Genome sequence of a clinical isolate of dermatophyte, *Trichophyton rubrum* from India

Chitra Latka$^{1,2}$, Sanchita Sanchaya Dey$^{1,2}$, Siddharth Mahajan$^1$, Ramachandira Prabu$^1$, Pramod Kumar Jangir$^3$, Chhavi Gupta$^4$, Shukla Das$^4$, Vishnampettai Ganapathysubramanian Ramachandran$^4$, Sambit Nath Bhattacharya$^5$, Rajesh Pandey$^6$, Rakesh Sharma$^{2,3}$, Srinivasan Ramachandran$^{1,2}$ and Bhupesh Taneja$^{1,2,*}$

$^1$Genome Informatics and Structural Biology Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi 110020, India, $^2$Academy of Scientific and Innovative Research, New Delhi 110025, India, $^3$Microbial Biotechnology and Genomics Unit, CSIR-IGIB, New Delhi 110020, India, $^4$Department of Microbiology, UCMS & GTB Hospital, Dilshad Garden, Delhi 110095, India, $^5$Department of Dermatology, UCMS & GTB Hospital, Dilshad Garden, Delhi 110095, India and $^6$CSIR Ayurgenomics Unit - TRISUTRA, CSIR-IGIB, New Delhi 110020, India

*Corresponding author: Genome Informatics and Structural Biology Unit, Room No. 324, CSIR-IGIB, Sukhdev Vihar, Mathura Road, New Delhi-110020, India. Tel: +91-11-29879324; E-mail: btaneja@igib.res.in

One sentence summary: The first draft genome sequence of the dermatophyte *Trichophyton rubrum* from India.

Editor: Stefanie Poeggeler

ABSTRACT

*Trichophyton rubrum* is one of the major causative agents of dermatophytosis in humans worldwide. We report the draft genome sequence of *T. rubrum* var. *raubitschekii* from Delhi, India, isolated from a patient presenting symptoms of onychomycosis. The total estimated genome size of the clinical isolate is 25.2 MB containing 8265 predicted protein-coding sequences, 91 tRNA and 15 rRNA genes. Sequence analysis of the secreted subtilases, one of the major virulence factors in dermatophytes, clusters them into three subfamilies with distinct sequence features. The genome sequence is a step in understanding diversity of dermatophytes worldwide and will aid in identification of virulence factors and dissecting mechanisms of pathogenesis among them.

Keywords: *Trichophyton rubrum* var. *raubitschekii*; dermatophytosis; subtilisin proteases; next-generation sequencing; filamentous fungi

INTRODUCTION

Dermatophytosis (invasion of keratinized tissue by dermatophytes) is one of the most common superficial fungal infections, affecting millions of people annually worldwide (Havlíchová, Czaika and Friedrich 2008). It is caused by a group of filamentous fungi belonging to the genera *Trichophyton*, *Epidermophyton* and *Microsporum* and collectively called dermatophytes. Dermatophytes invade stratum corneum of the epidermis and keratinized tissues derived from it and cause chronic infections that may last over long periods of time (Weitzman and
Summer 1995; Achterman and White 2012). Dermatophyte infections are communicable, easily transmitted and associated with considerable morbidity. In addition, increasing incidences of emerging drug resistance amongst the strains often result in failure of treatment and prevents complete clearance of the fungus (Mukherjee et al. 2003; Alvarez and Silverberg 2006). Among dermatophytes, Trichophyton rubrum, an anthropophile, accounts for most number of superficial mycoses such as tinea pedis, tinea capitis, tinea corporis, tinea inguinale, tinea manuum and tinea unguium in human tissues of immune competent as well as deep-seated dermatophytosis in immune-compromised individuals (Weitzman and Summerbell 1995; Elewski 1998; Smith, Welsh and Skelton 2001; Seckin, Arikant and Haberal 2004; Das, Goyal and Bhattacharya 2007). In order to degrade keratinized structures of skin, hair and nail during infection, T. rubrum secretes a repertoire of proteins including subtilisins and metalloendoproteases which are considered as potential virulence factors (Wang et al. 2006; Monod 2008; Peres et al. 2010). The recent availability of the genome sequence of seven different dermatophyte species, including one for T. rubrum, has led to renewed efforts for identification of unique sequence features that are likely to aid infection and pathogenesis (Burmester et al. 2011; Martinez et al. 2012). Trichophyton rubrum var. rauheitschekii was found to be prevalent in tinea corporis and onychomycosis infections around Delhi. Here, we present the draft genome sequence of a clinical isolate of the fungus, T. rauheitschekii var. rauheitschekii from Delhi isolated from the nail tissue of a patient presenting symptoms of onychomycosis and designate it as T. rubrum IGIB-SBL-C11.

GENOME SEQUENCE OF T. rubrum IGIB-SBL-C11

Trichophyton rubrum IGIB-SBL-C11 was grown on Sabouraud's dextrose agar at 30°C for 7 days and genomic DNA was isolated using Wizard Genomic DNA Purification Kit (Promega). The purity and concentration of the isolated DNA was checked using a Nanodrop Spectrophotometer ND-1000 and the integrity was checked by running a 0.8% agarose gel. The isolate was confirmed as T. rubrum by sequencing the region spanning nuclear ribosomal internal transcribed spacer regions 1 to 2 (Fig. S1, Supporting Information) using universal primers ITS1 and ITS4 (White et al. 1990). Trichophyton rubrum IGIB-SBL-C11 is available from the authors upon request.

Sequencing was carried out employing a whole genome shotgun strategy using 454 GS FLX+, which produced a total of 1215 576 reads. The reads were assembled using Newbler (version 2.9) into 382 large contigs (>500 bp) including five mitochondrial contigs. The total estimated size of the assembled genome of T. rubrum is 25.2 Mb with a G + C content of 48.2%. The assembled contigs hence represent an average sequence depth of 30-fold with the largest contig size of 883 kbp and an N50 length of 216 kbp. The genome harbors 91 tRNA genes and 15 rRNA genes as predicted by tRNAscan-SE 1.21 (Schattner, Brooks and Lowe 2005) and R Nammer (Lagesen et al. 2007), respectively. Sequencing statistics are given in Table 1A.

Putative protein-coding genes in the genome of T. rubrum IGIB-SBL-C11 were predicted by Augustus (Hoff and Stanke 2013) trained with cDNA sequences of the RefSeq strain, T. rubrum CBS 118892. Preliminary annotation identified 8265 predicted protein-coding sequences (cds) in the genome of Indian isolate of T. rubrum, which is slightly different from the 8706 predicted cds of T. rubrum CBS 118892 (Martinez et al. 2012), possibly due to the alternate prediction algorithms used and/or different stringency in defining alternate transcripts. The average length of cds of the predicted genes in T. rubrum IGIB-SBL-C11 is 1463 bp. A total of 4472 of the 8265 predicted protein-coding sequences could be grouped into at least one or more functional categories defined by ‘clusters of orthologous groups’ (COG) database (Table S1, Supporting Information) (Tatusov et al. 2001). 8.2% of the total predicted protein-coding sequences were placed in ‘information storage and processing’, 10.3% in ‘cellular processes and signaling’, 16.7% in ‘metabolism’ and 11.6% sequences containing conserved domains of uncharacterized function were placed in ‘poorly characterized’ category (Fig. S2, Supporting Information). Multiple functions could be ascribed to a small subset of the predicted sequences and were grouped into ‘multiple families’ category (7.3%).

Secreted proteases have earlier been suggested to be important virulence factors of dermatophytes (Jousson et al. 2004; Monod 2008) and found to be enriched in dermatophyte genomes (Burmester et al. 2011; Martinez et al. 2012). The secreted proteome or the ‘secretome’ of T. rubrum IGIB-SBL-C11 was identified with the help of SignalP 4.1 (Petersen et al. 2011), which predicted 575 secreted proteins. Secreted proteases constitute one of the major predicted enzymes of the secretome of T. rubrum IGIB-SBL-C11 with subtilisases (belonging to SB clan of Metalloendopeptidases, MEPs)

Table 1. Sequencing and gene prediction statistics of T. rubrum IGIB-SBL-C11.

<table>
<thead>
<tr>
<th>A. Genome sequencing</th>
<th>B. Protein prediction and annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size, Mb</td>
<td>25.2</td>
</tr>
<tr>
<td>GC content, %</td>
<td>48.2</td>
</tr>
<tr>
<td>tRNAs</td>
<td>91</td>
</tr>
<tr>
<td>rRNAs</td>
<td>15</td>
</tr>
<tr>
<td>Protein coding genes</td>
<td>8265</td>
</tr>
<tr>
<td>Average coding sequence length, bp</td>
<td>1463</td>
</tr>
<tr>
<td>Secreted proteins</td>
<td>575</td>
</tr>
<tr>
<td>SB Subtilases</td>
<td>16</td>
</tr>
<tr>
<td>S8A subfamily</td>
<td>12</td>
</tr>
<tr>
<td>S8B subfamily</td>
<td>1</td>
</tr>
<tr>
<td>S53 subfamily</td>
<td>3</td>
</tr>
<tr>
<td>Aminopeptidases</td>
<td>7</td>
</tr>
<tr>
<td>Asparticpeptidases</td>
<td>7</td>
</tr>
<tr>
<td>Carboxypeptidases</td>
<td>12</td>
</tr>
<tr>
<td>Metalloendopeptidases</td>
<td>9</td>
</tr>
</tbody>
</table>
Figure 1. Phylogenetic and domain analysis of secreted subtilases of *T. rubrum* IGIB-SBL-C11. (A) A phylogenetic tree of secreted subtilases identified in *T. rubrum* IGIB-SBL-C11 was constructed by the neighbor-joining method using MEGA. The secreted subtilases of *T. rubrum* IGIB-SBL-C11 cluster into three subfamilies. (B) Schematic presentation of the secreted subtilases highlighting the characteristic sequence and domain features of the three subfamilies; S8A (subtilisin-like) proteases, S8B (kexin) proteases and S53 (sedolisin) proteases. The sequence identities of the subtilisins in *T. rubrum* IGIB-SBL-C11 and *T. rubrum* CBS 118892 (RefSeq) are following: Sub1: TRSBL1_103, TERG_03400; Sub2: TRSBL1_3220, TERG_08260; Sub3: TRSBL1_6346, TERG_03815; Sub4: TRSBL1_4446, TERG_01617; Sub5: TRSBL1_7279, TERG_08201; Sub6: TRSBL1_4728, TERG_03900; Sub7: TRSBL1_2497, TERG_05699; Sub8: TRSBL1_7642, TERG_01617; Sub9: TRSBL1_4922, TERG_08008; Sub10: TRSBL1_3675, TERG_06625; Sub11: TRSBL1_530, TERG_00619, respectively.

Of the subtilisin-like proteases of *T. rubrum*, namely, Sub3 and Sub5 on keratin (Maranhao, Paiao and Martinez-Rossi 2007; Peres et al. 2010) or Sub3 and Sub4 in media containing human skin sections are upregulated during growth (Leng et al. 2009). Sub3, Sub4 and Sub7 were also found to be upregulated during growth on keratin in another closely related dermatophyte, *Arthroderma benhamiae* (Burmester et al. 2011) while Sub1, Sub2, Sub3, Sub6 and Sub7 were identified in the secretome of *A. benhamiae* in a guinea pig infection model (Staib et al. 2010).

In conclusion, the genome sequence of *T. rubrum* IGIB-SBL-C11 will prove to be a valuable resource for identifying the genes involved in various metabolic pathways along with
elucidation of molecular mechanisms of its pathogenesis and in comparative analyses of *T. rubrum* strains affecting other parts of the world. Detailed investigation of a more comprehensive repertoire of secreted enzymes of *T. rubrum* will help underline their importance in the growth of the fungal pathogen during infection.

Nucleotide sequence accession number: The draft genome sequence of *T. rubrum* has been deposited in GenBank under the accession no JGR00000000. The version described in this paper is version JGR01000000. The Bioproject designation of this project is PRJNA253358.

**SUPPLEMENTARY DATA**

Supplementary data is available at FEMSLE online.

**ACKNOWLEDGEMENTS**

We would like to thank Dr Mitali Mukerji for access to sequencing facility (CSIR-TRISUTRA), Dr Rajesh Gokhale for support for the project, Ajit Singh for help with figures and Deepak Tiwari for help in data upload to NCBI.

**FUNDING**

This work was supported by Council of Scientific and Industrial Research (grant number BSC0302 to BT). Financial support to RP is provided by Council of Scientific and Industrial Research (CSIR-Translational Research and Innovative Science ThRough Ayurveda project MLP901).

**Conflict of interest.** None declared.

**REFERENCES**