Computational analysis of conserved coil functional residues in the mitochondrial genomic sequences of dermatophytes

Bulbul Gupta & Jaspreet Kaur*

Biotechnology Branch, University Institute of Engineering and Technology (U.I.E.T), Punjab University, Chandigarh 160014, India. Jaspreet Kaur – E-mail: jaspreet_uiet@pu.ac.in; *Corresponding author

Abstract:
Dermatophyte is a group of closely related fungi that have the capacity to invade keratinized tissue of humans and other animals. The infection known as dermatophytosis, caused by members of the genera *Microsporum*, *Trichophyton*, and *Epidermophyton* includes infection to the groin (tinea cruris), beard (tinea barbae), scalp (tinea capitis), feet (tinea pedis), glabrous skin (tinea corporis), nail (tinea unguium), and hand (tinea manuum). The identification of evolutionary relationship between these three genera of dermatophyte is epidemiologically important to understand their pathogenicity. Mitochondrial DNA evolves more rapidly than a nuclear DNA due to higher rate of mutation but is very less affected by genetic recombination, making it an important tool for phylogenetic studies. Thus, here we present a novel scheme to identify the conserved coil functional residues of *Trichophyton rubrum*, *Trichophyton mentagrophytes*, *Epidermophyton floccosum* and *Microsporum canis*. Protein coding sequences of the mitochondrial genome were aligned for their similar sequences and homology modelling was performed for structure and pocket identification. The results obtained from comparative analysis of the protein sequences revealed the presence of functionally active sites in all the species of the genera *Trichophyton* and *Microsporum*. However in *Epidermophyton floccosum* it was observed in three protein sequences of the five studied. The absence of these conserved coil functional residues in *E. floccusum* may be correlated with lesser infectivity of this organism. The functional residues identified in the present study could be responsible for the disease and thus can act as putative target sites for drug designing.

Keywords: Dermatophyte, Genome, Homology, Protein structure, Protein sequences.

Background:

Dermatophytes represent the type of fungi that causes infection of the skin, hair and nails due to their ability to utilize keratin. The major types of dermatophyte infections include: Tinea capitis (scalp), Tinea pedis (feet), Tinea cruris (groin), and Tinea corporis (other body surfaces). Examples of infection include ringworms, jock-itch and athlete's foot. These types of infections are widespread and increasing in prevalence on a global scale. Tinea corporis and Tinea cruris were the most common type of dermatophytosis followed by Tinea pedis, Tinea unguium and Tinea capitis [1, 2, 3]. Despite the availability of new systemic antifungal therapies, nail infections (onychomycosis) are difficult to eradicate, with recurrence reported in number of cases. Research in this field has grown substantially in recent years [4]. A review elaborating the importance of study of the dermatophytes shows that in the United States alone, millions of individuals are treated for dermatophyte infections annually, translating into an economic burden estimated to exceed $400 million per year. Despite this, researchers lack a deeper understanding of dermatophyte pathogenesis. Other reports on dermatophyte pathogenesis and treatment indicate a need for new antifungal drugs that can overcome the limitations of existing drugs [5]. Also because of development of drug resistance in the pathogen, against the presently available anti-

dermatophytic antibiotics, there is an immediate requirement of newer and more effective therapeutics.

Dermatophytic fungi constituting three genera *Trichophyton*, *Epidermophyton*, and *Microsporum* commonly cause skin diseases in humans and animals due to their ability to obtain nutrients from keratinized material. Acid proteinases, elastase, keratinase, and other proteinases act as virulence factors [6, 7]. Lund and Deboer reported preparation of subunit vaccines based on virulence factors like the keratinases but with limited success. *Trichophyton* species play a major role in dermatophyte infections of both the feet and skin. Dermatophytosis is a relatively common disease in many countries. In Western world and India, *T. rubrum, T. mentagrophytes* and *Microsporum audouinii* were found to be most frequent pathogens associated with onychomycosis [8]. In countries with warm climates and other geographic regions, different dermatophyte species may be involved; however, *T. rubrum* remains one of the most frequent clinical isolates. Majority of infections were observed with *T. violaceum, T. rubrum, T. mentagrophytes* and *Microsporum canis* in patients with dermatophytosis [9, 10]. Further it was observed that incidence of dermatophytic infections is much higher in hot and humid climate. A case study presented that dermatophytic infections in immuno-suppressed patients is mainly caused by *Trichophyton tonsurans* [11].

Fungi belonging to the genera *Trichophyton* and *Microsporum* are most often isolated from clinical cases. However, the occurrence of *Epidermophyton floccosum* was less prevalent as compared to *Trichophyton mentographytes and Trichophyton rubrum* [12]. Despite the high incidence of dermatophyte infections, complete understanding of organism's biology and research in this area is lacking because of lack of genetic tools available for such studies. Complete genome sequences for different genera of dermatophyte have not been reported till date and only a few mitochondrial genomic sequences for this organism are available. The researchers have reported that organization of genes within the mtDNAs of the 6 dermatophytes were substantially identical, reinforcing the view that dermatophytes are closely related and constitute a highly conserved lineage of filamentous fungi [13]. Furthermore, mitochondrial DNA (mtDNA) has been reported to be less affected by genetic recombination, mainly due to a predominantly uniparental heredity and higher rate of evolution than the nuclear genome. Very few mtDNA sequence variations in fungal species are mainly because of variation in intergenic regions and intronic sequences. These features make mitochondrial DNA a potentially powerful source for such studies because of its highly conserved nature.

Comparative studies provide an insight into the similarities and differences between species. Thus here we present a computational mitochondrial genomic analysis of dermatophytes that reveals the presence of conserved coil residues in most of the strains studied. These residues act as active sites and are likely to

be responsible for infectivity thus, can serve the target sites for drug designing [14].

Methodology:
The identification of functional residues in dermatophytes was carried in seven steps viz. Genome retrieval, open reading frame determination, similar sequence identification, conserved sequence determination, structure determination, evaluation of loop residues, and determination of functional residues.

Genome Retrieval
The whole organism genome sequence provides an insight into all the possible proteins transcribed in the organism in different Open Reading Frames of the genes. The GenBank at NCBI was searched for the genome sequence of various strains of pathogenic dermaotophytes which mainly include: *Chrysosporium species, Epidermophyton species, Microsporum species* and *Trichophyton species.* Out of all these fungal genera, the mitochondrion genome of four organisms was chosen for the present study, as the complete mitochondrion genome sequence of only these dermatophytes was available at GenBank.

Open reading frame determination
Protein coding frames of the selected mitochondrion genome were determined using ORF Finder and the parameter used for calculations of the proteins was that at least 100 amino acids should be present in one protein sequence [15]. The search results showed all the possible proteins that could be translated by the organism. The sequences of these proteins were retrieved for further steps of study.

Similar sequence determination
The protein sequences similar to the proteins retrieved by the ORF Finder were determined using a search tool 'BLASTnr & BLASTp'. The results obtained from BLASTp were only considered further as it provided a refined research methodology to enquire about the structurally possible protein sequence. Out of the BLASTp results for all the protein sequences, only those were selected for further studies that had acceptable BLASTp result i.e. Alignment score >= 200.

Conserved sequence determination
After obtaining the similar sequences from the BLASTp results, multiple sequence alignment of the obtained protein sequences was performed to determine the conserved residues. Multiple sequence alignment was performed with five similar sequences by software Clustal W [16].

Structure determination
For the structure determination of the selected protein sequences, homology modelling was performed using software Modeller 9v6. Homology modelling involves taking a known sequence with an unknown structure and mapping it against a known structure of one or several homologous proteins. Knowledge of the three dimensional structure is a prerequisite for the rational

drug designing and also can be of great importance for site-directed mutagenesis.

Evaluation of loop residues
After obtaining the structure of the protein, evaluation of the loop residues present in the structure was done by two methods: SPDB Viewer and by using the online GOR V server [17]. The secondary structure prediction and the probabilities for each secondary state element at each position were determined by GOR V.

Determination of functional residues
Using Ligsite, pockets for all the protein sequences were determined. The protein surface can form pockets which are binding sites of small molecules/ ligands [18]. The determination of pockets on protein surface is therefore a prerequisite for identifying functional residues.

Results and Discussion:
Complete mitochondrial genomic sequence and the accession number of the selected four species obtained from NCBI are as follow: *Epidermophyton floccosum* (NCBI: NC_007394) of 30910 bp, *Microsporum canis* (NCBI: FJ385030) of 23943 bp, *Trichophyton mentagrophytes* (NCBI: NC_012826) of 24297 bp and *Trichophyton rubrum* (NCBI: NC_012824) of 26985 bp. Results of the protein sequences obtained from analysis of these genomic sequences shows that: for *Epidermophyton floccosum,* the number of protein sequences calculated is 66; for *Microsporum canis,* 43; for *Trichophyton mentagrophytes,* 44 and for *Trichophyton rubrum* it was 50, respectively.

The results of BLASTp and BLASTnr for these protein sequences were compared and sequences having acceptable BLASTp results were selected for further studies, shown in the **Table 1** with their sequence description. These results indicated that out of 66 protein sequences obtained from ORF Finder for *Epidermophyton floccosum*, only 5 sequences were chosen (with Alignment score >= 200) for further studies. Similarly for *Microsporum canis,* 5; for *Trichophyton mentagrophytes,* 4 and for *Trichophyton rubrum* 4 sequences were selected out of 43, 44 and 50 respectively.

Multiple sequence alignment was performed for the selected protein sequences, and conserved residues present among these sequences were identified. After determining the conserved residues, structure was developed for all the selected sequences using Modeller 9v6 for the identification of loop or coiled residue. **Figure 1** shows the presence of α-helices and coils in the structure model obtained.

In order to further understand the secondary structure of the selected proteins, the probability of secondary state element at each position was calculated, that indicated the presence of coiled residues in all the structures of selected proteins. Thus by comparing the results obtained from GOR V and Modeller 9v6, the position of coils in the structure could be marked. Finally the functional residues in the sequences were identified by analysing the presence of pockets, as shown in **Figure 2.** In the present study Ligsite was used for identification of pockets, however a report suggests the use of Pocket Picker for improved results [19].

Table 1: Selected number of sequences having acceptable **BLASTp** score.

Organism Name	# using ORF Finder	# using acceptable BLASTp score	Sequence number	Sequence description
Epidermophyton floccosum	66	5	14	lcl|Sequence 14 ORF:22556..22873 Frame +2
			20	lcl|Sequence 20 ORF:13854..14117 Frame +3
			22	lcl|Sequence 22 ORF:9737..9988 Frame +2
			25	lcl|Sequence 25 ORF:22256..22492 Frame +2
			36	lcl|Sequence 36 ORF:11805..11972 Frame +3
Microsporum canis	43	5	11	lcl|Sequence 11 ORF:15739..16026 Frame +1
			15	lcl|Sequence 15 ORF:10508..10771 Frame +2
			17	lcl|Sequence 17 ORF:7379..7630 Frame +2
			18	lcl|Sequence 18 ORF:15439..15675 Frame +1
			26	lcl|Sequence 26 ORF:8330..8497 Frame +2
Trichophyton mentagrophytes	44	4	9	lcl|Sequence 9 ORF:15562..15897 Frame +1
			14	lcl|Sequence 14 ORF:10342..10605 Frame +1
			17	lcl|Sequence 17 ORF:15262..15498 Frame +1
			24	lcl|Sequence 24 ORF:8283..8450 Frame +3
Trichophyton rubrum	50	4	12	lcl|Sequence 12 ORF:18240..18575 Frame +3
			18	lcl|Sequence 18 ORF:13021..13284 Frame +1
			21	lcl|Sequence 21 ORF:10002..10253 Frame +3
			22	lcl|Sequence 22 ORF:17940..18176 Frame +3

= number of sequences

Table 2: The conserved coil functional residues identified near the pockets.

Organism Name	# using ORF Finder	# using acceptable BLASTp score	Sequence Number	Sequence description	Start and stop positions for conserved coil functional residues.
Epidermophyton floccosum	66	5	14	lcl|Sequence 14 ORF:22556..22873 Frame +2	A-47 to P-53.
			20	lcl|Sequence 20 ORF:13854..14117 Frame +3	Q-61 to M-73, S-81 to L-87.
			22	lcl|Sequence 22 ORF:9737..9988 Frame +2	V-68 to A-74, S-22 to A-30.
			25	lcl|Sequence 25 ORF:22256..22492 Frame +2	No coiled residues
			36	lcl|Sequence 36 ORF:11805..11972 Frame +3	No coiled residues
Microsporum canis	43	5	11	lcl|Sequence 11 ORF:15739..16026 Frame +1	K-16 to F-24, I-84 to K-95.
			15	lcl|Sequence 15 ORF:10508..10771 Frame +2	Q-61 to P-74, A-40 toQ-47.
			17	lcl|Sequence 17 ORF:7379..7630 Frame +2	A-23 to A-30.
			18	lcl|Sequence 18 ORF:15439..15675 Frame +1	M-1 to G-6.
			26	lcl|Sequence 26 ORF:8330..8497 Frame +2	A-22 to A-24, H-31 to T-33.
Trichophyton mentagrophytes	44	4	9	lcl|Sequence 9 ORF:15562..15897 Frame +1	R-15 to N-23, L-48 to P-53.
			14	lcl|Sequence 14 ORF:10342..10605 Frame +1	S-21 to H-27, S-33 to L-34.
			17	lcl|Sequence 17 ORF:15262..15498 Frame +1	F-52 to P-54, A-56 to C-60.
			24	lcl|Sequence 24 ORF:8283..8450 Frame +3	A-22 to A-24, H-31 to Y-34.
Trichophyton rubrum	50	4	12	lcl|Sequence 12 ORF:18240..18575 Frame +3	L-14 to G-23, A-47 to R-52.
			18	lcl|Sequence 18 ORF:13021..13284 Frame +1	I-20 to H-27, Y-41 to R-44.
			21	lcl|Sequence 21 ORF:10002..10253 Frame +3	L-21 to I-29, G-69 to P-76.
			22	lcl|Sequence 22 ORF:17940..18176 Frame +3	M-53 to L-57, G-20 to F-27.

= number of sequences [a](A: Alanine, M: Methionine, P: Proline, Q: Glutamine, S: Serine, L: Leucine, K: Lysine, I: Isoleucine, F: Phenylalanine, G: Glycine, H: Histidine, T: Threonine, R: Arginine, N: Asparagine, C: Cysteine, Y: Tyrosine).
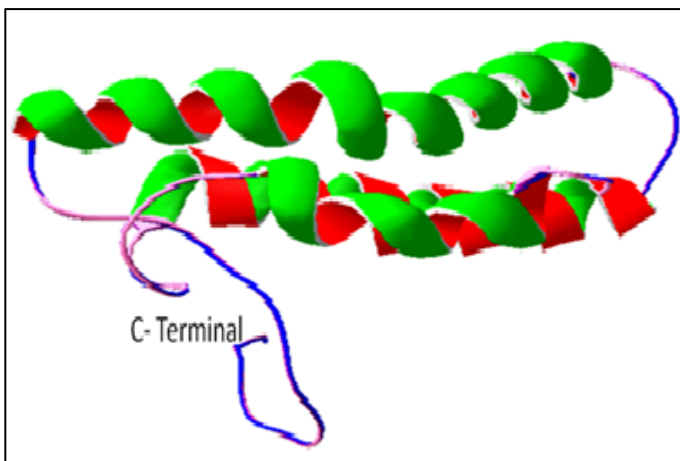

Figure 1: Structural model of one of the protein sequence of *M. canis* (sequence no. 11) obtained after homology modelling of the sequence by Modeller 9v6 indicating the presence of α-helices (red and green coloured) and coils (blue and pink coloured). The pictorial form was produced using SPDB Viewer, in which only C-Terminal can be viewed at this rotation of sequence.

The results obtained from ligsite were then compared with GOR V and MSA to identify the required conserved coil functional residues that are present near a pocket **(Table 2).** The results indicate the presence of functional coil residues ranging from 2-13 amino acids near the pockets, in the selected sequences of *Microsporum canis, Trichophyton mentagrophytes, Trichophyton rubrum* and *Epidermophyton floccosum.* However in *Epidermophyton floccosum,* functional residues could be identified in three of the chosen five sequences. These functional residues are likely to be responsible for the pathogenecity caused by these genera. The predominance of *Trichophyton* and *Microsporum* species in causing dermatophytosis is consistent with our results, as more functional residues have been identified in these fungi than in *Epidermophyton floccosum* **[20].** Previous reports support our hypothesis of relationship between the pathogenecity and occurrence of conserved coil functional residues.
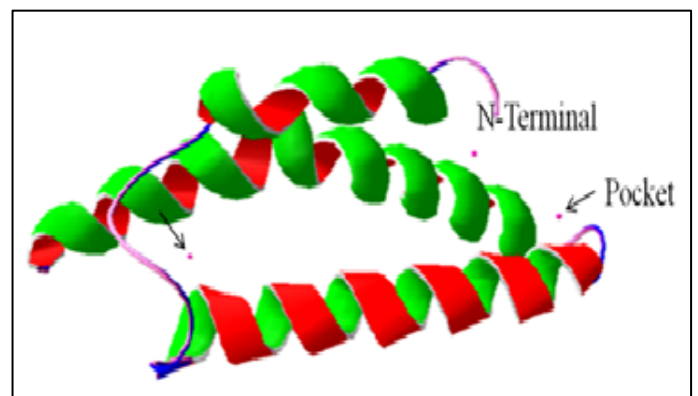

Figure 2: Structural model of one of the protein sequence of *M. canis* (sequence no. 11) indicating the presence of α-helices (red and green coloured), coils between the helices (blue and pink coloured) and pockets (pink coloured) that are present near the coil. The pictorial form was produced using SPDB Viewer, in which only N- Terminal can be viewed at this particular rotation.

Conclusion:
The conserved coil residues identified in the present study provide an insight into the protein functionality of three main

genera of Dermatophytes. The residues identified in the selected protein sequences from the mitochondrial genome are likely to be responsible for causing infections such as Tinea pedis, Tinea unguium, Tinea cruris, Tinea corporis and Tinea capitis. Therefore these sites can be the potential targets for drug designing and delivery.

References:
[1] White TC *et al. Eukaryotic Cell* 2008 **7:** 1238. [PMCID: PMC251977]
[2] Hanumanthappa H *et al. Indian J Dermatol.* 2012 **57:** 322. [PMID**:** 22837576]
[3] Mathur M *et al. Kathmandu Univ Med J.* 2012 **37:** 30. [PMID: 22971858]
[4] Akcaglar S *et al. Med Mycol*. 2011 **49:** 602. [PMID:21198349]
[5] Adefemi SA *et al. Niger J Clin Pract.* 2011 **14:**23. [PMID**:** 21493987]
[6] Martinez DA *et al. mBio.* 2012 **3**: e00259–12. [PMID: 22951933]
[7] Lund A & Deboer DJ, *J of Mycopathologia* 2008 **166:** 407. [PMID: 18478355]
[8] Seyfarth F *et al. Mycoses* 2007 **2:** 26. [PMID: 17681051]
[9] Vasconcellos C *et al. An Bras Dermatol*. 2013 **88:** 377. [PMICD: PMC3754368]
[10] Muller A *et al. Can Vet J.* 2011 **52:** 385. [PMID:21731090]
[11] Narang K et al. *Indian J Dermatol.* 2012 **57:**288. [PMID: 22837564]
[12] Yenisehirli G *et al. J of Mycopathologia* 2012 **174:** 327.[PMID:22576940]
[13] Wu Y *et al. BMC Genomics* 2009 **10:** 238. [PMID: 19457268]
[14] Jeon J *et al. PLoS Comput Biol.* 2009 **5:** e1000522. [PMID:19798434]
[15] http://www.ncbi.nlm.nih.gov/gorf/gorf.htm
[16] http://www.ebi.ac.uk/clustalw
[17] gor.bb.iastate.edu
[18] Ashford P *et al. BMC Bioinformatics* 2012 **13:**39. [PMID: 22417279]
[19] Dai T *et al. BMC Bioinformatics* 2011 **12:** S9. [PMID: 22373099]
[20] Surendran KAK *et al. Indian J Dermatol.* 2014 **59:** 262. [PMCID: PMC4037947]

Edited by P Kangueane

Citation: Gupta & Kaur, Bioinformation 12(3): 197-201 (2016)

BIOMEDICAL
INFORMATICS

BIOMEDICAL
©2016
INFORMATICS