 605.60 T 

Further research is however required to test the usefulness and a MYB DNA-binding domain-containing protein, could be more suitable to reconstruct the real species relationships. Further research is however required to test the usefulness of these markers.

The 18 newly sequenced genomes cover the five species and several phenotypic variants of the *T. rubrum* complex and were scaffolded on *T. rubrum* CBS 118892 (GCF_000151425.1). Although the latter assembly lacks the whole ribosomal region, it is the genome that is indicated as representative in the NCBI reference sequence database (RefSeq), and we therefore used it as the template for scaffolding our own genomes [41]. The completeness of the genomes that were assembled in this study, estimated by EuKCC [13], is above 99.5% and comparable to the publicly available genomes. We further performed a metagenomic binning to avoid including any residual bacterial contamination of our assemblies.

The ribosomal region, consisting of the subunits 18S, ITS1, 5.8S, ITS2 and 28S, is present in all new genome assemblies, with some exceptions. Interestingly, this ribosomal region was absent in almost all of the publicly available genomes. This absence in public genomes could be the result of metagenomic binning, as this process is known to discard the region from assemblies due to variation in k-mer frequencies and in coverage [42]. This was also true for our assemblies, in which we had to manually add the ribosomal region with the main fungal bin for ten out of our 18 new assemblies.

The phylogenomic analyses showed a topologically stable segregation of the six taxa. Nevertheless, variations in the branching patterns within taxa, in internal nodes, were observed. These variations happened not only between the DNA and protein jackknife trees, but also between the protein jackknife trees and the tree obtained with the large protein phylogenomic analysis. This suggests that the signal present in the 3105 core genes might not be sufficient to resolve the phylogenetic relationships between strains within each species. These discrepancies might be due to incongruence between individual gene histories, which possibly differ from each other. This can be due to incomplete lineage sorting [42, 43], gene duplication and loss [44], or hidden paralogy [45]. This can also be due to the very close genetic distances between strains of the *T. rubrum* complex since the ANI between morphotypes is >99.5%. These discrepancies are only found within the six taxa and did not impact the conclusion of this study. The large DNA phylogenomic analysis (Fig. S1) shows the same topology as the large protein phylogenomic analysis and the jackknife trees, with high support values (bootstrap >90%). Hence, congruence between all the trees inferred in this study confirms our conclusions. Nevertheless, for future analysis of the *T. rubrum* complex, focussing on intra-taxon relationships, pangenomic analyses using accessory genes and identifying key traits of strains [46] should be considered.

The comparison of phylogenomic analyses (using bootstrap and jackknife support values), with single-gene phylogeny has already been used with success to define marker genes [47]. The low phylogenetic resolution of the core genes explained why a threshold of 0.40 in R-F values led to the selection of only 11 best genes. Within these 11 genes, concatenations were needed to obtain a sufficient phylogenetic signal in comparison to our reference tree. The best approach was to
CONCLUSION

A new phylogeny of the *T. rubrum* complex, based on genomic data, suggests that six species should be recognized in this complex, namely *T. rubrum*, *T. kuryangei*, *T. megninii*, *T. soudanense*, *T. violaceum* and *T. yaoundei*. Genome-wide analyses also revealed two potential new gene markers to facilitate molecular differentiation of these dermatophyte species.

Funding information

This work was supported by a research grant (no. B/2/191 /P2/18CC GEN-ERA) financed by the Belgian Science Policy Office (BELSPO). Computational resources were provided through two grants to DB (University of Liège ‘Crédit de démarrage 2012’ SFRD-12/04; F.R.S.-FNRS ‘Crédit de recherche 2014’ CDR J.0080.15).

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Grösser Y, Kuipers AF, Presber W, de Hoog GS. Molecular taxonomy of the trichophyton rubrum complex. J Clin Microbiol 2000;38:3329–3336.
2. de Hoog GS, Dukik K, Monod M, Packeu A, Stubbe D, et al. Toward a novel multilocus phylogenetic taxonomy for the dermatophytes. Mycopathologia 2017;182:5–31.
3. Packeu A, Stubbe D, Roesems S, Goens K, Van Rooij P, et al. Lineages within the *Trichophyton rubrum* complex. Mycopathologia 2020;185:123–136.
4. Zhan P, Dukik K, Li D, Sun J, Stielow JB, et al. Phylogeny of dermatophytes with genomic character evaluation of clinically distinct *Trichophyton rubrum* and *T. violaceum*. Stud Mycol 2018;89:153–175.
5. Su H, Packeu A, Ahmed SA, Al-Hatmi AMS, Blechert O, et al. Species distinction in the *Trichophyton rubrum* complex. J Clin Microbiol 2019;57:e00352-19.
6. Faggi E, Pini G, Campesi E, Bertellini C, Difonzo E, et al. Application of PCR to distinguish common species of dermatophytes. J Clin Microbiol 2001;39:3382–3385.
7. Martinez DA, Oliver BG, Grösser Y, Goldberg JM, Li W, et al. Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. mBio 2012;3:e00295-12.
8. Burmester A, Shelest E, Glöckner G, Heddercott C, Schindler S, et al. Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biol 2011;12:R7.
9. Persinoti GF, Martinez DA, Li W, Dügen A, Billimyre RB, et al. Whole-genome analysis illustrates global clonal population structure of the ubiquitous dermatophyte pathogen *Trichophyton rubrum*. Genetics 2018;208:1657–1669.
10. Chen S, Zhou Y, Chen Y, Gu J, fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–i890.
11. Nurk S, Meleshko D, Korobeinikov A, Pevzner PA. metaSPades: a new versatile metagenomic assembler. Genome Res 2017;27:824–834.
12. Alnberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, et al. Binning metagenomic contigs by coverage and composition. Nat Methods 2014;11:1144–1146.
13. Saary P, Mitchell AL, Finn RD. Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. Genome Biol 2020;21:244.
14. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 2019;20:224.
15. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genome comparisons that improved error induced recovery from metagenomes through de-replication. ISME J 2017;11:2864–2868.
16. Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinformatics 2014;48:4.
17. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 2011;29:644–652.
18. Emmms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 2019;20:238.
19. Van Vlierberghen M, Philippe H, Baurain D. Broadly sampled orthologous groups of eukaryotic proteins for the phylogenetic study of plastid-bearing lineages. BMC Res Notes 2021;14:143.
20. Rodriguez A, Burgon JD, Lyra M, Irisarri I, Baurain D, et al. Ferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. Mol Phylogenet Evol 2017;115:16–26.
21. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 2013;30:772–780.
22. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009;25:1972–1973.
23. Roüe B, Rodriguez-Ezepeleta N, Philippe H. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenetic analyses. BMC Evol Biol 2007;7 Suppl 1:52.
24. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688–2690.
25. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. 2004. http://www.evolution.gs.washington.edu/phylip.html. https://ci.nii.ac.jp/naid/10027219536/
26. Lagesen K, Hallin P, Redland EA, Staerfeldt H-H, Rognes T, et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–3108.
27. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol 2013;4:n.
28. Robinson DF, Foulds LR. Comparison of phylogenetic trees. Mathematical Biosciences 1981;53:131–147.
29. Smith MR. Information theoretic generalized Robinson-Foulds metrics for comparing phylogenetic trees. Bioinformatics 2020;36:5007–5013.
30. R Core Team. R: a language and environment for statistical computing. 2014. https://www.R-project.org/
31. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 2019;35:526–528.
32. Vanbreuseghem R, Rosenthal SA. *Trichophyton kuryangei* n. sp., a new African dermatophyte. Ann Parasitol Hum Comp 1961;36:797–803.
33. Bodin E. Les Champignons parasites de l’homme. Paris: Masson; 1902.
34. Sabouraud R. Le trichophyton de la poule (Trichophyton rosaceum) et la maladie humaine qu’il determine. Arch Med Expér Anat Path 1909;274–298.
35. Kane J, Summerbell R, Krajden S, Sigler L, Land G. Laboratory Handbook of Dermatophytes (A Clinical Guide and Laboratory Manual of Dermatophytes and Other Filamentous Fungi from Skin, Hair and Nails). Belmont (USA): Star Publishing Company; 1997.
36. Cochet G, Doby-Dubois M, Deblock S, Doby JM, Vaiva C. (n.d.) Contribution à la connaissance des teignes infantiles du Cameroun. Annales de Parasitologie Humaine et Comparée; 32:580–589.
37. Capella-Gutiérrez S, Marcet-Houben M, Gabaldón T. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol 2012;10:47.
38. Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A, et al. A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera Aspergillus and Penicillium. mBio 2019;10:e00925-19.
39. Galindo LJ, López-García P, Torruella G, Karpov S, Moreira D. Phylogenomics of a new fungal phylum reveals multiple waves of reductive evolution across Holomycota. Nat Commun 2021;12:4973.
40. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 2007;35:D61-5.
41. Cornet L, Bertrand AR, Hanikenne M, Javaux EJ, Wilmotte A, et al. Metagenomic assembly of new (sub)polar Cyanobacteria and their associated microbiome from non-axenic cultures. Microb Genom 2018;4:e000212.
42. Degnan JH, Sailer LA. Gene tree distributions under the coalescent process. Evolution 2005;59:24–37.
43. Wang K, Lenstra JA, Liu L, Hu Q, Ma T, et al. Complete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. Commun Biol 2018;1:169.
44. Hallett M, Lagergren J, Tofigh A. Simultaneous identification of duplications and lateral transfers. In: Proceedings of the eighth annual international conference on Research in computational molecular biology. New York, NY, USA: Association for Computing Machinery. pp. 347–356.
45. Fernández R, Gabaldón T, Dessimoz C. Orthology: Definitions, Prediction, and Impact on Species Phylogeny Inference. Scornavacca C, Delsuc F and Galtier N (eds). In: Phylogenetics in the Genomic Era. No commercial publisher | Authors open access book; . p. 2.
46. Golicz AA, Bayer PE, Bhalla PL, Bailey J, Edwards D. Pangenomics comes of age: from bacteria to plant and animal applications. Trends Genet 2020;36:132–145.
47. Cornet L, Magain N, Baurain D, Lutzoni F. Exploring syntenic conservation across genomes for phylogenetic studies of organisms subjected to horizontal gene transfers: A case study with Cyanobacteria and cyanolichens. Mol Phylogenet Evol 2021;162:107100.

Five reasons to publish your next article with a Microbiology Society journal
1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.