

UNIVERSIDADE DE RIBEIRÃO PRETO - UNAERP
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA
DOUTORADO EM BIOTECNOLOGIA

MATHEUS ELOY FRANCO

IDENTIFICAÇÃO E ANÁLISE *IN SILICO* DE REPETIÇÕES DE DNA EM
FUNGOS DERMATÓFITOS ENVOLVIDOS COM PATOGENICIDADE

Ribeirão Preto - SP

2015

Matheus Eloy Franco

IDENTIFICAÇÃO E ANÁLISE *IN SILICO* DE REPETIÇÕES DE DNA EM
FUNGOS DERMATÓFITOS ENVOLVIDOS COM PATOGENICIDADE

Tese apresentada ao programa de pós-graduação
em Biotecnologia da Universidade de Ribeirão
Preto para obtenção do Título de Doutor em
Biotecnologia.

Orientadora: Prof^a Dr^a Ana Lucia Fachin
Saltoratto

Co-Orientador: Prof^o Dr Mozart de Azevedo
Marins

Ribeirão Preto - SP

2015

MATHEUS ELOY FRANCO

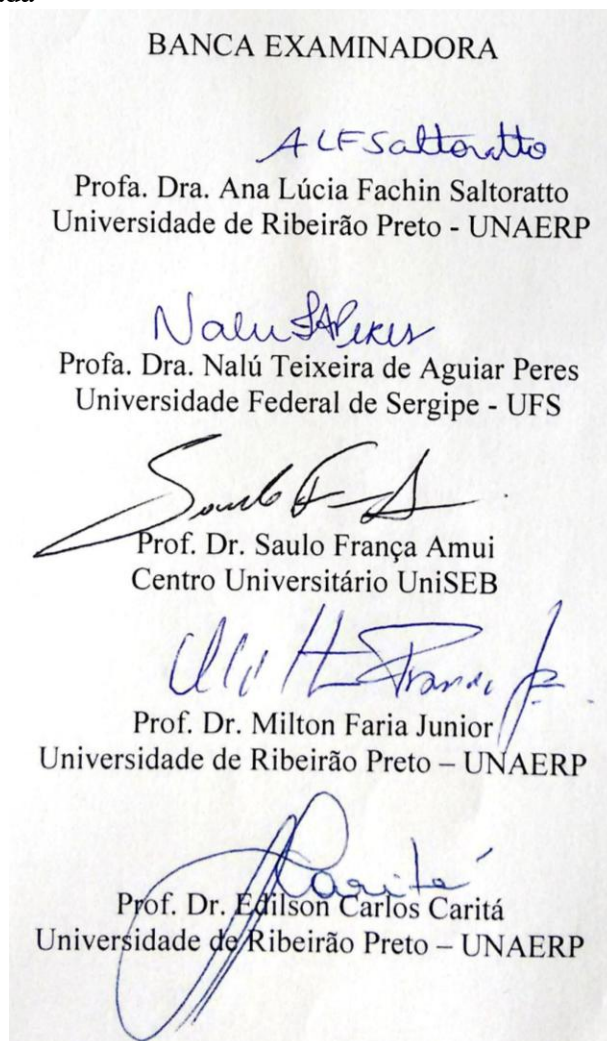
**IDENTIFICAÇÃO E ANÁLISE *IN SILICO* DE REPETIÇÕES DE DNA EM FUNGOS
DERMATÓFITOS ENVOLVIDOS COM PATOGENICIDADE**

Tese de Doutorado apresentada ao Programa de Pós-graduação
em Biotecnologia da Universidade de Ribeirão Preto, para
obtenção do título de Doutor em Biotecnologia.

Área de Concentração: Biotecnologia Aplicada à Saúde

Data da defesa: 07 de agosto de 2015-08-19

Resultado: Aprovada



**RIBEIRÃO PRETO - SP
2015**

Ficha catalográfica preparada pelo Centro de Processamento Técnico
da Biblioteca Central da UNAERP

- Universidade de Ribeirão Preto -

F825i Franco, Matheus Eloy, 1983-
Identificação e análise in silico de repetições de DNA em
fungos dermatófitos envolvidos com patogenicidade / Matheus
Eloy Franco. - - Ribeirão Preto, 2015.
114 f.: il. color.

Orientadora: Prof^a. Dr^a. Ana Lucia Fachin Saltoratto.

Tese (doutorado) - Universidade de Ribeirão Preto, UNAERP,
Biotecnologia. Ribeirão Preto, 2015.

1. Fungos dermatófitos. 2. *Trichophyton rubrum*. I. Título.

CDD 660

AGRADECIMENTOS

Agradeço a Deus pela vida e força concedida.

A minha família por todo apoio e compreensão, em especial, a minha esposa Daguiamar e minha filha Ana Laura.

À minha orientadora, Dr^a Ana Lúcia Fachin Saltoratto pelos ensinamentos, dedicação e paciência de se dispor a orientar um aluno de outra área de atuação.

Ao meu co-orientador, Dr. Mozart Marins de Azevedo, pelos ensinamentos e apoio para execução deste trabalho.

A minha amiga Tamires Aparecida Bitencourt pelos momentos de companheirismo e ajuda que foram essenciais. Ao meu amigo Pablo Sanches pelos ensinamentos na área de bioinformática e pelo tempo dispensado a mim.

Ao Professor Dr. Kevin Verstrepen pelos direcionamentos sobre análise de repetições adjacentes.

A todos meus colegas de pós-graduação, aos amigos do IFSULDEMINAS - *Campus* Machado que trilharam parte desta caminhada comigo, em especial, ao amigo Prof. Luciano Pereira Carvalho.

Ao IFSULDEMINAS - *Campus* Machado pelo incentivo a qualificação concedido.

A CAPES pela concessão de bolsa taxa como suporte financeiro.

A FAPESP pela bolsa concedida à Tamires Bitencourt e pelo projeto auxílio concedido aos experimentos de *microarray*.

E a todos que direta ou indiretamente me incentivaram e auxiliaram nesta etapa tão importante da minha vida.

Obrigado.

“Alguns nunca tentam, outros nunca terminam, os vencedores nunca desistem.”

Norman Vincent Peale

RESUMO

As dermatofitoses são as infecções fúngicas mais comuns no mundo, sendo caracterizada por lesões superficiais em pele, cabelo e unhas. As infecções causadas pelos dermatófitos são de grande importância clínica, pois também acometem indivíduos imunocomprometidos podendo causar infecções invasivas. O genoma de diferentes espécies e linhagens de dermatófitos foi recentemente sequenciado, neste cenário, a identificação e análise de repetições adjacentes tornam-se importante, pois elas contribuem significativamente para a variação genética e para patogênese de fungos. Dessa forma, este trabalho teve como objetivo a identificação e análise das regiões de repetições adjacentes em genes alvo que possam estar relacionados à patogenicidade e com a interação fungo-hospedeiro em dermatófitos. As sequências utilizadas foram obtidas através do domínio público do *Broad Institute* que compila os dados do Projeto Genoma Comparativo Dermatófitos. Um *pipeline* foi implementado utilizando tecnologias estabelecidas (Linux, HTML, CSS, Asp.NET) para automatização dos processos de identificação, armazenamento e consulta das repetições identificadas. A partir dos dados obtidos, foi possível se traçar o perfil das repetições adjacentes em dermatófitos, bem como comparar a variabilidade das repetições entre espécies e linhagens. A partir das comparações, pôde-se observar que alguns genes relacionados à adaptação ao ambiente e hospedeiro possuem alto padrão de variabilidade em suas repetições. Ainda, observou-se que repetições adjacentes com maior variabilidade são prevalentemente encontradas em genes envolvidos com a regulação da expressão ou, potencialmente envolvidos com fatores de virulência. A partir da análise *in silico* foram identificados genes ricos em repetições adjacentes variáveis potencialmente envolvidos com a adesão em fungos, especialmente o gene TERG_08771 de *Trichophyton rubrum* que contém sequências características de adesinas. Um experimento de *microarray* do nosso grupo de pesquisa comprovou a indução da expressão *in vitro* desse gene quando *T. rubrum* foi cultivado em meio de cultura contendo queratina. O gene TERG_08771 também foi induzido quando *T. rubrum* foi co-cultivado na presença de uma linhagem de queratinócitos humanos. Dessa forma esse gene se mostrou um importante alvo molecular a ser investigado, pois pode estar relacionado a adesão e patogenicidade em *T. rubrum*. Os resultados obtidos contribuem para anotação do genoma de *T. rubrum*, como também para um melhor entendimento dos mecanismos moleculares envolvidos com a patogenicidade em dermatófitos.

Palavras-chave: Fungos dermatófitos, *Trichophyton rubrum*, repetições adjacentes

ABSTRACT

Dermatophytoses are the most common fungal infections in the world, characterized by superficial lesions in skin, hair and nails. Infections caused by dermatophytes have great clinical importance because it also affect immunocompromised individuals, may cause invasive infections. The genomes of different species and strains of dermatophytes was recently sequenced, in this scenario, the identification and analysis of tandem repeats are important because these repeats contribute significantly to genetic variation and fungal pathogenesis. Thus, the aim of study was identify and analyze the tandem repeats in target genes that may be related to pathogenicity and dermatophyte-host interaction. Data was retrieved from public domain of Broad Institute that storage Dermatophyte Comparative Database. A pipeline was developed using established technologies (Linux, HTML, CSS, Asp.NET) to process automation of identification, storage and query tandem repeats. From the data obtained, was possible generate the profile of tandem repeats in dermatophytes and compare the variability between species and strains. Through the comparisons, it could be observed that some genes related to adaptation to environment and host have high standard of variability in their repetitions. Furthermore, it was observed that tandem repeats with higher variability are predominantly found in genes involved in regulation of expression or potentially involved in virulence factors. *In silico* gene analysis identified tandem repeats potentially involved in fungal adhesion, especially TERG_08771 gene (*T. rubrum* CBS strain) containing sequences characteristics of adhesins. Microarray experiment of our research group demonstrated the induction of this gene *in vitro* when *T. rubrum* was grown in medium containing keratin substrate. The gene TERG_08771 also was induced when *T. rubrum* was co-cultured in the presence of human keratinocytes cell line. Thus, this gene proved to be an important molecular target to be investigated, as it may be related to adhesion and pathogenicity in *T. rubrum*. The results contribute to annotation of *T. rubrum* genome, and also to a better understanding of the molecular mechanisms involved in pathogenicity of dermatophytes.

Keywords: Dermatophyte Fungi, *Trichophyton rubrum*, tandem repeats

LISTA DE FIGURAS

Figura 1- Genes ortólogos em fungos dermatófitos	22
Figura 2- Estrutura genérica de adesinas fúngicas	26
Figura 3- Repetições adjacentes em regiões codificantes modulam a adesão celular.....	27
Figura 4- Classificação de DNA repetitivo	28
Figura 5- Características gerais das repetições adjacentes	29
Figura 6- Mecanismos de expansão e contração de repetições adjacentes.....	30
Figura 7- Repetições adjacentes no gene <i>Runx-2</i> promovem uma rápida evolução em cães...35	
Figura 8- Repetições adjacentes como mecanismo de ajuste para ciclos circadianos.....	37
Figura 9- Esquema de busca e armazenamento das repetições adjacentes.....	45
Figura 10- Diagrama de caso de uso do sistema desenvolvido	46
Figura 11- Interface inicial do sistema	51
Figura 12- Configurações do sistema	51
Figura 13- Envio e processamento das repetições.....	52
Figura 14- Interface do sistema para envio e processamento das sequências.	53
Figura 15- Perfil de repetições adjacentes de um organismo selecionado.	53
Figura 16- Listagem com repetições mais variáveis de um organismo.....	54
Figura 17- Consulta e filtro de repetições adjacentes.....	54
Figura 18- Repetições adjacentes encontradas em diferentes sequências	55
Figura 19- Consulta categorias funcionais dos genes.....	55
Figura 20- Abundância relativa das repetições agrupadas	59
Figura 21- Repetições adjacentes em <i>T. rubrum</i>	61
Figura 22- Classificação funcional dos genes transcritos com unidades de repetições adjacentes variáveis maiores que 10-pb de acordo com a tabela Functat.....	62
Figura 23- Repetição adjacente variável do gene TERG_08771. (A) Aminoácidos codificados pela sequência repetitiva. (B) Variabilidade da extensão da repetição dentre diferentes espécies e linhagens de dermatófitos.....	73
Figura 24- Domínios presentes na repetição do gene TERG_08771	73
Figura 25- Proteína resultante do gene TERG_08771. (A) Alinhamento com outras adesinas descritas. (B) Região homóloga com <i>Mad1</i> de <i>Metarhizium anisopliae</i> . (C) Região repetitiva homóloga com <i>cspA</i> de <i>Aspergillus fumigatus</i>	75
Figura 26- Via de interação de proteínas rica em poli-Q	77
Figura 27- Categorias funcionais de genes com repetições adjacentes variáveis.....	79

Figura 28- Repetições adjacentes em regiões <i>upstream</i> dos genes	93
Figura 29- Genes anotados como <i>extracellular serine-threonine rich protein</i> . (A) Gene de <i>T. rubrum</i> alinhado com <i>Adhesin-like protein</i> [<i>Metarhizium anisopliae</i>] e <i>GPI-anchored protein</i> [<i>Aspergillus niger</i>]. (B) Gene de <i>M. canis</i> alinhado com <i>Adhesin-like protein</i> [<i>Metarhizium anisopliae</i>] e <i>GPI-anchored protein</i> [<i>Aspergillus niger</i>] (C) Região repetitiva rica em glicina do gene MCYG_04810.....	94
Figura 30- Sequência de aminoácidos proveniente da repetição adjacente do gene TESHG_04591	95
Figura 31- Sequência de aminoácidos proveniente da repetição adjacente do gene TEQG_03078.	96
Figura 32- Sequência de aminoácidos proveniente da repetição adjacente do gene MGYG_07037.....	96
Figura 33- Sequência de aminoácidos proveniente da repetição adjacente do gene TRV_07130	97

LISTA DE TABELAS

Tabela 1- Dados de isolados de <i>T. rubrum</i> no banco de dados TrED	20
Tabela 2- Características gerais dos genomas de fungos dermatófitos	21
Tabela 3- Proteínas/Genes potencialmente envolvidas na virulência de dermatófitos	24
Tabela 4- Ferramentas para detecção de <i>tandem repeats</i>	43
Tabela 5- Perfil de repetições adjacentes em genes transcritos de fungos dermatófitos	56
Tabela 6- Ocorrência das unidades de repetições adjacentes e abundância relativa	57
Tabela 7- Padrões de repetição consenso mais prevalentes	59
Tabela 8- Categorias e genes com repetições de minissatélites de maior interesse	62
Tabela 9- Repetições adjacentes variáveis com unidade ≥ 10 -pb em <i>T. rubrum</i>	63
Tabela 10- Anotação dos genes transcritos com repetições de variáveis de minissatélites.	66
Tabela 11- Genes com domínios repetitivos em <i>T. rubrum</i>	69
Tabela 12- Regiões repetitivas variáveis em genes modulados	71
Tabela 13- Genes ortólogos de <i>T. rubrum</i> : TERG_08771	72
Tabela 14- Genes ortólogos de <i>T. rubrum</i> : TERG_05576	80
Tabela 15- Genes ortólogos de <i>T. rubrum</i> : TERG_00768	82
Tabela 16- Genes ortólogos de <i>T. rubrum</i> : TERG_07370	84
Tabela 17- Genes ortólogos de <i>T. rubrum</i> : TERG_03736	86
Tabela 18- Genes ortólogos de <i>T. rubrum</i> : TERG_05189	88
Tabela 19- Genes ortólogos de <i>T. rubrum</i> : TERG_08141	90
Tabela 20- Genes ortólogos de <i>T. rubrum</i> : TERG_01042	92

LISTA DE ABREVIATURAS

BLAST	Basic Local Alignment Search Tool
Mb	Mega bases - corresponde a um milhão de pares de base
NCBI	National Center for Biotechnology Information
pb	Pares de base
PCR	Reação em Cadeia da Polimerase
RNA	Ácido Ribonucleico
RNA-seq	Tecnologia de sequenciamento de nova geração para revelar o perfil transcricional de RNA a partir de um genoma em um determinado momento ou condição
SNP	Single Nucleotide Polymorphism - Polimorfismo de base única
TR	Tandem Repeats - Repetições adjacentes
TRF	Tandem Repeat Finder
VNTR	Variable number of tandem repeats - Número variável de repetições adjacentes

SUMÁRIO

I. INTRODUÇÃO	15
1.1. FUNGOS DERMATÓFITOS	16
1.2. <i>TRICHOPHYTON RUBRUM</i>	18
1.3. A GENÔMICA DOS DERMATÓFITOS	20
1.4. GENES RELACIONADOS À PATOGENICIDADE EM FUNGOS DERMATÓFITOS	23
1.5. PROTEÍNAS DE ADESÃO E REPETIÇÕES ADJACENTES	25
1.6. CONCEITOS BÁSICOS SOBRE REPETIÇÕES ADJACENTES	28
1.7. REPETIÇÕES ADJACENTES	30
1.8. MICROSSATÉLITES E MINISSATÉLITES	32
1.9. REPETIÇÕES ADJACENTES VARIÁVEIS	34
1.10. FERRAMENTAS PARA IDENTIFICAÇÃO DE REPETIÇÕES ADJACENTES	38
1.10.1. TANDEM REPEAT FINDER (TRF)	40
1.10.2. MREPS	40
II. HIPÓTESE.....	42
III. OBJETIVOS	42
3.1. OBJETIVO GERAL	42
3.2. OBJETIVOS ESPECÍFICOS	42
IV. METODOLOGIA.....	43
4.1. LEVANTAMENTO DAS FERRAMENTAS/ALGORITMOS PARA BUSCA DE REPETIÇÕES ADJACENTES	43
4.2. COMPARAÇÃO DAS FERRAMENTAS/ALGORITMOS PARA BUSCA DE REPETIÇÕES ADJACENTES	44
4.3. OBTENÇÃO DAS SEQUÊNCIAS	44
4.4. METODOLOGIA DE BUSCA E ARMAZENAMENTO DE REPETIÇÕES ADJACENTES	44
4.5. DESENVOLVIMENTO DO SISTEMA PARA BUSCA E ARMAZENAMENTO DE REPETIÇÕES ADJACENTES	46
4.6. PARÂMETROS UTILIZADOS NA BUSCA POR REPETIÇÕES ADJACENTES	46
4.7. VARIABILIDADE DAS REPETIÇÕES ADJACENTES	47
4.8. CONSERVAÇÃO DAS REPETIÇÕES	48
4.9. TENTATIVA DE ANOTAÇÃO DAS SEQUÊNCIAS	48
4.10. IDENTIFICAÇÃO DE REGIÕES REPETITIVAS EM GENES DE <i>TRICHOPHYTON RUBRUM</i> MODULADOS EM MEIO DE CULTURA QUE MIMETIZA A INFECÇÃO DA PELE HUMANA PELA TÉCNICA DE MICROARRAY	48
4.11. PREDIÇÃO DE ADESINAS	48
V. RESULTADOS E DISCUSSÃO	50
5.1. FERRAMENTAS PARA BUSCA DE REPETIÇÕES ADJACENTES	50

5.2.	SISTEMA PARA ARMAZENAMENTO E ANÁLISE DE REPETIÇÕES ADJACENTES EM FUNGOS DERMATÓFITOS.....	50
5.3.	CONSULTA POR CATEGORIAS FUNCIONAIS DOS GENES DE <i>T. RUBRUM</i> ...	55
5.4.	PERFIL DE REPETIÇÕES ADJACENTES EM FUNGOS DERMATÓFITOS.....	56
5.5.	REPETIÇÕES ADJACENTES NO DERMATÓFITO <i>T. RUBRUM</i>	61
5.6.	CATEGORIZAÇÃO FUNCIONAL DAS REPETIÇÕES ADJACENTES NO DERMATÓFITO <i>T. RUBRUM</i>	61
5.7.	DOMÍNIOS REPETITIVOS EM <i>T. RUBRUM</i>	68
5.8.	IDENTIFICAÇÃO DE REGIÕES REPETITIVAS EM GENES DE <i>TRICHOPHYTON RUBRUM</i> MODULADOS EM MEIO DE CULTURA QUE MIMETIZA A INFECÇÃO DA PELE HUMANA PELA TÉCNICA DE MICROARRAY	70
5.9.	ADESÃO CELULAR E REPETIÇÕES ADJACENTES VARIÁVEIS EM <i>T. RUBRUM</i>	75
5.10.	REPETIÇÕES ADJACENTES EM SISTEMAS DE TRANSDUÇÃO DE SINAL E REGULAÇÃO DE TRANSCRIÇÃO EM <i>T. RUBRUM</i>	76
5.11.	COMPARAÇÃO DOS GENES DE <i>TRICHOPHYTON RUBRUM</i> CONTENDO REPETIÇÕES ADJACENTES VARIÁVEIS COM OUTROS DERMATÓFITOS.....	78
5.12.	GENE TERG_05576	79
5.13.	GENE TERG_00768	81
5.14.	GENE TERG_07370	83
5.15.	GENE TERG_03736	85
5.16.	GENE TERG_05189	87
5.17.	GENE TERG_08141	89
5.18.	GENE TERG_01042	90
5.19.	REPETIÇÕES ADJACENTES VARIÁVEIS EM REGIÕES <i>UPSTREAM</i> DE <i>TRICHOPHYTON RUBRUM</i>	93
5.20.	GENES COM REPETIÇÕES VARIÁVEIS EM OUTRAS ESPÉCIES DE FUNGOS DERMATÓFITOS	93
5.20.1.	<i>MICROSPORUM CANIS</i>	93
5.20.2.	<i>T. TONSURANS</i>	95
5.20.3.	<i>T. EQUINUM</i>	95
5.20.4.	<i>M. GYPSEUM</i>	96
5.20.5.	<i>T. TONSURANS</i>	96
5.20.6.	<i>A. BENHAMIAE</i>	97
VI.	CONCLUSÕES.....	98
	REFERÊNCIAS	100
	APÊNDICE A - DIAGRAMA ENTIDADE RELACIONAMENTO DO BANCO DE DADOS	108
	APÊNDICE B - LISTA DE GENES COM REPETIÇÕES ADJACENTES VARIÁVEIS (VARSCORE >0) COM PREDIÇÃO DE ADESÃO CELULAR.....	109

I. INTRODUÇÃO

As dermatofitoses são infecções provocadas por fungos dermatófitos que constituem um grupo de fungos filamentosos capazes de invadir e colonizar tecidos queratinizados em humanos e animais, sendo a infecção fúngica mais comum no mundo (GARBER, 2011). São especializados em infecções de tecidos queratinizados como unhas, pele e cabelo; podem também ser divididos de acordo com seu habitat de escolha como geofílicos, zoofílicos e antropofílicos. As micoses provocadas são denominadas tineas, sendo normalmente seguidas pela localização específica da infecção fúngica, possuindo importância tanto para a medicina humana quanto veterinária (WEITZMAN; SUMMERBELL, 1995).

As pesquisas sobre a genômica de dermatófitos é relativamente nova, pois até abril de 2007 apenas 524 seqüências de nucleotídeos deste gênero eram encontradas no GenBank (YANG et al., 2007). Porém, em 2006 foi anunciado pelo *National Human Genome Research Institute* o início de um projeto para o sequenciamento do genoma de cinco dermatófitos, incluindo a espécie *Trichophyton rubrum* (NHGRI, 2006). Os avanços nas tecnologias de sequenciamento veêm mudando o ritmo de projetos de genoma e têm contribuído para o crescente volume de dados genômicos. Isso tem sido acompanhado pelo aumento na capacidade dos recursos computacionais para processar e traduzir os dados brutos em informações significativas.

Atualmente, o genoma dos dermatófitos está completo com boa cobertura, mas ainda pode haver alguns ajustes, principalmente de anotação (Dr^a Nilce Martinez-Rossi, comunicação pessoal). O sequenciamento do genoma das espécies de dermatófitos selecionados têm permitido diferentes estudos sobre seus determinantes de virulência, como os fatores envolvidos nas interações fungo-hospedeiro.

Com o genoma dos principais dermatófitos disponível, diferentes estudos podem ser realizados, como o de repetições adjacentes que constituem uma parte importante de todos os genomas. De acordo com Boeynaems (2012), a ocorrência destas repetições em regiões codificantes e promotoras não é aleatória: genes que contêm repetições em série são enriquecidos em classes funcionais específicas. Estas repetições adjacentes (*tandem repeats*) são seqüências repetitivas que não ocorrem apenas sequencialmente, mas também diretamente adjacente, e de acordo com Levdansky et al. (2010) mostram estar relacionadas à patogenicidade em diferentes microrganismos.

Desta maneira, a identificação e análise de repetições adjacentes no genoma de fungos dermatófitos pode contribuir para compreensão das bases moleculares relacionados à sua virulência. Estas análises são importantes, pois as estratégias para controlar esse tipo de infecção apresentam limitações como o número restrito de alvos celulares disponíveis (BALDO et al., 2012).

Assim este trabalho teve como objetivo realizar o levantamento de ferramentas de bioinformática para identificar e analisar de repetições adjacentes, desenvolver um *pipeline* para automatização dos processos de identificação, armazenamento e consultas das repetições, bem como traçar um perfil das repetições relacionando-as à *motifs* funcionais destes importantes fungos patogênicos.

1.1. FUNGOS DERMATÓFITOS

Fungos dermatófitos são classificados em três gêneros de acordo com a formação e morfologia dos micro ou macroconídios: *Trichophyton*, *Microsporum* e *Epidermophyton*. As espécies de dermatófitos são divididas em geofílicas (solo), antropofílicas (humanos) e zoofílicas (animais) dependendo de seu habitat principal (WEITZMAN; SUMMERBELL, 1995). As espécies zoofílicas representam cerca de 30% das patologias causadas por dermatófitos em humanos, enquanto que as espécies antropofílicas são responsáveis por cerca de 70% das infecções nesses hospedeiros, causando infecções recorrentes e de progressão lenta. Cerca de 30 espécies de dermatófitos estão identificadas como patógenos humanos (WHITE et al., 2008).

Fungos dermatófitos são a causa mais comum de infecções fúngicas em todo o mundo, acometendo milhões de pessoas anualmente. Nos Estados Unidos, isso se traduz em um impacto econômico que se estima ultrapassar US\$ 400 milhões por ano somente para seu tratamento (SMITH; FLEISCHER; FELDMAN, 1998). As infecções causadas pelos dermatófitos são de grande importância clínica, pois acometem indivíduos imunocomprometidos podendo causar infecções invasivas ou externas atingindo tecidos queratinizados, como pele, pelos e cabelos. As infecções podem ser assintomáticas, leves, crônicas ou graves, causando normalmente manchas na pele e deformações nas unhas (WHITE et al., 2008).

A dermatofitose pode variar de crônica para altamente inflamatória, dependendo do agente causador e a localização do corpo afetada. A doença é descrita com a palavra "tinea",

seguido de um termo referindo-se ao local do corpo infectado, por exemplo: *tinea pedis* (pés), *tinea capitis* (couro cabeludo ou na cabeça), *tinea corporis* (corpo ou tronco) e *tinea unguium* (unhas, também chamada de onicomicose). As principais espécies antropofílicas proeminentes, *Trichophyton rubrum*, *Trichophyton mentagrophytes*, *Trichophyton interdigitale* e *Trichophyton tonsurans*, são na sua maioria associadas com infecções crônicas, menos inflamatórias. Em contraste, as espécies zoofílicas, como, *Microsporum canis*, *Arthroderma benhamiae*, *Arthroderma vanbreeu-seghemii*, *Trichophyton verrucosum* e *Trichophyton erinacei*, bem como dermatófitos geofílicos como *Microsporum gypseum*, frequentemente causam infecções inflamatórias em humanos (DEGREEF, 2008).

Embora existam diversas terapias antifúngicas, as dermatofitoses, particularmente as infecções de unhas são de difícil tratamento, sendo que elas apresentam uma taxa de reinfecção ou reativação de 25 a 40% (HAY, 2001). Por este motivo, diversos estudos vêm sendo realizados com dermatófitos para melhor compreender sobre a epidemiologia dessas infecções.

Um levantamento realizado na Índia mostrou que de 165 isolados de infecções micóticas, 48,5% são causadas por dermatófitos, no qual *T. rubrum* é o mais prevalente em infecções de pele, seguido por *Trichophyton mentagrophytes*. Já em infecções causadas no couro cabeludo/cabelo *Trichophyton violaceum* é o agente mais comum (KANNAN; JANAKI; SELVI, 2006).

Nos Estados Unidos, *Trichophyton tonsurans* é um patógeno comum de *tinea capitis* em crianças e adultos (especialmente entre os afro-americanos e hispânicos). Surto de *tinea corporis* e *tinea capitis gladiatorum* ocorrem devido as infecções causadas por *T. tonsurans* rotineiramente em crianças e adolescentes que participam de esportes de combate. Na Alemanha, *tinea capitis* é causada principalmente por dermatófitos zoofílicos, em particular *Microsporum canis*. Fungos zoofílicos, principalmente da espécie *Arthroderma benhamiae*, devem ser levados em considerações no diagnóstico diferencial de *tinea capitis*, *tinea faciei*, e *tinea corporis*, pois podem causar infecções graves em crianças/adolescentes e pacientes imunodeprimidos (NENOFF et al., 2014).

No Brasil, *T. rubrum* e *T. mentagrophytes* são os dermatófitos mais comumente isolados, sendo que nas regiões Sul e Sudeste há um predomínio de *T. rubrum*, *M. canis* e *T. mentagrophytes* (AQUINO; CONSTANTE; BAKOS, 2007). No estado de São Paulo, um estudo realizado coletou 280 amostras de unhas ou pele das regiões dos pés ou mãos em

estudantes universitários obtendo a ocorrência de dermatófitos em 18,2% da população estudada. O único gênero isolado foi *Trichophyton*, apresentando predominância da espécie *T. rubrum* (80%), seguida de *T. mentagrophytes* (20%) (SIQUEIRA et al., 2006). Já o levantamento realizado por Schoeler et al. (2010) no estado de Santa Catarina apontou a prevalência de *T. mentagrophytes*, seguido por *T. rubrum*, mesmo resultado obtido em um levantamento realizado no estado do Pará por Silveira-Gomes et al. (2013), que também apontou maior incidência de *T. mentagrophytes*, seguido por *T. rubrum*.

Novas abordagens genéticas têm permitido o conhecimento sobre as bases moleculares em quase todas as áreas de investigação da patogênese microbiana (MCMANUS; COLEMAN, 2014). No entanto, tais conhecimentos em dermatófitos ainda não foram estabelecidos, em contraste com outros fungos patogênicos clinicamente importantes, como *Candida albicans*, *Aspergillus fumigatus* ou *Cryptococcus neoformans*. Por conseguinte, o conhecimento sobre a patogenicidade de dermatófitos em nível molecular é limitada (GRUMBT; MONOD; STAIB, 2011).

Tendo em vista o sequenciamento e a anotação do genoma das principais espécies de fungos dermatófitos (MARTINEZ et al., 2012), além dos recentes avanços em técnicas para manipulação genética destes fungos (PERSINOTI et al., 2014), espera-se elucidar os mecanismos envolvidos em sua virulência. Assim, novos estudos são necessários para prover um melhor entendimento de suas bases moleculares, em especial da espécie *Trichophyton rubrum*, pois de acordo com Leng et al. (2008) este é o maior causador de dermatofitoses em humanos.

1.2. *Trichophyton rubrum*

Trichophyton rubrum é um fungo dermatófito patogênico filamentosos, de espécie antropofílica responsável por cerca de 70% das dermatofitoses em humanos (LENG et al., 2008). De acordo com Marconi et al. (2010), um fator agravante das infecções causadas por este dermatófito é que em pacientes imunocomprometidos (pacientes submetidos a transplantes de órgãos, em tratamento quimioterápico ou pacientes com síndrome da imunodeficiência adquirida), *T. rubrum* pode causar infecções invasivas podendo tornar-se profundas e generalizadas.

Desta maneira, evidencia-se a importância clínica deste dermatófito, porém o conhecimento de suas bases moleculares e mecanismos de invasão ainda são limitados

(BALDO et al., 2012), devido ao fato das dificuldades em sua manipulação genética e de sua natureza pouco agressiva e não fatais em infecções superficiais. Assim, pesquisas acerca de sua interação com o hospedeiro e expressão gênica são necessárias para fornecer perspectivas sobre os mecanismos moleculares deste patógeno (KAUFMAN et al., 2005).

Para se compreender os aspectos da biologia desse fungo, diferentes estudos foram realizados como o da expressão gênica a partir da construção e do sequenciamento de bibliotecas de cDNA. O trabalho desenvolvido por Wang et al. (2006) sequenciou e categorizou 11.085 ESTs de *T. rubrum* revelando genes envolvidos em diferentes processos celulares, permitindo um melhor conhecimento das atividades fisiológicas como seu metabolismo, ciclo celular, resistência a drogas, entre outros. Outros trabalhos como o de Paião et al. (2007) e Maranhão et al. (2009) possibilitaram um melhor entendimento dos mecanismos de ação de drogas e fatores de virulência deste fungo.

Em comparação a outros fungos, dermatófitos mostram-se menos susceptíveis a manipulação genética. Como consequência, a mutagênese dirigida em espécies de dermatófitos foi evidenciada apenas em um número muito limitado de casos. O trabalho realizado por Fachin et al. (2006) demonstraram que é possível a transformação para estudos genéticos do dermatófito *T. rubrum*, no qual os autores realizaram o rompimento do gene *TruMDR2*, que codifica uma proteína de membrana da família dos transportadores ABC relacionado a fatores de virulência. Já Ferreira-Nozawa et al. (2006) realizaram a disrupção do gene *pacC*, observando no mutante *pacC-1* a redução da secreção de proteases queratinolíticas e a capacidade de crescer em fragmentos de unhas humanas como a única fonte de nutrição.

Os trabalhos citados demonstram um esforço para o melhor entendimento das bases moleculares deste fungo, porém, com o objetivo de obter-se um entendimento mais abrangente, o projeto genoma comparativo de dermatófitos sequenciou inicialmente o genoma de cinco espécies de dermatófitos, entre elas *T. rubrum*. Posteriormente, foi sequenciada outra espécie (*T. interdigitale*) e diversas linhagens de *T. rubrum*. Os dados resultantes do projeto genoma comparativo de dermatófitos estão disponíveis para acesso, sendo possível realizar o *download* de arquivos contendo o genoma completo, genes, transcritos, 1.000 bases *upstream* e *downstream* dos genes preditos, dentre outros arquivos (BROAD INSTITUTE, 2014). Outro projeto independente realizado anteriormente sequenciou outras duas espécies de

dermatófitos, *Trichophyton verrucosum* e *Arthroderma benhamiae* (BURMESTER et al., 2011).

1.3. A GENÔMICA DOS DERMATÓFITOS

Dentre as pesquisas sobre a genômica de dermatófitos destaca-se o banco de dados de expressão gênica da espécie *Trichophyton rubrum* - TrED, este é construído sobre um banco de dados relacional¹, com uma interface *web* oferecendo funções analíticas e acesso integrado aos dados de expressão gênica de várias linhagens do dermatófito *T. rubrum*. O banco de dados integra dados genéticos, transcritos e metabólicos de *T. rubrum*, bem como resultados de genômica comparativa de dermatófitos. TrED armazena três tipos de dados: (i) sequências EST de *T. rubrum*, (ii) perfis de transcrição de *T. rubrum* e (iii) resultados de CGH de 22 dermatófitos (YANG et al., 2007). A Tabela 1 apresenta os dados das três linhagens de *T. rubrum* armazenados no banco de dados TrED.

Tabela 1- Dados de isolados de *T. rubrum* no banco de dados TrED

Isolado	Dados de		Quantidades		Autor
	ESTs	Microarray	de bibliotecas	de ESTs	
BMU01672	Sim	Sim	14	36992	<u>Jin Q</u>
CHUV862.00	Sim	Não	1	3562	<u>Zaugg C</u>
MYA-3108	Sim	Não	2	63	<u>Martinez-Rossi NM</u>

Fonte: (TRED, 2012)

Até fevereiro de 2012 este banco de dados contava com 40.617 sequências expressas de três isolados de *T. rubrum* (TRED, 2012), enquanto o GenBank contava com 9.131 sequências da espécie *T. rubrum* depositadas (NCBI, 2012). Em uma consulta ao site do NCBI (GenBank) em junho de 2015, o mesmo contava com 8967 genes e 47885 sequências expressas (ESTs) de *T. rubrum*.

O artigo de Martinez et al. (2012) relatou o sequenciamento do genoma completo de cinco espécies de fungos dermatófitos. Até então o conhecimento sobre o modo que estes agentes patogênicos causam doenças é insuficiente, talvez devido a falta de modelos e dados para a análise molecular destes organismos (GRUMBT; MONOD; STAIB, 2011). O sequenciamento do genoma dos dermatófitos está auxiliando no desenvolvimento de novas

¹ Conceito abstrato que define maneiras de armazenar, manipular e recuperar dados estruturados na forma de tabelas, construindo um banco de dados.

ferramentas para caracterização e entendimento sobre sua patogenicidade. As características gerais dos genomas sequenciados são apresentadas na Tabela 2.

Tabela 2- Características gerais dos genomas de fungos dermatófitos

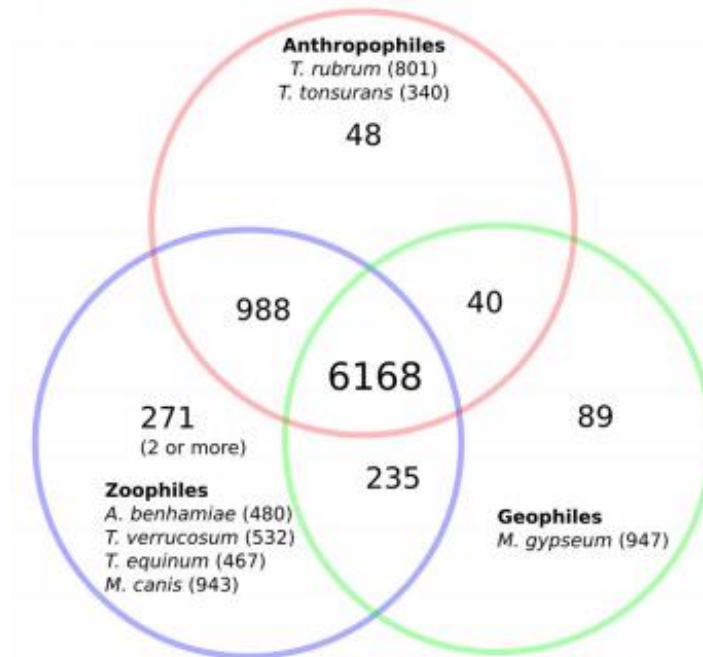
Características	<i>T. rubrum</i>	<i>T. tonsurans</i>	<i>T. equinum</i>	<i>M. canis</i>	<i>M. gypseum</i>	<i>T. verrucosum</i>	<i>A. benhamiae</i>
Prefixo do Gene	TERG	TESG	TEQG	MCYG	MGYG	TRV	ARB
Isolado	CBS118892	CBS112818	CBS127.97	CBS113480	CBS118893	HKI0517	CBS112371
Tamanho (Mb)	22.5	23.0	24.1	23.1	23.2	22.5	22.2
GC (%)	48,31	48,15	47,39	47,5	48,5	48,24	48,75
Nr. genes cod. preditos	8.707	8.523	8.679	8.915	8.907	8.024	7.980
Tam. médio das sequências cod.	1.393	1.409	1.371	1.459	1.436	1.468	1.483

Fonte: MARTINEZ et al. (2012, p. 2)

O genoma, bem como, algumas caracterizações das espécies *Trichophyton rubrum*, *Trichophyton tonsurans*, *Trichophyton equinum*, *Microsporum gypseum*, *Microsporum canis*, estão disponíveis na página da internet do *BROAD INSTITUTE*: http://www.broadinstitute.org/annotation/genome/dermatophyte_comparative. Os genomas foram sequenciados utilizando a tecnologia de Sanger com uma cobertura de 8 a 10x para *T. rubrum*, *M. gypseum* e *M. canis*, para *T. tonsurans* e *T. equinum* utilizou-se uma cobertura de 5 a 6x (MARTINEZ et al., 2012). Posteriormente, foram sequenciados os genomas de duas linhagens da espécie *Trichophyton interdigitale* e de outras dez linhagens de *T. rubrum* (dados não publicados - *Broad Institute*). Além dos genomas descritos anteriormente, também se encontram disponíveis nas análises comparativas as espécies *Arthroderma benhamiae* e *Trichophyton verrucosum* já sequenciadas em outro projeto prévio (BURMESTER et al., 2011).

Segundo Martinez et al. (2012), os genomas dos dermatófitos apresentam um elevado grau de conservação, pois, a média de identidade dos aminoácidos entre as proteínas de *T. rubrum* comparada com a de cada um dos outros dermatófitos vai de 79% para *M. canis* e *M. gypseum* (cobrindo 89% das proteínas de *T. rubrum*) a elevados 97% de *T. verrucosum* (cobrindo 85% das proteínas de *T. rubrum*). Os dermatófitos também são semelhantes em seus genes, pois, um conjunto de 6.168 grupos ortólogos comuns aos sete genomas de dermatófitos foi identificado (Figura 1). O número total de ortólogos encontrados em todos os dermatófitos varia entre 69% (*M. canis*) a 77% (*A. benhamiae*). Dos grupos ortólogos identificados, 308 são únicos para os dermatófitos.

Figura 1- Genes ortólogos em fungos dermatófitos



Fonte: MARTINEZ et al. (2012, p. 4).

Na análise comparativa realizada por Martinez et al. (2012) relata-se que os dermatófitos sequenciados possuem tamanhos similares, sendo 22.5 Mb para *T. rubrum* a 24.1 para *T. equinum*. Os autores também relatam quatro grupos gênicos associados à patogenicidade dos dermatófitos, algo que corrobora com as análises iniciais dos dois primeiros dermatófitos sequenciados. Estes grupos gênicos são: (i) as proteases secretadas para degradar substratos proteicos da pele que atuam como fatores de virulência, (ii) quinases, possivelmente envolvidas na sinalização necessária para adaptação à pele e unha, (iii) genes envolvidos no metabolismo secundário, que atuam como toxinas ou sinais de interação entre fungos e hospedeiro e (iv) uma classe de proteínas (*LysM*) que possivelmente se ligam a carboidratos para mascarar os componentes da parede celular e hidratos de carbono, evitando assim, a resposta imune do hospedeiro aos fungos.

O sequenciamento do genoma dos principais fungos dermatófitos está auxiliando na identificação de novos fatores de virulência destes importantes organismos, contribuindo significativamente para o conhecimento de suas bases moleculares, como o conhecimento de repetições adjacentes encontradas em seu genoma.

1.4. GENES RELACIONADOS À PATOGENICIDADE EM FUNGOS DERMATÓFITOS

A compreensão dos mecanismos envolvidos em uma infecção dermatofítica é a base para o desenvolvimento de novas estratégias para seu tratamento (MARTINEZ-ROSSI et al., 2004). A adesão de microrganismos aos tecidos do hospedeiro é um passo importante no estabelecimento da maioria das infecções e um pré-requisito para infecções por dermatófitos. A aderência dos fungos as células hospedeiras é mediada através de adesinas fúngicas e suas interações com receptores do hospedeiro, porém pouco se sabe sobre os fatores que mediam a adesão de dermatófitos, no entanto, estudos indicam que as proteases secretadas podem facilitar ou serem necessárias para a adesão eficiente (BALDO et al., 2012).

Várias proteases já foram isoladas a partir de diferentes espécies de dermatófitos e mostraram atividades queratinolítica e/ou colagenolítica. O trabalho de Brouta et al. (2002) descreve uma família de genes (MEP1, MEP2 e MEP3) codificando três endo metaloproteínases em *Microsporum canis*. Outros autores identificaram duas metaloproteases, MEP4 e MEP5, em *M. canis* como parte de uma família de cinco genes (chamado MEP1-5), também identificadas em *Trichophyton rubrum* e *Trichophyton mentagrophytes* (JOUSSON et al., 2004).

Mais tarde, sete genes (SUB1-7) foram isolados na espécie *T. rubrum* por Jousson et al. (2004). Os autores relataram também a identificação de genes ortólogos em outras espécies de dermatófitos, três deles semelhantes aos genes SUB de *M. canis*. Estudos realizados por Dong et al. (2005) demonstraram que as atividades das metaloproteases MEP-2 e MEP-9 alcançaram um alto nível após três dias de infecção (72 h), também sendo possível correlacionar estas proteínas à virulência do dermatófito.

Levantamentos realizados por Maranhão et al. (2009) sugerem que o gene responsável pela codificação do transportador ATPase de cobre deve estar relacionado com a patogenicidade de alguns organismos, inclusive de *T. rubrum*. Ademais, genes codificadores de transportadores do tipo ABC foram identificados como importantes fatores de virulência nos estágios iniciais da infecção por fungos filamentosos, como observado em *M. grisea* (gene ABC1), *Botrytis cinerea* (BcatrB) e *Gibberella puricales* (VERMEULEN; SCHOONBEEK; DE WAARD, 2001).

O crescimento, fisiologia, diferenciação e a capacidade de infecção de todos os organismos, incluindo os fungos dermatófitos tem suas respostas adaptativas afetadas em

função do pH ambiente (MARTINEZ-ROSSI et al., 2004). Uma via metabólica composta por no mínimo sete genes é responsável pela resposta ao pH, o gene *pacC* que expressa o fator de transcrição *PacC* e os genes de sinalização *palA*, *palB*, *palC*, *palF*, *palH* e *palI* (DENISON, 2000; OREJAS et al., 1995). Experimentos realizados por Ferreira-Nozawa et al. (2006) e Silveira et al. (2010) demonstraram que o gene *pacC* de *T. rubrum* está relacionado à virulência, pois após ser nocauteado em uma linhagem mutante, teve sua capacidade infectante reduzida. Martinez-Rossi et al. (2012) sugeriram que o gene *pacC* do dermatófito *T. rubrum* tenha um papel fundamental na regulação de genes necessários a uma infecção eficiente, sendo importante sua relação com a virulência de fungos filamentosos como *T. rubrum*. Na Tabela 3 há uma lista de proteínas e genes identificados em dermatófitos provavelmente envolvidos com sua virulência (PERES et al., 2010).

Tabela 3- Proteínas/Genes potencialmente envolvidas na virulência de dermatófitos

Gene/proteína	Função em fungos
Isocitrato liase	Enzima do ciclo do glioxalato
Malato sintase	Enzima do ciclo do glioxalato
Citrato sintase	Enzima do ciclo do glioxalato
Fosfolipase B	Inativação gênica causa atenuação na virulência de <i>Cryptococcus neoformans</i> e <i>C. albicans</i>
Protease subtilisina (Sub3)	Sub3 é a principal protease secretada por <i>T. rubrum</i> durante a infecção no hospedeiro
Protease subtilisina (Sub 5)	Protease secretada é um importante fator de virulência
Metaloprotease (Mep3)	MEP3 é produzida por <i>M. canis</i> durante infecção em porquinho da Índia
Metaloprotease (Mep4)	Mep4 é a principal metaloprotease secretada por <i>T. rubrum</i> durante a infecção no hospedeiro
Carboxipeptidase	Importante para a assimilação de substratos durante a infecção
Dipeptidil-peptidase V	Fatores de virulência em potencial de <i>M. canis</i> , auxiliando na degradação de substratos.
ATPase P-type associada a resistência ao cobre	Inativação gênica leva à atenuação de virulência de <i>Listeria monocytogenes</i> e <i>C. neoformans</i>
Transportador ABC <i>TruMDR2</i>	Inativação gênica atenua a virulência de <i>T. rubrum</i> <i>in vitro</i>

Gene/proteína	Função em fungos
Manosiltransferase	Inativação gênica atenua a virulência de <i>C. albicans</i> e <i>A. fumigatus</i>
Urease	Inativação gênica reduz a quantidade de amônia secretada in vitro e atenua a virulência de <i>Coccidioides posadasii</i>
Glucosamina-6-fosfato deaminase	Inativação gênica atenua a virulência de <i>C. albicans</i> em modelo murino
Gliceraldeído-3-fosfato desidrogenase (GAPDH)	GAPDH contribui para a adesão de <i>Paracoccidioides brasiliensis</i> no tecido hospedeiro e para a disseminação da infecção
Fator de transcrição <i>PacC</i>	Inativação gênica atenua a virulência de <i>T. rubrum</i> in vitro
Tiorredoxina TrxA	Provável fator de virulência de <i>T. mentagrophytes</i>

Fonte: PERES et al. (2010, p. 662)

O conhecimento acerca das bases moleculares dos fatores de virulência em fungos patogênicos constitui um importante passo na identificação de novos alvos celulares para o desenvolvimento de novas drogas.

1.5. PROTEÍNAS DE ADESÃO E REPETIÇÕES ADJACENTES

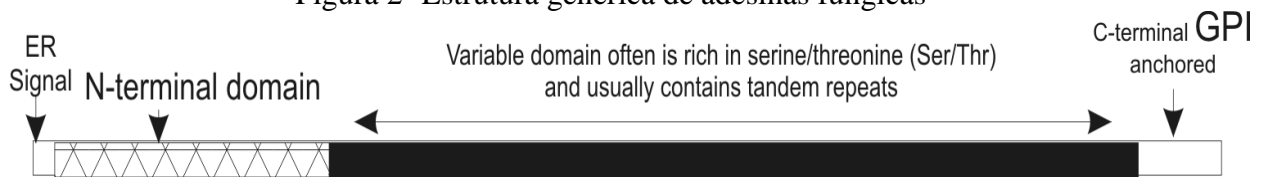
Adesinas são uma classe especializada de proteínas de superfície celular encontradas em fungos patogênicos, sendo responsáveis pela invasão da célula hospedeira e, portanto, determinantes para a virulência (NOBILE et al., 2008).

Recentemente, estudos sobre adesinas fúngicas utilizando nanotecnologia a partir de cristalografia de raios-X, ressonância magnética nuclear e microscopia de força atômica vêm auxiliando na obtenção de informações detalhadas a respeito de sua estrutura de ligação em organismos como *C. albicans* e *C. glabrata* (BEAUSSART et al., 2012). Estas novas abordagens têm demonstrado que adesinas fúngicas são responsáveis por outras funções como a formação do biofilme e respostas imunitárias do hospedeiro, o que contribui para o aumento da resistência a medicamentos e persistência das infecções (DE GROOT et al., 2013).

As adesinas fúngicas mais conhecidas possuem uma região de âncora GPI (glicosilfosfatidilinositol). A estrutura primária de precursores proteicos GPI possuem características conservadas, as quais podem ser utilizadas para identificar adesinas putativas por meio de bioinformática (RAMANA; GUPTA, 2010). Na sua extremidade N-terminal, possuem um peptídeo sinal para a entrada no retículo endoplasmático, e na sua extremidade C-terminal possui um peptídeo de ancoragem a um lipídeo GPI. A região N-terminal é

seguida por um domínio variável de baixa complexidade, que muitas vezes é rica em serina/treonina (Ser/Thr) e, geralmente, contém repetições em *tandem* (TRs) como pode ser observado na Figura 2. As regiões repetitivas apresentam polimorfismos significativos intraespécies devido a eventos de recombinação durante a replicação do DNA, sendo que regiões repetitivas mais longas podem conferir uma maior aderência, enquanto que regiões com repetições mais curtas podem resultar na redução da adesão (VERSTREPEN; KLIS, 2006).

Figura 2- Estrutura genérica de adesinas fúngicas



Fonte: Adaptado DE GROOT et al. (2013).

Repetições adjacentes intragênicas em adesinas podem desencadear eventos de recombinação e formação de novas adesinas, oferecendo aos fungos um amplo reservatório de propriedades de adesão. Estes aspectos da adesão de fungos exemplificam sua plasticidade fenotípica, permitindo-lhes adaptar-se rapidamente a ambientes estressantes e explorar novas oportunidades (VERSTREPEN; KLIS, 2006).

Diversas adesinas fúngicas já relatadas são ricas em repetições adjacentes variáveis (VNTR – *Variable Number of Tandem Repeats*), sendo um amplo objeto de estudo em *C. albicans*. Segundo Hoyer et al. (2008), na família de adesinas ALS de *C. albicans*, o número de cópias do domínio central repetitivo em *tandem* varia em cada gene ALS de acordo com o isolado. Segundo os autores, o gene desta família mais comumente isolado ALS1, possui o domínio central repetitivo variável entre 4 e 37 cópias por isolado, já o gene ALS7 possui uma variação entre 14 e 17 repetições em *tandem* do domínio central, sendo esta variabilidade associada a sua função.

Segundo Oh et al. (2005), a proteína ALS3 de *C. albicans* possui mais comumente 9 ou 12 cópias de uma sequência de repetição em *tandem*. Após realizarem experimentos de adesão *in vitro*, os autores observaram que proteínas com 12 cópias repetidas contribuem mais para a adesão de *C. albicans* a células endoteliais ou epiteliais do que aquelas com 9 cópias.

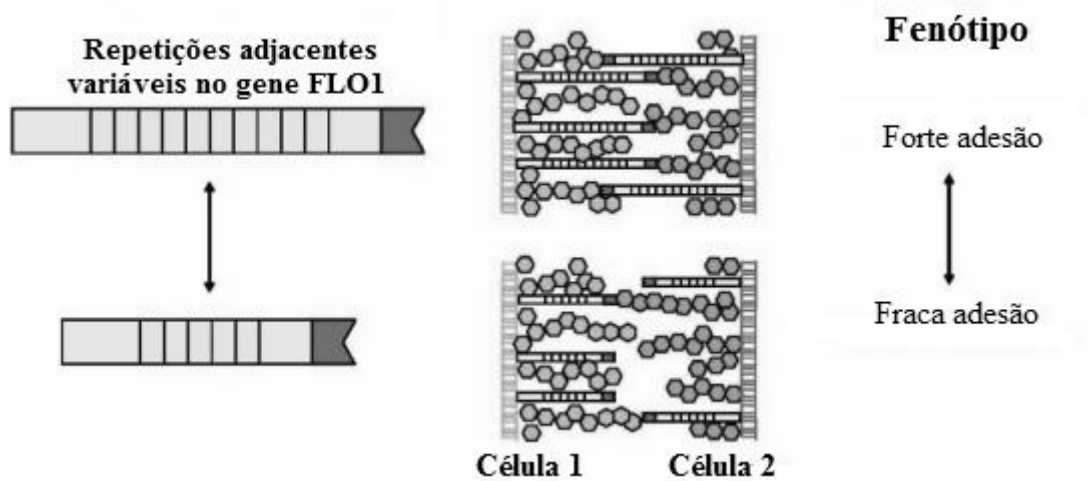
Proteínas de superfície que possuem domínios de repetições em *tandem* têm características geralmente "pegajosas" na interação proteína-proteína. As repetições em *tandem* existentes na proteína ALS5 de *C. albicans* mostram-se envolvidas com sua

capacidade de agregação, pois ensaios de adesão *in vitro* mostraram que a extensão das repetições adjacentes atuam em sua adesão (FRANK et al., 2010). Rauceo et al. (2006) realizaram a um estudo de mutagênese dirigida no qual os autores removeram as repetições adjacentes em um gene homólogo de ALS5 em *S. cerevisiae* observando uma redução significativa em sua capacidade de agregação. Portanto, a presença de repetições adjacentes é importante para maior capacidade de agregação de ALS5.

Em *Aspergillus fumigatus*, Levdansky et al. (2007) demonstraram que genes com repetições adjacentes desempenham um papel importante na interação patógeno-hospedeiro. Os autores realizaram a disrupção do gene Afu3g08990 que possui uma repetição adjacente de unidade igual a 18-pb que se repete por 32 vezes. A supressão da proteína até então caracterizada como hipotética resultou em um fenótipo com menor capacidade de adesão.

Já Verstrepren et al. (2005) realizaram um experimento em laboratório no qual os autores alteraram o número de repetições adjacentes do gene FLO1 em uma linhagem de *S. cerevisiae* para quantificar a alteração no fenótipo. Observou-se que o aumento do número de repetições resultou em uma adesão mais forte. Na Figura 3 ilustra-se o impacto da variação das repetições no gene FLO1 de que codifica uma adesina de superfície celular que media a adesão célula-célula e célula-superfície.

Figura 3- Repetições adjacentes em regiões codificantes modulam a adesão celular



Fonte: Adaptado GEMAYEL et al. (2012)

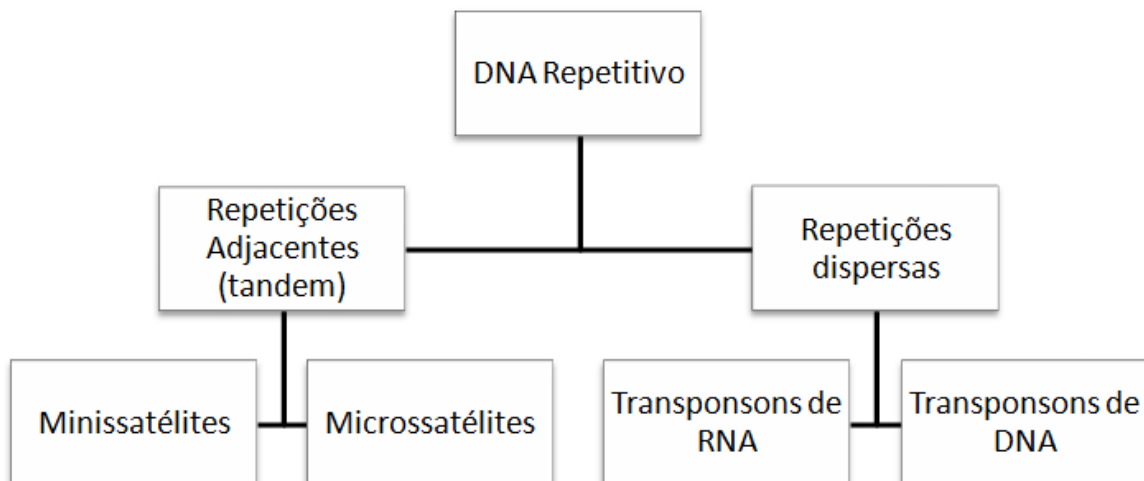
Em dermatófitos, o conhecimento dos mecanismos envolvidos com sua adesão ao hospedeiro permanecem limitados (BALDO et al., 2012), não existindo até o presente nenhum estudo que faça relação entre fatores de virulência e padrões de repetição existentes

em seu genoma. No entanto, uma melhoria neste campo de pesquisa tem sido obtida nos últimos anos devido ao desenvolvimento de novos modelos experimentais e graças à disponibilidade de seus dados genômicos.

1.6. CONCEITOS BÁSICOS SOBRE REPETIÇÕES ADJACENTES

Repetições adjacentes ou *tandem repeats* são sequências repetidas que não ocorrem apenas sequencialmente, mas também diretamente adjacente. Diferentes categorias de DNA repetitivos são conhecidas. Uma característica de distinção é se estas repetições estão agrupadas de forma adjacente ou se estão espalhadas por todo o genoma. Os tipos de repetições adjacentes são chamados coletivamente de DNA satélites, variando em relação a sua localização no genoma, ao comprimento total da repetição e ao tamanho da unidade de repetição (NUSSBAUM; MCINNES; WILLARD, 2008). O esquema presente Figura 4 representa a classificação de repetições de DNA adaptado de Brown (2002).

Figura 4- Classificação de DNA repetitivo



Fonte: Adaptado BROWN (2002)

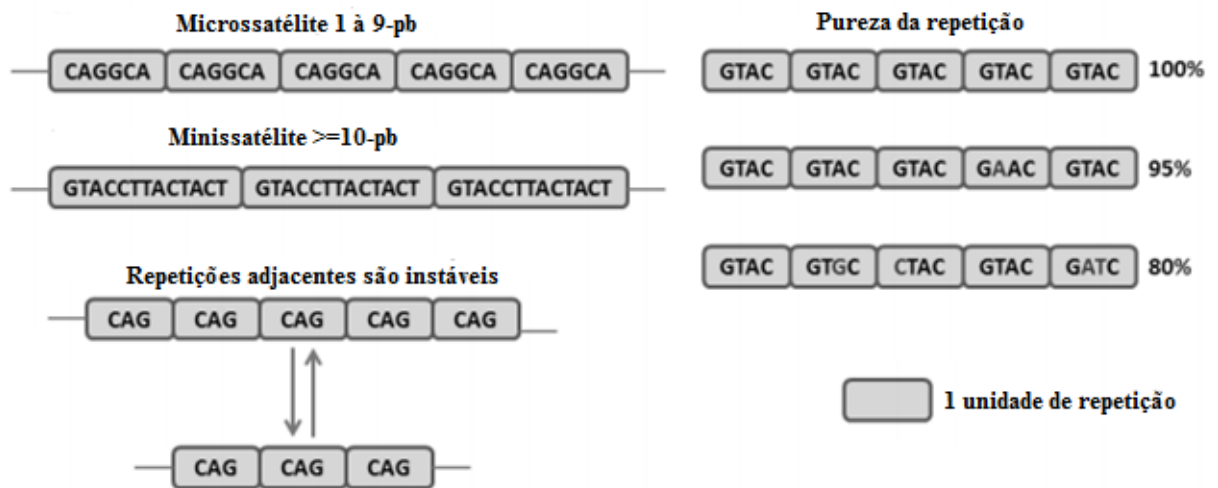
Uma unidade de repetição é considerada como uma sequência de nucleotídeos que se expande por uma determinada quantidade de vezes. Considerando a sequência de bases **AATAT-AATAT-AATAT-AATAT** temos a unidade de repetição de 5-pb **AATAT** que se repete por 4 vezes, formando assim, uma região repetitiva com comprimento total de 20-pb.

As repetições adjacentes podem ser classificadas como perfeitas ou imperfeitas (também classificada como pureza), dependendo se elas são cópias exatas umas das outras ou possuem alguma incompatibilidade mutacional de inserção ou exclusão de pares de base (ELLEGREN, 2004). Um exemplo de repetição imperfeita seria a sequência **AATAG-**

AATAT-AATAT-ATTAT em contraposição a repetição perfeita apresentada anteriormente.

Além disto, outra característica das repetições adjacentes é que a quantidade de unidades de repetição é altamente variável entre espécies e linhagens, desta maneira são comumente classificadas como instáveis (LEGENDRE et al., 2007). A Figura 5 apresenta as características gerais das repetições adjacentes de acordo com Boeynaems (2012).

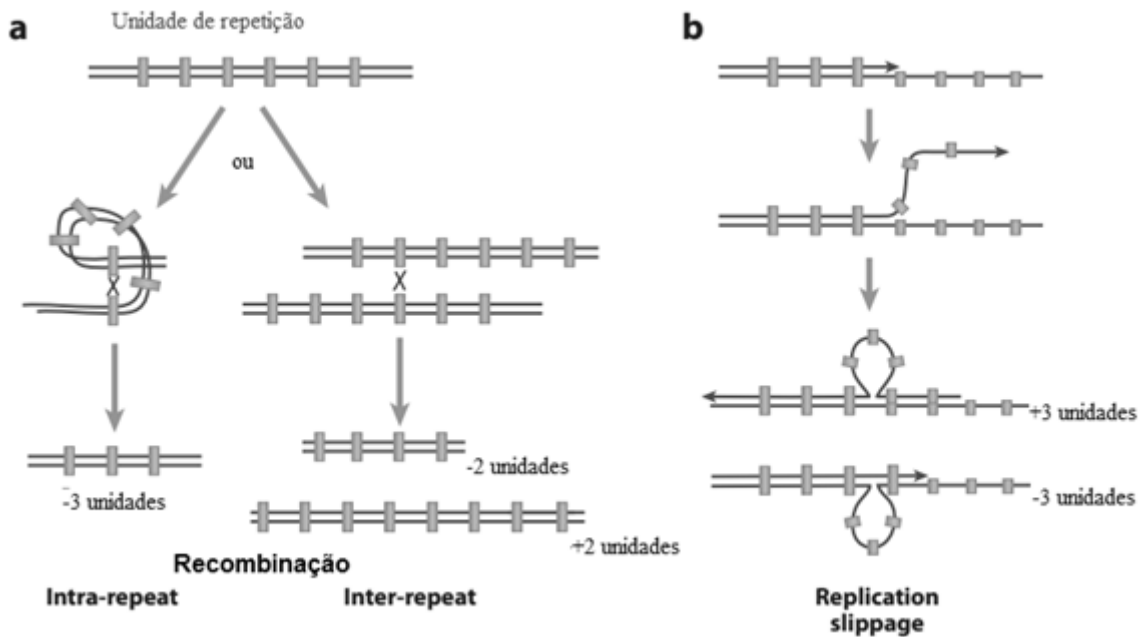
Figura 5- Características gerais das repetições adjacentes



Fonte: Adaptado BOEYNAEMS (2012).

Repetições adjacentes tem como principal origem eventos de recombinação e replicação. Na replicação pode ocorrer um erro chamado de “*strand slippage replication*” onde ocorre um erro no pareamento entre a fita molde do DNA e a fita que está sendo sintetizada. Assim, se um *loop* se formar na fita molde ocorre uma contração da repetição, se o *loop* se formar na fita sintetizada ocorre uma expansão na extensão da repetição (Figura 6a). Na recombinação, o pareamento e eventos “*crossing-over*” desiguais podem expandir ou contrair a sequência (Figura 6b) (GEMAYEL et al., 2010) .

Figura 6- Mecanismos de expansão e contração de repetições adjacentes



Fonte: Adaptado GEMAYEL et al. (2010).

Em sequências codificantes, repetições com unidades que contêm múltiplo de três nucleotídeos, como tri e hexanucleotídeos, são mais comuns, provavelmente pela seleção contra mutações da janela de leitura que ocorreria em repetições com unidades que não contêm um múltiplo de três nucleotídeos (METZGAR; BYTOF; WILLS, 2000).

1.7. REPETIÇÕES ADJACENTES

Sequências repetitivas de DNA constituem uma parte importante de todos os genomas, assim o recente sequenciamento completo de diversos organismos levou a um interesse renovado em repetições adjacentes. Diversos pesquisadores demonstraram que a ocorrência destas repetições em regiões codificantes e promotoras não é aleatória: genes que contêm repetições em série são enriquecidos em classes funcionais específicas (BOEYNAEMS, 2012).

Estas repetições são comuns em organismos eucariotos, o genoma humano, por exemplo, contém aproximadamente 260.000 repetições, o que consiste em cerca de 7% do genoma (RICHARD; KERREST; DUJON, 2008). Análises já realizadas demonstram que mais de 30% dos genes no genoma humano contêm repetições em regiões codificantes (LEGENDRE et al., 2007).

Existe grande interesse biológico em encontrar padrões de repetições adjacentes em um genoma, pois estes padrões estão associados com uma ampla gama de doenças humanas (MIRKIN, 2007), incluindo síndrome do X frágil, doença de Huntington, ataxia de Friedreich, certas formas de câncer e mais de 40 outras doenças neurológicas, neurodegenerativas e neuromusculares (USDIN, 2008).

A presença de *tandem repeats* como agentes causadores de várias doenças humanas levou a crença de que essa sequência repetitiva só pode conferir fenótipos nocivos ou na melhor das hipóteses um "DNA lixo neutro". Entretanto, pesquisas realizadas na última década como as de Verstrepen et al. (2005), Vincens et al. (2009), Boeynaems (2012) e Duitama et al., (2014) apontam que *tandem repeats* também possuem funções benéficas. Por meio de sua instabilidade inerente, estas sequências podem conferir uma variabilidade funcional útil. Os benefícios para o organismo também pode explicar a conservação através da evolução (LEVDANSKY; SHARON; OSHEROV, 2008).

Levantamentos realizados por Gemayel et al. (2012) descrevem que sequências repetitivas estão relacionadas a funções biológicas e estruturais importantes no genoma. Segundo os autores a instabilidade destas repetições em regiões de codificantes, promotores e outras regiões reguladoras podem eventualmente proceder para alterar a função do gene. Em outro trabalho Gemayel et al. (2010), descrevem que repetições adjacentes não são aleatoriamente presentes em sequências codificantes, pois genes de regulação, de parede celular e proteínas induzidas pelo estresse são particularmente enriquecidas com repetições.

Em vírus, estudos de genômica comparativa de cepas atenuadas e virulentas de *Gallid herpesvirus2* (GaHV-2) identificaram diferenças no número de repetições nos genes UL36 e UL47 que se mostraram correlacionados à sua virulência (SPATZ; SILVA, 2007). Em bactérias são importantes para sua adaptação ao ambiente, pois para se proteger contra mudanças ambientais, genes que contêm estas repetições mutam a extensão das mesmas, fazendo com que estes genes sejam alternados entre um estado de atividade e inatividade (MOXON; BAYLISS; HOOD, 2006).

De acordo com Levdansky; Sharon e Oshero (2008) repetições adjacentes são encontradas em diferentes as classes de proteínas fúngicas. A análise genômica mostra que estas repetições encontradas regiões de codificação podem ser classificadas em três grupos com base em motivos funcionais: (i) Proteínas destinadas ao transporte para a membrana plasmática, e/ou relacionados à parede celular; (ii) Proteínas contendo uma sequência de

peptídeo de sinalização, que são destinados principalmente para a secreção; (iii) Proteínas que não possuem estes motivos, sendo localizadas no interior da célula. O primeiro grupo de genes que codificam proteínas de parede celular ricas em repetições possui segundo os autores potencial para mediar a interação fungo-hospedeiro, sendo estas repetições mais comumente encontradas nas proteínas da parede celular de fungos do que em outras classes de proteínas.

De acordo com Verstrepen et al. (2005), a variação nestas repetições em regiões intragênicas na proteína *FLO1* de *S. cerevisiae* está associada com o aumento da adesão célula a célula como já descrito anteriormente, e as variações de repetições na proteína *FLO11* contribui para a formação de biofilme.

Algumas repetições mostram ter evoluído de um modo dependente em genes homólogos. Por exemplo, alguns homólogos, como *SIZ1* e *SIZ2* de *Saccharomyces cerevisiae*, mostram a mesma sequência de repetição deslocada em algumas posições no gene, e, em outros casos, como nos genes *WSC2* e *WSC3*, uma repetição semelhante (mas não igual) ocorre na mesma posição em um par de genes homólogos (GEMAYEL et al., 2010).

Portanto, a identificação e caracterização dessas sequências repetitivas tornou-se uma parte importante em projetos genômicos, pois este fenômeno pode contribuir significativamente para a variação genética e patogênica entre os organismos. Porém, apesar da sua prevalência em regiões funcionais do genoma e sua associação com várias doenças neurológicas humanas, e apesar da sua utilidade como marcadores genéticos, estudos sobre repetições em *tandem* permanecem escassos (DUITAMA et al., 2014).

Quanta a sua extensão, estas repetições podem ser classificadas como minissatélites ou microssatélites de acordo com o tamanho de sua unidade de repetição, sendo que cada tipo está associado à determinadas funções no genoma.

1.8. MICROSSATÉLITES E MINISSATÉLITES

Microssatélites e minissatélites são *tandem repeats* com repetições de unidades curtas, sendo a unidade de repetição a principal característica para sua classificação. Entretanto, segundo Richard; Kerrest e Dujon (2008), não existe um consenso sobre a definição para ambos os tipos de repetições. Alguns autores não consideram repetições de mononucleotídeos como microssatélites, e para outros o corte entre micro e minissatélites fica entre unidades de 6 a 10-pb.

Muitos estudos consideram microssatélites como repetições de unidades sendo <10-pb, minissatélites com unidades de 10<100-pb e satélites DNA ou megassatélites repetições com unidades sendo >100-pb . Além disso, não existe um consenso da quantidade mínima de repetições das unidades para ser considerado um *tandem repeat*. Os autores divergem que são necessárias de duas a cinco repetições das unidades no mínimo. Nos últimos anos, houve um aumento na quantidade de trabalhos relacionados à microssatélites publicados, podendo-se observar que diferentes algoritmos e parâmetros foram utilizados, não se observando um consenso (RICHARD; KERREST; DUJON, 2008).

Microssatélites são mais prevalentes em genomas eucariotos, sendo amplamente utilizados para tipagem molecular, tendo suas pesquisas iniciadas para identificação de cepas de *S. cerevisiae* (RICHARD; DUJON, 1996), sendo também utilizados para outros fungos patogênicos como *Candida albicans* e *Candida glabrata*. Estes *repeats* também são utilizados por geneticistas para estudos epidemiológicos, populacionais e evolutivos (MCMANUS; COLEMAN, 2014).

Quando encontrados perto de início da transcrição de um gene, ou seja, no promotor, microssatélites podem aumentar ou diminuir os níveis de expressão gênica por expansão e contração de seu comprimento. Desta maneira, microssatélites são frequentemente relacionados com a regulação da expressão (LEGENDRE et al., 2007).

Repetições de microssatélites de poliglutamina (poliQ) são observados nos genes de um grande número de pacientes com diferentes doenças genéticas, tais como várias ataxias e doença de Huntington. A agregação proteica, que é uma característica fundamental da maioria destas doenças, mostra-se desencadeada por estas sequências poliQ expandidas em genes relacionadas com a doença, como no gene HTT, crucial na doença de Huntington. No entanto, intervalos de poliQ é uma característica de muitas proteínas normais humanos, o que sugere que eles possuem uma importante função celular. Para esclarecer a potencial função de repetições poliQ em sistemas biológicos, foram analisadas sistematicamente informações disponíveis armazenados em bancos de dados de sequência e interação proteica. Ao integrar genômica, filogenética, rede de interação de proteínas e informação funcional, obteve-se evidências de que o trato poliQ nas proteínas é responsável por estabilizar interações proteicas (SCHAEFER; WANKER; ANDRADE-NAVARRO, 2012).

Minissatélites (repetições com unidades maiores ou iguais a 10-pb) foram objeto de estudos Verstrepen et al. (2005) e Richard e Dujon (2006) no qual os autores observaram que

entre 50% e 60% dos genes de fungos que contêm repetições de minissatélites codificam proteínas de parede celular e adesão celular. Mais importante, este enriquecimento funcional é conservada ao longo da evolução. Além disto, outro trabalho de Thierry et al. (2008), demonstrou que grandes minissatélites no fungo patogênico *Candida glabrata* ocorre em genes diretamente envolvidos com a adesão e patogenicidade.

Segundo Legendre et al. (2007), em organismos eucariotos, minissatélites são frequentemente encontrados em genes que codificam proteínas extracelulares, tais como mucinas, enquanto microssatélites são na sua maioria presentes em genes reguladores. Vários estudos indicam que polimorfismos em minissatélites podem confere variabilidade funcional. O gene de *S. cerevisiae* FLO1 codifica uma adesina da superfície celular que possui uma repetição de minissatélites de aproximadamente 100-pb, sendo esta fortemente relacionada à capacidade de adesão.

Assim, a compreensão sobre a função destas repetições nos organismos é importante, pois, as mesmas constituem uma importante fonte de variação fenotípica e adaptação ao ambiente por possuírem uma natureza de instabilidade na extensão de suas repetições (VINCES et al., 2009).

1.9. REPETIÇÕES ADJACENTES VARIÁVEIS

Genomas podem mutar de diferentes maneiras, que variam de grandes rearranjos a substituições de um único nucleotídeo. Dependendo onde ocorrem, o impacto destas mutações podem variar. A maior parte do genoma tem uma baixa taxa de substituição de nucleotídeos, mas algumas regiões do genoma apresentam taxas elevadas de mutação, e muitas destas são ricas em repetições adjacentes (RANDO; VERSTREPEN, 2007).

Repetições em *tandem* são instáveis devido a sua natureza repetitiva. Comparado a outros *loci*² do genoma, as taxas de mutação de repetições em *tandem* são de 10 a 10.000 vezes maiores. Desta maneira, o que diferencia essas repetições de outras regiões do genoma é sua instabilidade, sendo que a grande maioria das mutações é de alterações no número de unidades de repetição. Devido a esta variabilidade, o termo VNTR (*Variable Number of Tandem Repeat*) - número variável de repetições adjacentes é comumente usado na literatura (VERSTREPEN et al., 2005).

² Região do genoma onde está localizado determinado gene ou marcador genético

Polimorfismos de base única ou SNPs (*Single Nucleotide Polymorphism*) são sequências já conhecidas como fonte de diversidade genética e evolução morfológica, no entanto, o estudo realizado por Fondon e Garner (2004) propôs que repetições adjacentes são outra fonte de variação. O estudo comparou os padrões de repetição adjacente em genes regulatórios de 92 raças de cães, encontrando evidências que certas variações morfológicas estão relacionadas a sequências repetitivas em regiões de codificação. Analisando o gene *Runx-2* (conhecido por regular a diferenciação de osteoblastos - células responsáveis pela formação óssea) os autores identificaram duas extensões de repetições variáveis, uma que codifica uma sequência de poliglutamina e outra de polialanina. Uma comparação estatística do comprimento total de ambas estas repetições (poliQ + poliA) com vários parâmetros de tamanho do crânio canino revelou que as repetições mostram-se correlacionadas com diferenças do dorsoventral, nariz e comprimento da face entre diferentes raças de cães como pode ser observado na Figura 7 (FONDON; GARNER, 2004).

Figura 7- Repetições adjacentes no gene *Runx-2* promovem uma rápida evolução em cães



Fonte: FONDON; GARNER (2004, p. 18062).

A contração e expansão de repetições adjacentes podem ocorrer durante o processo de replicação e recombinação do DNA, pois há uma probabilidade relativamente alta de que o deslizamento da polimerase irá fazer com que o número de repetições aumente ou diminua (FONDON; GARNER, 2004). Embora todas as repetições adjacentes sejam inerentemente

instáveis, as taxas de mutação variam muito entre elas, dependendo de múltiplos fatores, principalmente do tamanho da unidade, comprimento e perfeição da repetição. Dessa maneira, quanto mais longa e pura for a repetição, maior será a frequência de mutação, ao passo que quanto menor e menos pura, mais baixa será a frequência de mutação (GEMAYEL et al., 2010). Legendre et al. (2007) utilizaram técnicas de aprendizado de máquina para desenvolver um modelo algorítmico que estima a variabilidade de uma dada repetição (utilizado neste trabalho). Segundo os autores, o número de unidades de repetição é a característica que mais impacta na instabilidade da repetição, acima da pureza e do tamanho da unidade.

A alta prevalência de repetições de microssatélites em regiões reguladoras em muitos organismos sugere que a variação das repetições pode ser um comum, evolutivamente conservado, mecanismo de regulação da expressão gênica (GEMAYEL et al., 2010). Diversos estudos relatam correlações entre repetições de microssatélites e alterações na expressão genética, como o trabalho de Vines et al. (2009) no qual os autores apontam que cerca de 25% de todos os promotores da levedura *Saccharomyces cerevisiae* contêm repetições adjacentes, em especial do dinucleotídeo CA, sendo que a variação destas repetições está associada com a expressão gênica dentro e entre as espécies de leveduras.

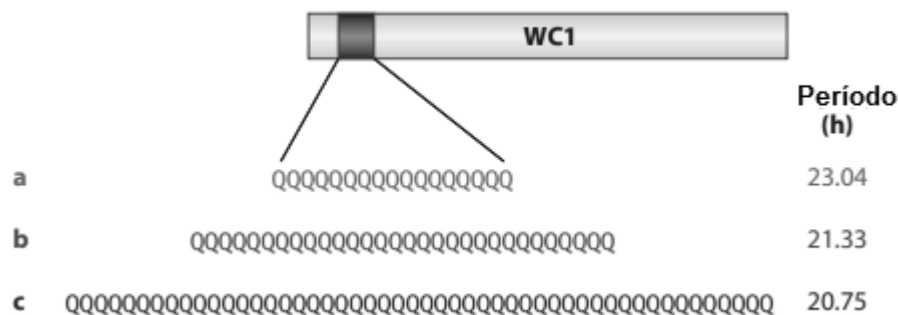
Um exemplo da relação das repetições com a expressão gênica é encontrada no fungo patogênico *C. albicans*. Este fungo expressa, entre outros fatores de virulência, proteases secretadas que o ajudam a alcançar nutrientes e invadir o tecido do hospedeiro. Tal protease é codificada pelo gene SAP2, que contém no seu promotor dois *tandem repeats* de cinco unidades variáveis (GCTTT e TTGAT/A). A alteração do número de unidades de repetição no promotor afeta a expressão de cada alelo durante a infecção, bem como o nível de indução em laboratório. O interessante neste promotor é que o número específico de unidades em cada repetição individual não determina a expressão, mas sim o número total de cópias das unidades combinadas das duas repetições (STAIB et al., 2002).

Em regiões codificantes, como descrito anteriormente, repetições adjacentes são conhecidas na sua maioria pela associação com um número de doenças neurodegenerativas humanas. Além disso, é bem descrito que repetições de minissatélites são enriquecidas em genes de superfície celular, sendo esta característica conservada através da evolução, desde leveduras até seres humanos, onde estudos indicam que as variações nesses minissatélites conferem uma real variabilidade funcional (VERSTREPEN et al., 2005).

Repetições adjacentes não são encontradas em muitos genes, existindo uma tendência para estas serem encontradas em genes que respondem às mudanças das condições ambientais, sendo que algumas dessas repetições em série podem atuar como mecanismos de ajuste ao ambiente através de alterações fenotípicas (GEMAYEL et al., 2012).

No fungo filamentoso *Neurospora crassa* o fator de transcrição WC-1 (*White collar-1*) induz a expressão de um componente chave para seu ciclo circadiano (mecanismo de tempo celular que sincroniza diferentes processos biológicos com o ambiente externo). A região N-terminal da proteína de WC-1 contém um padrão de repetição de poliglutamina (poliQ) necessária para o ciclo circadiano funcional em condições escuras contínuas. Em isolados naturais de *N. crassa*, o período de duração do ciclo circadiano variam entre eles, sendo esta variabilidade associada com o comprimento da repetição de poliglutamina em WC-1, onde quanto maior a repetição, menor a duração do ciclo (Figura 8). Esta associação foi validada através de do cruzamento entre duas linhagens com comprimentos diferentes de repetição de poliglutamina (MICHAEL et al., 2007).

Figura 8- Repetições adjacentes como mecanismo de ajuste para ciclos circadianos



Fonte: Adaptado (GEMAYEL et al., 2012)

Em *Aspergillus fumigatus*, o gene *cspA* (*cell surface protein A*) codifica uma região repetitiva variável de 188 aminoácidos de comprimento rica em serina-treonina-prolina [P-G-Q-P-S-(A/V)], sendo que o número de repetições varia de 18 a 47 entre os isolados. A disrupção do gene *cspA* resultou em um fenótipo caracterizado pela rápida germinação de conídios e reduzida aderência à matriz extracelular (LEVDANSKY et al., 2010).

Observações semelhantes também foram feitas para o fungo patogênico *Candida albicans*. Nos genes da família ALS que codificam glicoproteínas de superfície celular, o comprimento do domínio repetitivo central mostra-se correlacionado com a propensão de

aderência (repetições mais longas geram células mais adesivas), ainda, eventos de recombinação entre os diferentes genes de ALS podem gerar células com repetições variáveis necessárias para evadir a resposta imune (HOYER et al., 2008) .

Assim, a instabilidade das repetições adjacentes em regiões codificadoras, promotores, e outras regiões reguladoras podem, eventualmente, proceder para alterar a função dos genes. Desta maneira, em vez de sequências neutras sem um valor funcional, repetições adjacentes podem atuar como "botões de ajuste" para responder a mudanças ambientais, contribuindo de forma significativa para a variação funcional e na expressão gênica (VINCES et al., 2009) .

Levando-se em consideração a importância destas repetições, diferentes autores já relataram o desenvolvimento de ferramentas e metodologias para identificação de repetições adjacentes.

1.10. FERRAMENTAS PARA IDENTIFICAÇÃO DE REPETIÇÕES ADJACENTES

Dado o considerável interesse em repetições adjacentes, é possível encontrar ferramentas de bioinformática com diferentes abordagens desenvolvidas para detectar e caracterizar estas repetições. Levantamentos já realizados revelaram mais de 25 ferramentas existentes, e outras abordagens continuam a surgir. Algumas destas ferramentas são baseadas em expressões regulares sendo escritas em Perl como MsatFinder, SSRIT e MISA, enquanto outras utilizam uma abordagem combinatória como Sputnik e Mreps. Já a ferramenta Tandem Repeat Finder utiliza uma abordagem probabilística/estatística (MERKEL; GEMMELL, 2008).

Do ponto de vista do usuário, a identificação de repetições adjacentes dentro de uma sequência possui duas opções básicas: em primeiro lugar, se a pesquisa será padrão específico, ou inespecífico (com base na natureza repetitiva de uma única sequência) e, segundo, qual o tipo de repetição será procurada (perfeitas ou imperfeitas) (MERKEL; GEMMELL, 2008).

Do ponto de vista do programador deve-se definir uma abordagem. A maioria das ferramentas de busca por repetições adjacentes adota uma abordagem em duas fases baseadas em pesquisa e filtragem (MERKEL; GEMMELL, 2008). Na primeira fase, dois tipos de algoritmos são comumente implementados: (i) um algoritmo de busca combinatória que realiza exaustivas comparações de cada subsequência adjacente em seções de toda a sequência para identificar potenciais repetições de todos os tamanhos (KOLPAKOV; BANA;

KUCHEROV, 2003) ou (ii) um algoritmo estatístico ou baseado em heurística que usa janelas de segmento pequenas para procurar através da sequência e detectar possíveis repetições (normalmente pequenas repetições perfeitas) para então determinar se essa repetição pode ser mais longa (BENSON, 1999).

Na segunda fase, diferentes métodos de filtragem podem ser utilizados para identificar e extrair repetições com um significado biológico. A filtragem pode ser baseada em um filtro de comprimento simples ou em métodos mais complexos como heurística (*k-means clustering*) (KOLPAKOV; BANA; KUCHEROV, 2003) ou métodos estatísticos.

Através da análise *in silico* de cinco ferramentas amplamente utilizadas (Tandem Repeat Finder-TRF, Mreps, Sputnik, STAR e RepeatMasker), Levdansky; Sharon e Oshero (2008) demonstram que ajustes de parâmetros afetam a distribuição das repetições detectadas em número e comprimento. Os autores observaram que o comprimento mínimo, pesos de alinhamento e ajustes pontuação mínima possuem influência marcante, resultando em grandes discrepâncias entre as detecções realizadas.

Merkel e Gemmell (2008) destacam que as ferramentas variam muito na definição da repetição, algoritmo de pesquisa e método de filtragem, resultando em vários conceitos diferentes. Na comparação de cinco ferramentas de busca comumente utilizadas (TRF, Mreps, Sputnik, Msatfinder e Troll), eles também observaram diferenças significativas nos resultados, devido às opções de parâmetros.

Lim et al. (2012) realizaram uma análise algorítmica de consistência de desempenho em sete ferramentas de busca por repetições, sendo três amplamente utilizadas (TRF, Mreps e Sputnik) e outras quatro mais recentemente desenvolvidas (ATRHunter, iMEx, T-reks e INVERTER). Sputnik, Mreps, iMEx e INVERTER utilizam uma abordagem combinatória, enquanto dos demais utilizam um método estatístico/heurístico. Suas análises demonstraram que a ferramenta Mreps aparece no topo de desempenho, seguido por TRF, INVERTER, ATRHunter, T-reks, iMEx e Sputnik. Seus estudos ainda demonstram que ferramentas como iMEx que são projetados para busca de repetições curtas tendem a ter um melhor desempenho para genomas ricos em microssatélites, enquanto ferramentas como ATRHunter que são otimizados para repetições de maior comprimento, possui melhor desempenho em genomas ricos em tais repetições, enquanto TRF e Mreps por outro lado, demonstram executar igualmente bem em ambas as situações, como ferramentas para o uso mais generalizado.

Estas duas ferramentas (TRF e Mreps) foram selecionadas para a realização de testes e utilização neste trabalho.

1.10.1. TANDEM REPEAT FINDER (TRF)

TRF é provavelmente a ferramenta mais comum e amplamente utilizada para encontrar repetições sendo base de desenvolvimento para outras ferramentas similares (MERKEL; GEMMELL, 2008; PELLEGRINI; RENDA; VECCHIO, 2010).

A ferramenta é baseada em um modelo estatístico/probabilístico que utiliza um algoritmo de janela deslizante para procurar nucleotídeos separados por uma distância comum correspondente. Como no algoritmo de Smith e Waterman (1981), TRF utiliza *matches*³ parciais entre as cópias, chamadas k-tuplas. Para cada k-tupla encontrada, a distância de informação e localização é armazenada em um índice. Para selecionar os candidatos relevantes, uma variedade de critérios estatísticos são aplicados, que são derivados de distribuições de probabilidade (comprimento do padrão, probabilidade de correspondência, probabilidade de *indels*⁴ e tamanho k-tupla). O resultado não é uma busca exaustiva, mas suficiente para um processamento rápido e razoável de grandes conjuntos de dados, como por exemplo, os genomas de mamíferos. TRF encontra repetições com períodos entre 1-2000 nucleotídeos. Se uma repetição contém muitas cópias, o mesmo será detectado em vários tamanhos de período, porém isto é restrito as três melhores pontuações de período (BENSON, 1999).

1.10.2. MREPS

Mreps identifica repetições adjacentes baseado na técnica de extensão da repetição. Inicialmente, repetições exatas são detectadas então, dependendo dos parâmetros definidos pelo usuário, e então a mesma é estendida. Todos os resultados são submetidos a verificações de redundância e verificados estatisticamente com base em uma distribuição real (MERKEL; GEMMELL, 2008).

Os parâmetros do Mreps são os comprimentos mínimos e máximos de detecções (em pb), o número mínimo de repetições, e os comprimentos mínimos e máximos motivo. Esses parâmetros não afetam a execução do algoritmo, mas são usados para filtrar resultados finais retornados para o usuário. O algoritmo de distância de *Hamming* usado pelo Mreps percorre

³ Termo que define o padrão de correspondência entre diferentes segmentos de DNA.

⁴ Termo da biologia molecular para a inserção ou a deleção de bases no DNA de um organismo

k-valores que são independentes do período de repetição em *tandem*. (KOLPAKOV; BANA; KUCHEROV, 2003).

Mreps possui um parâmetro de resolução que permite identificar repetições não perfeitas a partir da lógica difusa, que consiste na extensão da lógica booleana admitindo valores intermediários entre o falso (0) e verdadeiro (1). O algoritmo de Mreps possui a limitação de não considerar *indels* (inserções / deleções de nucleotídeos), mas apenas substituições. Como resultado, *indels* são tratados de forma indireta, e certas repetições contendo *indels* podem não ser identificadas. (KOLPAKOV; BANA; KUCHEROV, 2003).

II. HIPÓTESE

Este trabalho partiu da hipótese que a utilização de ferramentas computacionais para identificação e análise de repetições adjacentes pode auxiliar na identificação de genes possivelmente relacionados à virulência em fungos dermatófitos bem como relacionar os mesmos a categorias funcionais específicas ricas em repetições.

III. OBJETIVOS

Nesta seção são apresentados os objetivos geral e específicos deste trabalho.

3.1. OBJETIVO GERAL

Tendo em vista a importância clínica e o recente sequenciamento do genoma das principais espécies de fungos dermatófitos, este trabalho teve como objetivo identificar e analisar *in silico* regiões adjacentes repetitivas intra e entre espécies utilizando ferramentas de bioinformática em fungos dermatófitos, especialmente em *T. rubrum*.

3.2. OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho foram:

- Analisar as ferramentas computacionais disponíveis atualmente para identificação de repetições adjacentes;
- Desenvolver um sistema *web* (*pipeline*) para automatizar os processos de identificação e armazenamento das repetições adjacentes encontradas;
- Traçar um perfil das repetições adjacentes identificadas e armazenadas pelo sistema desenvolvido;
- Relacionar as repetições à *motifs* funcionais dos genes;
- Comparar a variabilidade das repetições entre diferentes espécies e linhagens de fungos dermatófitos;
- Através da análise *in silico* de regiões repetitivas, identificar potenciais genes que possam estar relacionados à patogenicidade em dermatófitos;

IV. METODOLOGIA

Tendo em vista a necessidade da utilização de diferentes ferramentas para o desenvolvimento deste trabalho, nesta seção são descritas os procedimentos metodológicos definidos para emprego das mesmas.

4.1. LEVANTAMENTO DAS FERRAMENTAS/ALGORITMOS PARA BUSCA DE REPETIÇÕES ADJACENTES

A partir de pesquisas na literatura foram levantadas as principais ferramentas para a busca de repetições adjacentes, sendo estas apresentadas na Tabela 4.

Tabela 4- Ferramentas para detecção de *tandem repeats*

Ferramenta	Linguagem	Sistema Operacional	URL	Interface
TandemRepeatFinder	N.D. ¹	Windows, Linux, Mac	http://tandem.bu.edu/trf/trf.html	gui/web
Mreps	ANSI-C	Windows, Linux, Mac	http://bioinfo.lifl.fr/Mreps/	console/web
etandem/equicktandem	C	Unix, Windows, Mac	http://emboss.sourceforge.net/	gui/console
RepeatMasker	Perl	Unix	http://www.repeatmasker.org/	console/web
Misa	Perl	Linux	http://pgrc.ipk-gatersleben.de/misa/	console
SSRIT	Perl	Linux	http://www.gramene.org/db/searches/ssrtool	web
ATRHunter	C	Windows, Sun/Solaris, Linux	http://bioinfo.cs.technion.ac.il/atrhunter/ATRHunter.htm	web
TRING	C	UNIX, Windows, Mac	http://www.caspur.it/~castri/STRING/index.htm	console
TROLL	C++	Unix, Linux	http://finder.sourceforge.net	console

I: Linguagem não descrita na literatura.

A partir do levantamento das ferramentas, observando-se os resultados de experimentos comparativos realizados por Lim et al. (2012); Merkel e Gemell (2008) foram selecionadas as ferramentas Tandem Repeat Finder (BENSON, 1999) e Mreps (KOLPAKOV; BANA; KUCHEROV, 2003) para a identificação das repetições em fungos dermatófitos. Elas foram definidas levando em consideração a amplitude das referências na literatura de sua utilização para identificação de repetições adjacentes, desempenho e diferentes abordagens de codificação: probabilística e combinatória respectivamente.

4.2. COMPARAÇÃO DAS FERRAMENTAS/ALGORITMOS PARA BUSCA DE REPETIÇÕES ADJACENTES

A partir da definição das ferramentas para testes, utilizou-se os genes transcritos de *T. rubrum* (12,13 Mbp) para comparar o desempenho de TRF e Mreps. Para utilização da ferramenta TRF utilizou-se os parâmetros: *matching weight 2, mismatching penalty 5, indel penalty 5, match probability 0.8, indel probability 0.1, score ≥ 40 e maximum period 500*. Para utilização da ferramenta Mreps utilizou-se os parâmetros *-minsize 10 -minp 1 -res 1 -exp 3.0*. Para aferição do tempo de execução utilizou-se o comando *time* do sistema operacional Linux.

4.3. OBTENÇÃO DAS SEQUÊNCIAS

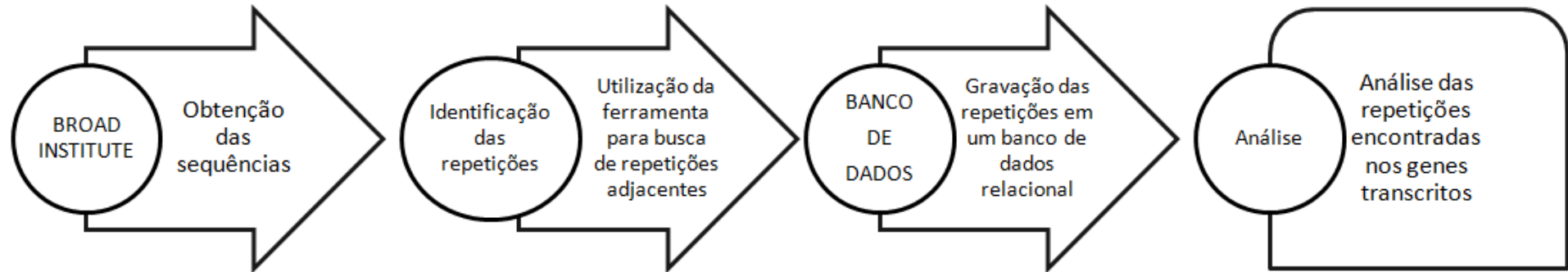
As sequências dos organismos analisados neste estudo foram obtidas na página da Internet do *Broad Institute* em maio de 2014 (http://www.broadinstitute.org/annotation/genome/dermatophyte_comparative). Este endereço apresenta os dados relativos ao projeto de sequenciamento dos principais fungos dermatófitos. Estes dados também estão disponíveis em bases de dados públicas como NCBI.

4.4. METODOLOGIA DE BUSCA E ARMAZENAMENTO DE REPETIÇÕES ADJACENTES

Partindo da hipótese que padrões de repetições adjacentes podem ter uma importante relação com a interação fungo-hospedeiro, buscou-se os dados de genes transcritos e regiões *upstream* dos dermatófitos *Trichophyton rubrum*, *Trichophyton tonsurans*, *Trichophyton equinum*, *Microsporum gypseum*, *Microsporum canis*, *Arthroderma benhamiae* e *Trichophyton verrucosum* disponíveis no banco de dados do projeto *Dermatophyte Comparative Sequencing Project* do *Broad Institute* (2014) para busca por repetições adjacentes.

As sequências utilizadas estão disponíveis através de arquivos multi-fasta⁵. O esquema apresentado na Figura 9 representa a metodologia utilizada para busca e armazenamento de repetições adjacentes.

Figura 9- Esquema de busca e armazenamento das repetições adjacentes.



Fonte: Elaborado pelo autor

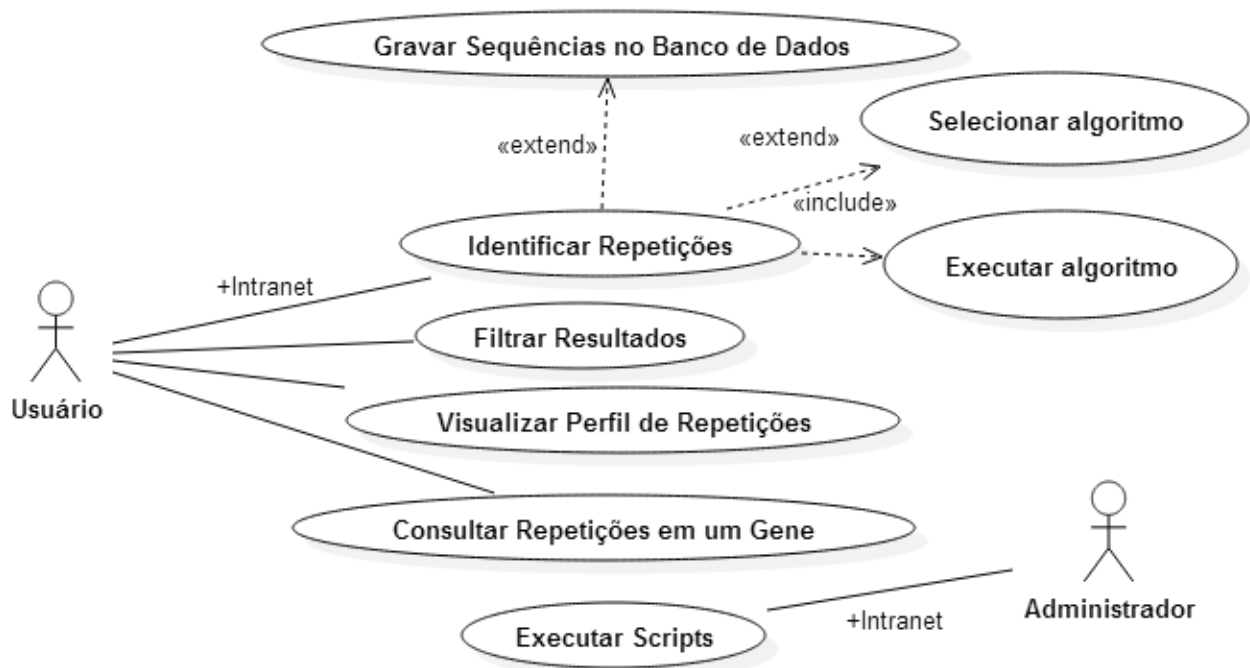
A partir da obtenção das sequências, elas são submetidas à ferramenta para identificação das repetições. Os resultados provenientes da ferramenta são armazenados em um banco de dados relacional, o que possibilita a posterior geração de relatórios e análise dos dados.

⁵ O formato multi-fasta é um arquivo texto composto por várias sequências de formato FASTA.

4.5. DESENVOLVIMENTO DO SISTEMA PARA BUSCA E ARMAZENAMENTO DE REPETIÇÕES ADJACENTES

Tendo em vista a necessidade de automatização dos processos de busca e armazenamento em um banco de dados relacional das repetições adjacentes identificadas pelas ferramentas utilizadas, foi implementado um *pipeline* - sistema baseado em um ambiente *web* - utilizando a tecnologia de desenvolvimento para Internet ASP.NET *Web Forms* por meio da linguagem de programação C#. Para isso, foi utilizado o ambiente de desenvolvimento integrado Microsoft Visual Studio 2012[®]. Para o armazenamento das repetições identificadas utilizou-se o sistema de gerenciamento de banco de dados relacional MySQL 5.2. O diagrama entidade relacionamento do banco de dados é apresentado no Apêndice A. Após a publicação dos resultados a ferramenta será disponibilizada sob licença GPL⁶. As ferramentas TRF e MREPS ficam disponíveis através de um servidor Ubuntu Linux versão 10.04, o qual é acessado de forma remota pela aplicação. O diagrama de caso de uso ⁷apresentado na Figura 10 descreve as principais funções executadas pelo sistema.

Figura 10- Diagrama de caso de uso do sistema desenvolvido



4.6. PARÂMETROS UTILIZADOS NA BUSCA POR REPETIÇÕES ADJACENTES

Os parâmetros utilizados na busca de repetições adjacentes para a ferramenta Mreps são apresentados na sequência: mreps -minsize 10 -minp 1 -res 1 -exp 3.0 ,sendo:

⁶ General Public License (Licença Pública Geral), é a designação da licença para software livre, permite que os programas sejam distribuídos e reaproveitados de forma livre, mantendo, porém, os direitos do autor.

⁷ Descreve a funcionalidade proposta para um sistema, sendo amplamente utilizado para o levantamento de requisitos funcionais.

- -minsize : Tamanho total mínimo da extensão da repetição .
- -minp: Tamanho mínimo do *repeat* a ser considerado.
- -res: O parâmetro de resolução (-res) especifica o "grau de imprecisão" de repetições que Mreps é capaz de encontrar. Quanto maior for a resolução, mais repetições degeneradas serão consideradas.
- -exp: Quantidade mínima de vezes que um *repeat* deve ocorrer em *tandem* para ser considerado na saída.

Através dos parâmetros definidos, a ferramenta permite a identificação de repetições perfeitas e imperfeitas. A detecção de repetições imperfeitas é possível pela utilização do parâmetro resolução, utilizou-se o valor mínimo igual a 1, recomendado para um baixo grau de degeneração da repetição. Foram consideradas repetições adjacentes com um comprimento mínimo de dez pares de base levando-se em consideração os parâmetros mínimos definidos em experimentos já realizados (KOLPAKOV; BANA; KUCHEROV, 2003; LECLERCQ; RIVALS; JARNE, 2007; LIM et al., 2012).

Os parâmetros utilizados na busca de repetições adjacentes para a ferramenta TRF foram definidos com base nos trabalhos de Legendre et al. (2007) e Vinces et al. (2009). Os parâmetros utilizados foram *matching weight 2, mismatching penalty 5, indel penalty 5, match probability 0.8, indel probability 0.1, score ≥ 40 e maximum period 500*, sendo:

- *matching weight*: Pontuação para *match* de base como utilizado no algoritmo de busca por similaridade de Smith e Waterman (1981).
- *mismatching penalty*: Penalidade para base diferente.
- *indel penalty*: Penalidade para *gaps* (inserções ou deleções de base)
- *match probability, indel probability*: Parâmetros padrões que não devem ser alterados.
- *score*: Pontuação mínima para a repetição ser reportada.
- *maximum period*: Tamanho máximo da unidade de repetição a ser reportada.

4.7. VARIABILIDADE DAS REPETIÇÕES ADJACENTES

Para avaliar a variabilidade genotípica das repetições, foi calculado um "*score*" de variabilidade para cada repetição utilizando o algoritmo SERV (LEGENDRE et al., 2007). Este algoritmo utiliza técnicas de aprendizagem de máquina baseada em máquina de vetores de suporte (*Support Vector Machines - SVM*) para fornecer para cada repetição uma pontuação, que é baseado no comprimento da unidade de repetição, comprimento total de repetição e pureza (perfeição da repetição).

Repetições com uma pontuação maior ou igual a 1 ($\text{VARScore} \geq 1$) são classificadas como extremamente mutáveis e repetições com pontuação entre 0 e 1 são classificadas como variáveis (DUITAMA et al., 2014).

4.8. CONSERVAÇÃO DAS REPETIÇÕES

A conservação das repetições entre as espécies foi realizada através de um alinhamento local com a ferramenta Blast com *e-value* de $1e-05$ (SHARON et al., 2005). Repetições que apresentaram homologia com ao menos uma espécie foi considerada conservada. O percentual de conservação foi realizado dividindo-se a quantidade de repetições homólogas com o total de repetições do organismo.

4.9. TENTATIVA DE ANOTAÇÃO DAS SEQUÊNCIAS

Como a maioria das sequências analisadas encontra-se anotada como hipotéticas, utilizou-se diferentes ferramentas de bioinformática para melhor compreender as bases moleculares das sequências analisadas. Para tentativa de anotação utilizou-se a ferramenta Blast2Go (CONESA et al., 2005), com *e-value* de $1e-05$ (SHARON et al., 2005), para buscar informações sobre domínios funcionais foram utilizadas as bases MIPS Funcat Pedant (RUEPP et al., 2004) e *Conserved domains database - CDD* - (MARCHLER-BAUER et al., 2011)

4.10. IDENTIFICAÇÃO DE REGIÕES REPETITIVAS EM GENES DE *Trichophyton rubrum* MODULADOS EM MEIO DE CULTURA QUE MIMETIZA A INFECÇÃO DA PELE HUMANA PELA TÉCNICA DE MICROARRAY

Para a identificação das repetições adjacentes apresentadas no perfil comparativo e em genes modulados em meio de cultura que mimetiza a infecção da pele humana pela técnica de microarray foi utilizada a ferramenta *Tandem Repeat Finder* com os parâmetros definidos no item 4.6. Para caracterização das sequências foram seguidos os passos definidos no item 4.9.

4.11. PREDIÇÃO DE ADESINAS

Tendo em vista a relação das repetições adjacentes com proteínas de adesão, buscou-se identificar genes que possam estar envolvidos com a adesão celular em dermatófitos. Para predição de adesinas utilizou-se a ferramenta *FaaPred* baseada em máquina de vetores de suporte (*Support Vector Machines - SVMs*) (RAMANA; GUPTA, 2010). A ferramenta baseou seu aprendizado em 75 adesinas de fungos descritas e características estruturais como *N-terminal carbohydrate* ou *peptide-binding domain*, domínios centrais glicosilados ricos em serina-treonina, região C-terminal e âncora glicosilfosfatidilinositol (GPI). O mesmo foi utilizado por apresentar alta acurácia e

especificidade (superior a 86%), o parâmetro de corte utilizado foi o padrão da ferramenta (*threshold* > -0,8).

4.12. PREDIÇÃO DE VIAS DE INTERAÇÃO PROTEICA

Para realizar a predição de vias de interação proteica em genes de *T. rubrum* que possuem repetições adjacentes variáveis utilizou-se a ferramenta STRING (SZKLARCZYK et al., 2015). Esta ferramenta realiza a predição de vias de interação a partir de diferentes fontes como: (i) Dados experimentais de bases de dados primárias; (ii) *Pathways* de bases de dados com curadoria manual; (iii) Mineração de texto automatizada em resumos e artigos do Pubmed; e (iv) Algoritmos para predição a partir de dados genômicos.

V. RESULTADOS E DISCUSSÃO

5.1. FERRAMENTAS PARA BUSCA DE REPETIÇÕES ADJACENTES

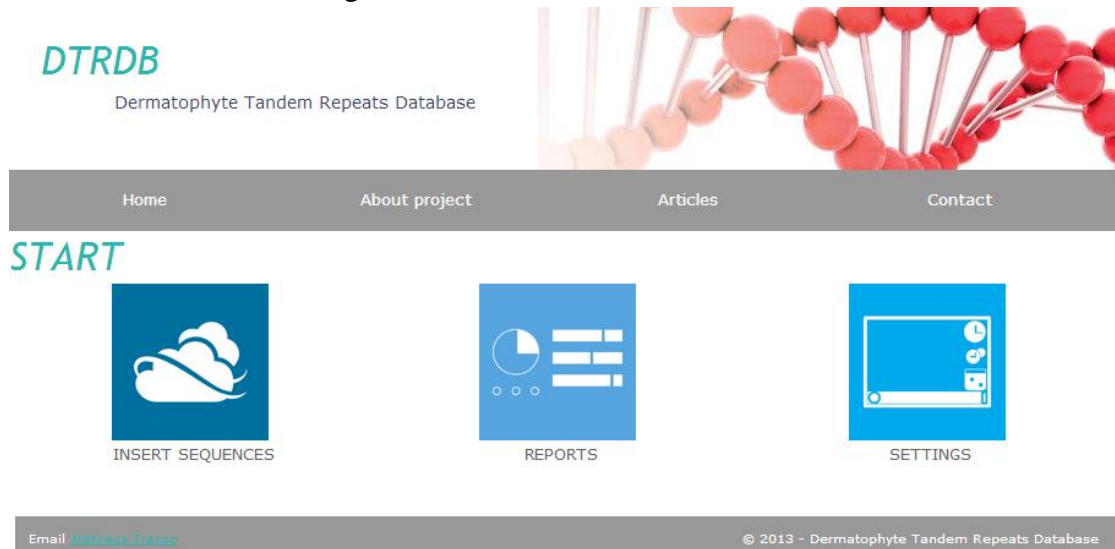
A partir das comparações realizadas, a ferramenta Mreps apresentou melhor desempenho de tempo de execução (12%), e de número de repetições encontradas (27%) maior que TRF. Considerando a análise comparativa realizada por Lim et al. (2012) e os testes realizados, inicialmente definiu-se a utilização da ferramenta Mreps para utilização no *pipeline*, pois o mesmo apresentou melhor desempenho em termos de tempo de execução e quantidade de repetições encontradas.

Entretanto, a ferramenta Mreps possui a limitação de não considerar *indels* na identificação das repetições. Além disso, observou-se através dos testes com as ferramentas que existe uma diferença significativa nos resultados devido às opções de parâmetros o que corrobora com Merkel e Gemmell (2008), mais ainda, as considerações sobre as repetições variam de acordo com a ferramenta. Por exemplo, uma repetição entre o locos 1300 e 2000 para TRF pode ser considerada como apenas uma repetição enquanto para Mreps pode representar algumas repetições com diferentes tamanhos de unidade. Assim, as duas ferramentas passaram a ser consideradas para utilização no sistema desenvolvido, podendo ser definida qual utilizar através de configurações de parâmetros.

5.2. SISTEMA PARA ARMAZENAMENTO E ANÁLISE DE REPETIÇÕES ADJACENTES EM FUNGOS DERMATÓFITOS

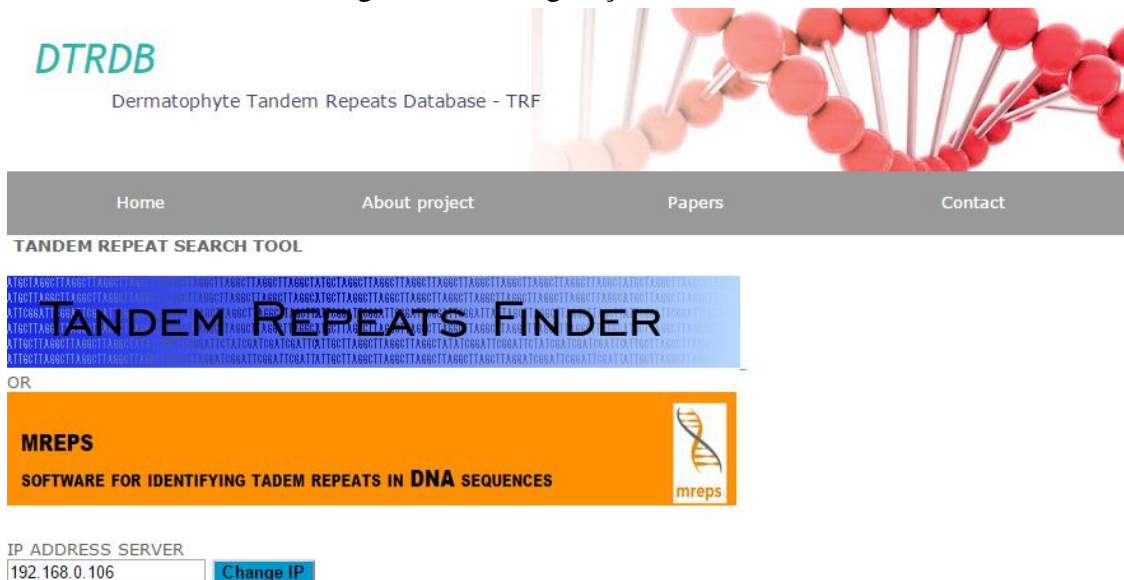
A partir da definição das ferramentas para busca de repetições adjacentes, realizou-se a implementação de um *pipeline* que possibilitou a automatização do processo de busca e armazenamento de repetições adjacentes. Dentre as diferentes tecnologias de desenvolvimento para internet disponíveis, selecionou-se o *framework* ASP.NET *Web Forms* versão 4.5 na linguagem C# pelos recursos disponíveis e conhecimento prévio na mesma. O sistema foi hospedado em um servidor Microsoft® Windows Server 2012 64 *bits* utilizando-se o Servidor Web IIS 8. Na Figura 11 apresenta-se a interface inicial do sistema desenvolvido e disponível através do endereço <http://comp.mch.ifsuldeminas.edu.br/dtrdb>.

Figura 11- Interface inicial do sistema



Na interface inicial apresentada na Figura 11 é possível selecionar uma ação a executar no sistema, como inserir novas sequências para busca e armazenamento de repetições, emitir relatórios, realizar consultas e definir configurações gerais do sistema. Na definição de configurações do sistema (Figura 12), pode-se selecionar o algoritmo a ser executado para identificação de repetições adjacentes (Mreps ou TRF - padrão) além de definir o endereço do servidor que é responsável por armazenar e executar estas ferramentas.

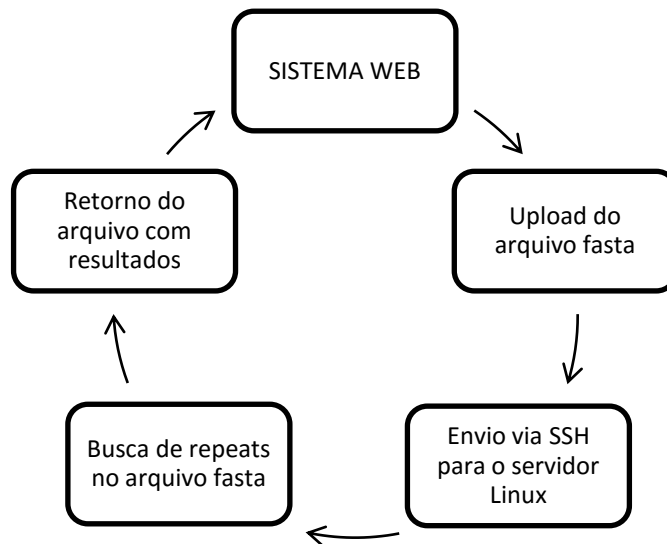
Figura 12- Configurações do sistema



A execução das ferramentas para identificação de repetições adjacentes ocorre em um servidor Ubuntu Linux versão 10.04. Tendo em vista que o *pipeline* é executado em um servidor Microsoft® Windows Server, para comunicação entre a aplicação e a ferramenta para busca de

repetições foi utilizado o serviço de *Secure Shell* - SSH⁸. O diagrama apresentado na Figura 13 descreve o fluxo de comunicação entre a aplicação e o servidor Ubuntu Linux.

Figura 13- Envio e processamento das repetições.




Através de uma interface intuitiva, o *pipeline* desenvolvido permite o envio de arquivos no formato FASTA⁹ para identificação e posterior armazenamento de repetições adjacentes, dispensando do usuário conhecimentos no sistema operacional Ubuntu Linux. Na Figura 14 apresenta-se a interface de envio e armazenamento das repetições. Na interface o usuário pode selecionar ou inserir um novo organismo, opcionalmente definir uma descrição para as sequências, o que possibilita a posterior realização de consultas com filtros. Após o processamento das sequências, em um segundo passo é possível realizar o *download* do arquivo resultante contendo as repetições identificadas, armazenar os *repeats* no banco de dados e visualizar o registro de gravação.

⁸ SSH (Secure Shell) é um protocolo de rede que realiza a conexão criptografada com outro computador na rede de forma a permitir execução de comandos de uma unidade remota.


⁹ Em bioinformática, FASTA é um formato baseado em texto para representar sequências de nucleótidos e sequências de peptídeos.

Figura 14- Interface do sistema para envio e processamento das sequências.


1º Step - Select a Fasta or Multi-Fasta file

Select the organism:  ADD Organism


File description:

 Record Sequences in DataBase


2º Step - Download XML File / Record in DataBase



DOWNLOAD XML



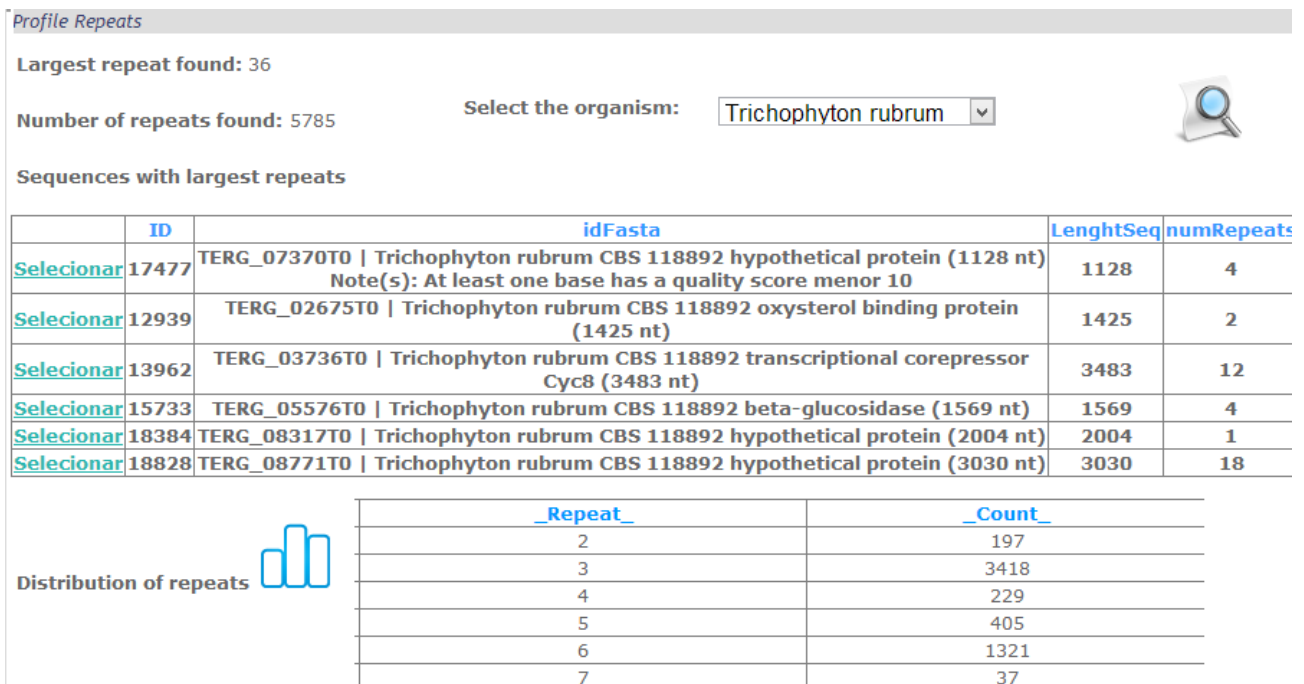
RECORD REPEATS IN DATABASE



PREVIEW REPORT RECORD

O processo de busca e armazenamento das repetições restringe-se a um ambiente de rede interna (*intranet*), sendo que as demais funcionalidades e acesso ao banco de dados podem ser realizadas através da internet. Após o processamento e armazenamento das sequências, é possível emitir relatórios e consultar as repetições identificadas. Por meio da opção relatório na interface inicial, pode-se visualizar o perfil de repetições adjacentes de um dado organismo selecionado.

Figura 15- Perfil de repetições adjacentes de um organismo selecionado.



Na Figura 15 pode-se observar o perfil de repetições adjacentes de um organismo previamente processado. Neste perfil é possível visualizar a maior repetição adjacente encontrada, o

número total de repetições, as sequências que possuem as maiores repetições do organismo, além da distribuição das repetições por tamanho de unidade do *repeat*. Quando definida a utilização pela ferramenta TRF (padrão), também é calculado um *score* de variabilidade das repetições, sendo apresentadas as dez repetições com maior *score* identificadas (Figura 16).

Figura 16- Listagem com repetições mais variáveis de um organismo.

Sequences with more variable tandem repeats

	ID	GeneName	Unit_Size	Expoent	Varscore
Selecionar	8466	TERG_08771	45	23	4,96904238114483
Selecionar	7877	TERG_07370	36	17	2,53983543943731
Selecionar	4544	TERG_00509	57	3	2,12140382557567
Selecionar	6166	TERG_03723	54	3	1,92936040916255
Selecionar	45613	TERG_08040	46	9	1,86834930386639
Selecionar	6036	TERG_03464	60	3	1,77637166435149
Selecionar	6195	TERG_03736	6	55	1,77340597682024
Selecionar	6921	TERG_05480	63	4	1,71980482796695
Selecionar	45614	TERG_08040	69	6	1,49042490857983
Selecionar	6793	TERG_05189	6	44	1,43193829376627

Outra opção do sistema é a consulta das repetições encontradas em um determinado organismo. Como observado na Figura 17, a partir da seleção de um organismo, é possível especificar itens de filtro como tamanho da unidade (*Repeat Unit*), vezes que a unidade se repete (*Expoent*), percentual de identidade (*MatchP*) e variabilidade da repetição (*VarScore*). A partir das sequências listadas, pode-se selecionar uma para que seja possível visualizar as repetições desta sequência.

Figura 17- Consulta e filtro de repetições adjacentes

Consult Repeats

Select the organism:

	Gene Name	Putative Annot	Repeats Count
Selecionar	TERG_00241T0	hypothetical protein TERG_00241	1
Selecionar	TERG_00241T1	hypothetical protein TERG_00241	1
Selecionar	TERG_00396T0	f-box domain containing protein	2
Selecionar	TERG_00509T0	protein	1
Selecionar	TERG_00768T0	serine threonine protein kinase	1
Selecionar	TERG_00789T0	hypothetical protein H100_08242	1
Selecionar	TERG_00819T0	pe repeat family protein	1

Outra funcionalidade é o processamento de uma repetição, para que sejam apresentadas somente sequências com repetições comuns a outras sequências. Na Figura 18 é apresentada uma repetição que existe em diferentes sequências armazenadas no banco de dados.

Figura 18- Repetições adjacentes encontradas em diferentes sequências

ID Fasta:TERG_01472T0

Repeat Sequence: CAGCAACAACAGCAGCAGCAGCAGCAGCAG


Repeat 4986 - CAG found in sequences below

	gene_name	repeat_consensus
Selecionar	TERG_00027T0	CAG
Selecionar	TERG_00139T0	CAG
Selecionar	TERG_00149T0	CAG
Selecionar	TERG_00226T0	CAG
Selecionar	TERG_00265T0	CAG
Selecionar	TERG_00265T0	CAG
Selecionar	TERG_00271T0	CAG

5.3. CONSULTA POR CATEGORIAS FUNCIONAIS DOS GENES DE *T. rubrum*

No sistema foram armazenadas informações sobre categorias funcionais de *T. rubrum* baseadas nas tabelas Funcat obtidas por meio dos bancos de dados MIPS PEDANT (<http://pedant.gsf.de/>) (RUEPP et al., 2004), Pfam¹⁰ (PUNTA et al., 2012) e *Gene Ontology* (ASHBURNER et al., 2000). A partir da descrição de um gene é possível consultar as repetições adjacentes existentes nele além de suas categorizações de acordo com as tabelas descritas (Figura 19).

Figura 19- Consulta categorias funcionais dos genes

Gene or Transcript ID: 

Transcripts list

ID	Description
Selecionar TERG_00838	TERG_00838T0 Trichophyton rubrum CBS 118892 zinc finger transcription factor pacC (2448 nt)

Transcript repeats

comprimento	sequence
27	CAG CAG CAG CAA CAT CAG CAT CAG CAG
27	TGC CGC TGC TGC TGC CGC CGC CGT CGC

Funcat Terms

codFuncat	description
10	CELL CYCLE AND DNA PROCESSING
11	TRANSCRIPTION
16	PROTEIN WITH BINDING FUNCTION
32	CELL RESCUE, DEFENSE AND VIRULENCE
32.1	STRESS RESPONSE
34	INTERACTION WITH THE ENVIRONMENT
43	CELL TYPE DIFFERENTIATION
34.11.03.11	pH RESPONSE

Pfam Terms

pfam_name	pfam_description
zf-C2H2	Zinc finger, C2H2 type

Gene Ontology Terms

id_go	go_desc
GO:0005737	cytoplasm
GO:0003677	DNA binding
GO:0045449	regulation of transcription
GO:0005634	nucleus
GO:0008270	zinc ion binding

¹⁰ O banco de dados Pfam é uma das coleções mais importantes de informação para a classificação de proteínas

5.4. PERFIL DE REPETIÇÕES ADJACENTES EM FUNGOS DERMATÓFITOS

Através do sistema desenvolvido é possível identificar, armazenar e consultar repetições em um dado organismo. Com o intuito de conhecer os padrões de repetições adjacentes em regiões codificantes dos principais fungos dermatófitos, utilizou-se os genes transcritos das espécies *Trichophyton rubrum*, *Trichophyton tonsurans*, *Trichophyton equinum*, *Microsporum gypseum*, *Microsporum canis*, *Arthroderma benhamiae* e *Trichophyton verrucosum* obtidos através da página da Internet do *Broad Institute* (2014). Para traçar o perfil das repetições utilizou-se os resultados oriundos da ferramenta TRF com os parâmetros descritos na metodologia.

Tabela 5- Perfil de repetições adjacentes em genes transcritos de fungos dermatófitos

	<i>T. rubrum</i>	<i>T. tonsurans</i>	<i>T. equinum</i>	<i>M. gypseum</i>	<i>M. canis</i>	<i>T. verrucosum</i>	<i>A. benhamiae</i>
Tamanho dos genes transcritos (Mbp ¹¹)	13,54	12,00	11,90	12,79	13,00	11,78	11,83
Qtde de genes transcritos ¹²	10418	8523	8679	8907	8915	8024	7980
<i>Repeats</i>	4616	4518	4634	4829	3724	6536	6720
Conservação ¹³	19,50%	43,25%	42,25%	2,96%	2,44%	22,77%	22,02%
Maior unidade de repetição	228	405	309	378	296	220	233

Para melhor entender a biologia comparativa destes importantes fungos patogênicos, primeiramente levantou-se a distribuição das repetições. Na Tabela 5 apresenta-se o total de genes transcritos, quantidade de repetições adjacentes identificadas em cada organismo, percentual de conservação das repetições e a maior unidade de repetição.

Foram identificadas repetições com diferentes tamanhos de repetição: microssatélites ou *short repeats* (1 a 9-pb), minissatélites (10 a 100-pb) e megassatélites ou repetições longas (acima 100-pb por unidade de repetição). Algumas repetições estão distribuídas em sequências já caracterizadas, porém a maioria das repetições ocorre em sequências ainda não caracterizadas.

A quantidade de repetições identificadas varia de 3724 para *M. canis* até 6720 para *A. benhamiae*. Desta maneira, observou-se que não existe correlação entre o tamanho das sequências com a quantidade de repetições adjacentes, pois *T. rubrum* apresenta 10416 genes transcritos (13,54

¹¹ milhões de pares de bases ou megabases

¹² quantidade de sequências de genes transcritos obtidos pela página do *Broad Institute* em outubro de 2014

¹³ percentual de conservação das repetições em relação as demais espécies

Mbp), sendo identificados 4616 repetições, enquanto *A. benhamiae* apresentou 6720 repetições em 7980 sequências (11,83 Mbp), algo também descrito no trabalho de Mayer; Leese e Tollrian (2010).

Tabela 6- Ocorrência das unidades de repetições adjacentes e abundância relativa*

Unidade	<i>T. rubrum</i>	<i>T. tonsurans</i>	<i>T. equinum</i>	<i>M. gypseum</i>	<i>M. canis</i>	<i>T. verrucosum</i>	<i>A. benhamiae</i>
2	-	-	11 (0,8)	-	-	74 (5,5)	91 (6,7)
3	471 (34,8)	536 (39,6)	527 (38,9)	392 (29)	259 (19,1)	848 (62,6)	902 (66,6)
4	-	17 (1,3)	23 (1,7)	19 (1,4)	-	65 (4,9)	67 (4,9)
5	14 (1,3)	10 (0,7)	20 (1,5)	25 (1,8)	-	70 (5,2)	59 (4,4)
6	459 (33,9)	498 (36,8)	513 (37,9)	554 (40,9)	271 (20,1)	684 (50,5)	710 (52,4)
7	14 (1,3)	27 (2)	29 (2,1)	34 (2,5)	12 (0,9)	68 (5,2)	67 (4,9)
8	24 (1,8)	42 (3,1)	45 (3,3)	39 (2,9)	23 (1,7)	136 (10,4)	122 (9,1)
9	578 (42,7)	561 (41,4)	561 (41,4)	584 (43,1)	420 (31,2)	728 (53,8)	713 (52,7)
10	54 (4)	62 (4,6)	77 (5,7)	71 (5,2)	50 (3,7)	139 (10,3)	127 (9,4)
11	101 (7,5)	108 (8)	115 (8,5)	122 (9,1)	124 (9,2)	221 (16,3)	241 (17,8)
12	927 (68,5)	874 (64,5)	881 (65,7)	919 (67,9)	757 (56)	1050 (77,5)	1087 (80,3)
13	153 (11,3)	119 (8,8)	128 (9,5)	147 (10,9)	130 (9,6)	192 (14,2)	203 (15)
14	129 (9,5)	119 (8,8)	127 (9,4)	136 (10,4)	109 (8,5)	186 (13,7)	189 (14)
15	497 (36,8)	460 (34)	464 (34,3)	501 (37,1)	423 (31,2)	576 (42,5)	584 (43,1)
16	92 (6,8)	74 (5,5)	75 (5,5)	71 (5,2)	67 (4,9)	116 (8,6)	132 (9,7)
17	44 (3,2)	46 (3,4)	54 (4)	56 (4,1)	53 (3,9)	89 (6,6)	89 (6,6)
18	359 (26,5)	299 (22,8)	316 (23,3)	352 (26)	298 (22,9)	395 (29,2)	402 (29,7)
19	33 (2,4)	27 (2)	31 (2,3)	26 (1,9)	23 (1,7)	54 (4)	75 (5,5)
20	30 (2,2)	25 (1,8)	28 (2,7)	43 (3,2)	21 (1,6)	57 (4,3)	61 (4,6)
21	211 (15,6)	190 (14,3)	199 (14,7)	191 (14,2)	191 (14,2)	220 (16,2)	212 (15,7)
22	20 (1,5)	29 (2,1)	26 (1,9)	40 (3)	22 (1,6)	43 (3,2)	45 (3,3)
23	23 (1,7)	18 (1,3)	16 (1,2)	24 (1,8)	20 (1,5)	37 (2,7)	36 (2,7)
24	121 (8,9)	105 (7,8)	99 (7,3)	120 (8,9)	129 (9,5)	144 (10,6)	148 (10,9)
25	-	-	-	12 (0,9)	-	21 (1,6)	18 (1,3)
26	-	-	-	11 (0,8)	-	20 (1,5)	19 (1,4)
27	53 (3,9)	46 (3,4)	45 (3,3)	66 (4,9)	45 (3,3)	64 (4,7)	55 (4,6)
28	-	-	-	-	-	11 (0,8)	20 (1,5)
29	-	-	-	-	-	11 (0,8)	-
30	43 (3,2)	53 (3,9)	43 (3,2)	38 (2,9)	35 (2,6)	48 (3,5)	53 (3,9)
31	-	-	-	-	-	-	10 (0,7)
33	23 (1,7)	17 (1,3)	22 (1,6)	30 (2,2)	20 (1,5)	17 (1,3)	18 (1,3)
36	20 (1,5)	20 (1,5)	22 (1,6)	17 (1,3)	21 (1,6)	31 (2,3)	22 (1,6)
39	-	-	-	18 (1,3)	12 (0,9)	11 (0,8)	17 (1,3)
42	17 (1,3)	14 (1,3)	12 (0,9)	13 (1)	-	14 (1,3)	19 (1,4)
45	15 (1,2)	11 (0,8)	-	13 (1)	-	-	-
48	-	-	-	14 (1,3)	-	-	-
51	-	-	10 (0,7)	-	-	-	-

* Abundância relativa (em parenteses) é o total de *repeats* por megabase da sequência analisada.

Através do sistema é possível obter a distribuição dos *repeats* por unidade de repetição. A Tabela 6 apresenta a quantidade de repetições por tamanho de unidade que tenham ocorrido ao menos 10 vezes. Para obter-se a abundância relativa em megabase foi realizada a divisão da quantidade de *repeats* pelo tamanho dos genes transcritos em megabase (Mbp).

Pôde-se observar que a maior quantidade de repetições adjacentes nos genes transcritos dos fungos dermatófitos estão agrupados em unidades de repetição que são divisíveis por três, assim o resultado sugere que as repetições mais prevalentes não alteram a janela de leitura (METZGAR; BYTOF; WILLS, 2000).

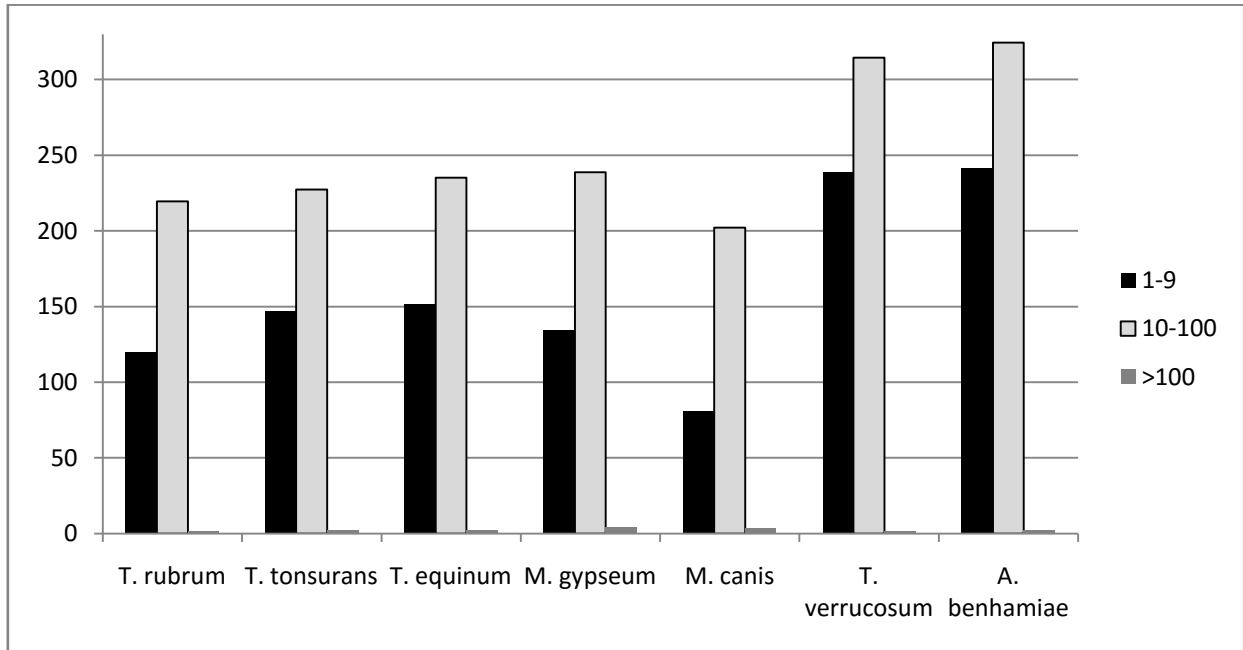
Realizou-se o agrupamento dos resultados de acordo com suas unidades de repetições em 1 a 9-bp, 10 a 100-bp e acima de 100-bp (Figura 20). Através do agrupamento pôde-se visualizar uma prevalência de repetições de minissatélites, especialmente entre 10 a 40-pb por unidade, observando-se ainda que conforme a unidade de repetição aumenta as ocorrências diminuem, algo também citado pelo trabalho de Gibbons e Rokas (2009), no qual os autores realizaram uma análise de repetições adjacentes em dez espécies de *Aspergillus*.

Além disso, os resultados demonstram conservação nos padrões de unidades de repetição nos dermatófitos analisados, o que apoia o trabalho de Martinez et al. (2012) sobre a alta proporção de ortólogos compartilhados entre as espécies de dermatófitos. Apenas *T. verrucosum* e *A. benhamiae* apresentam uma maior quantidade de repetições que as outras cinco espécies, porém mantendo o mesmo padrão na distribuição por unidades de repetições.

No que tange a quantidade de repetições identificadas, não existe outro levantamento em dermatófitos, porém existe uma grande variabilidade nas considerações das repetições e nos parâmetros utilizados em diferentes trabalhos como descrito por Leclercq; Rivals e Jarne (2007). No trabalho de Karaoglu; Le e Meyer, (2005) os autores realizaram um levantamento de repetições adjacentes curtas (1 a 6-bp por unidade de repetição) perfeitas com extensão mínima de 10-pb no genoma de nove espécies de fungos utilizando um programa próprio baseado em Python¹⁴. Como resultado, no genoma de *Neurospora crassa* (38 Mbp) foram identificadas 14.319 repetições com uma abundância relativa de 377 repetições por megabase. Outro trabalho identificou 13.292 repetições curtas (1 a 6-bp por unidade de repetição) no genoma de *Neurospora crassa*, porém considerando também repetições imperfeitas utilizando a ferramenta *Phobos* desenvolvida pelos autores (MAYER; LEESE; TOLLRIAN, 2010).

¹⁴ Linguagem de programação

Figura 20- Abundância relativa das repetições agrupadas



Observando-se as repetições consenso mais abundantes, os padrões nos sete genomas transcritos são similares. O *repeat* CAG é prevalente em todas as espécies de dermatófitos, o que ocorre também no genoma de *Puccinia triticina* de acordo com Singh et al. (2011). O *repeat* CAG também é o mais frequente *Drosophila* de acordo com Huntley e Clark (2007), no qual o genoma de 12 diferentes organismos foram analisados mostrando que o *repeat* CGA é o mais prevalente em regiões codificantes. Na Tabela 7 apresenta-se os padrões de repetição mais prevalentes (acima de 20 ocorrências consenso) nas sete espécies de dermatófitos.

Tabela 7- Padrões de repetição consenso mais prevalentes

<i>T. rubrum</i>		
Repetição consenso	Tamanho Unidade	Quantidade
CAG	3	125
GCA	3	51
CAA	3	35
CAGCAGCAA	9	28
AGC	3	25
<i>T. tonsurans</i>		
Repetição consenso	Tamanho Unidade	Quantidade
CAG	3	98
GCA	3	48
CAA	3	46
AGA	3	31
AGC	3	27
GAA	3	25

<i>T. equinum</i>		
Repetição consenso	Tamanho Unidade	Quantidade
CAG	3	96
CAA	3	46
GCA	3	43
GAA	3	31
AGA	3	28
AGC	3	26
<i>M. gypseum</i>		
Repetição consenso	Tamanho Unidade	Quantidade
CAG	3	81
GAA	3	38
GCA	3	34
AGA	3	29
CAA	3	22
<i>M. canis</i>		
Repetição consenso	Tamanho Unidade	Quantidade
CAG	3	71
GAA	3	25
<i>T. verrucosum</i>		
Repetição consenso	Tamanho Unidade	Quantidade
CAG	3	109
GAA	3	94
AGA	3	73
AAG	3	60
TCT	3	48
GCA	3	45
CAA	3	44
CTT	3	36
TTC	3	34
AGC	3	29
CTG	3	25
ACA	3	22
AG	2	22
CAGCAA	6	21
<i>A. benhamiae</i>		
Repetição consenso	Tamanho Unidade	Quantidade
CAG	3	123
GAA	3	105
AGA	3	73
AAG	3	56
TCT	3	47
CTT	3	46
CAA	3	42
GCA	3	42
TTC	3	33

CTG	3	30
AG	2	29
TC	2	26
CAGCAA	6	21

5.5. REPETIÇÕES ADJACENTES NO DERMATÓFITO *T. rubrum*

Levando-se em consideração a importância clínica do dermatófito *T. rubrum* já descrito neste trabalho, é apresentado na Figura 21 um quadro com a descrição das repetições adjacentes identificadas neste organismo.

Figura 21- Repetições adjacentes em *T. rubrum*

Total de genes transcritos (inclui isoformas)	Total de repeats (genes)	Repeats em genes transcritos já anotados (genes)	Repeats em genes transcritos hipotéticos (genes)
10418	4616 (2348)	425 (273)	4191 (2075)
	Microsatélites (genes)	Minissatélites (genes)	VNTR (Micro/Minissatélites)
	1571 (832)	3045 (1597)	795 (717/78)

Observando as repetições adjacentes identificadas no dermatófito *T. rubrum* obteve-se um total de 4616 *repeats*. Estas repetições estão distribuídas em 2348 sequências do total de 10418 genes transcritos armazenados, o que representa uma densidade de 22,53% de repetições nas sequências de genes transcritos. Do total de 4616 *repeats*, 4191 foram identificados em 2075 sequências hipotéticas, sendo o restante de 425 *repeats* identificados em 273 sequências já anotadas. Desta maneira, as repetições adjacentes concentram-se prevalentemente em genes hipotéticos.

No que tange a distribuição das repetições entre microsatélites e minissatélites, observa-se a prevalência de repetições com extensão maior que 10-pb, porém, quando observado apenas repetições potencialmente variáveis (VNTR), existe a ampla prevalência de repetições curtas (microsatélites).

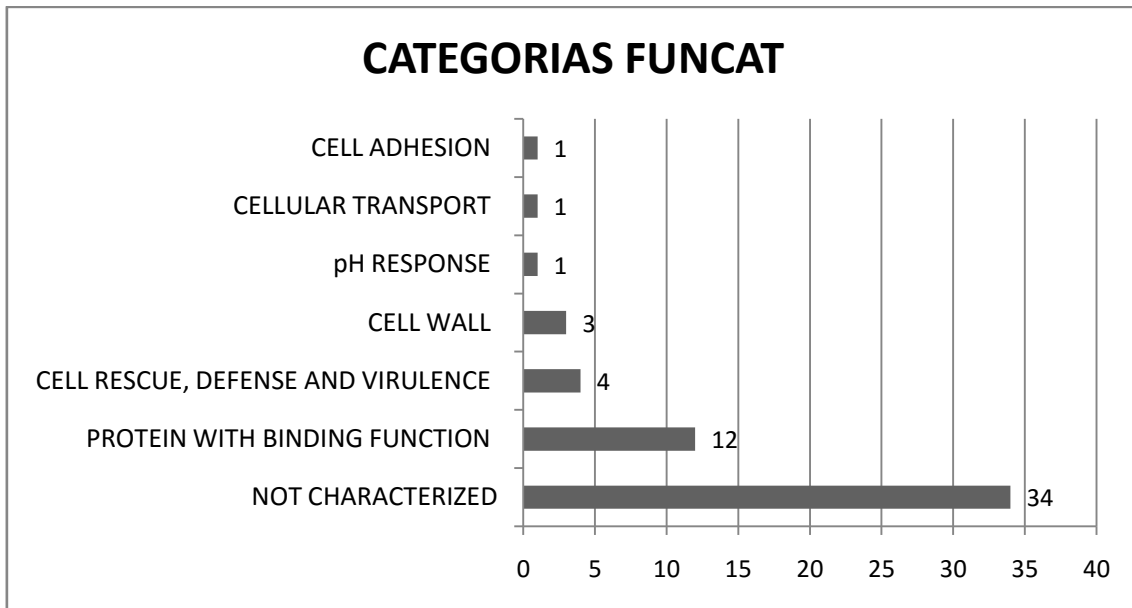
5.6. CATEGORIZAÇÃO FUNCIONAL DAS REPETIÇÕES ADJACENTES NO DERMATÓFITO *T. rubrum*

Tendo em vista que a maioria das repetições concentra-se em sequências hipotéticas, utilizou-se bases de dados funcionais para categorização dos genes transcritos que possuem repetições adjacentes. Inicialmente utilizou-se as categorias definidas na base de dados MIPS FunCat PEDANT (RUEPP et al., 2004) disponível em <http://pedant.gsf.de/>. Realizou-se a seleção de genes com unidades de repetição variáveis maiores ou iguais a 10-pb e *varscore* ≥ 0

(minissatélites variáveis). Assim, obteve-se um total 78 repetições em 54 genes para categorização. Os dados das repetições e identificação dos genes são apresentados na Tabela 9.

A Figura 22 apresenta a classificação funcional dos genes com unidades de repetições adjacentes variáveis maiores que 10-pb de acordo com a tabela Funcat, sendo que um determinado gene pode estar relacionado a mais de uma categoria.

Figura 22- Classificação funcional dos genes transcritos com unidades de repetições adjacentes variáveis maiores que 10-pb de acordo com a tabela Funcat.



A Tabela 8 apresenta a identificação de genes e categorias de maior interesse de acordo com a tabela Funcat.

Tabela 8- Categorias e genes com repetições de minissatélites de maior interesse

Categoria	ID da Sequência
CELL ADHESION	TERG_08771
CELL RESCUE, DEFENSE AND VIRULENCE	TERG_00153, TERG_03464, TERG_05663, TERG_07180
CELL WALL	TERG_00768, TERG_00890, TERG_05576
pH RESPONSE	TERG_05663
PROTEIN WITH BINDING FUNCTION	TERG_00819, TERG_00962, TERG_01459, TERG_02482, TERG_02675, TERG_05956, TERG_06432, TERG_07482, TERG_07711

A Tabela 9 apresenta as repetições consenso encontradas nas sequências com unidades de repetição maior ou igual a 10-pb.

Tabela 9- Repetições adjacentes variáveis com unidade ≥ 10 -pb em *T. rubrum*

ID da Sequência	Unidade	Repetições	Varscore	Repetição consenso
TERG_00153T0	36	2	0,14	CAGCACCTCCGCCCCCTCCAGCACCTCCTGGTAGTG
TERG_00241T0	63	9	1,34	GAGTTGACGAGTCCTTTGGCAAGGTAGACAACCTTACATGAGGACATGAAAGCCTTTCTCGGAC
TERG_00396T0	18	8	0,05	GGGTGAATGGAGTGAAT
TERG_00396T0	45	4	0,47	AAATGGAGTGAATGGGACGAACGGAGTGAATGGGACGAACGGGGT
TERG_00509T0	57	3	2,12	GATACATGGACAGCGCAGCAAAGCAACCTGAACTCCGTGCCTGTCTCTCTTAACACC
TERG_00768T0	12	7	0,33	CTGAGGAAAACA
TERG_00789T0	12	9	0,19	ATGATGAAGAA
TERG_00819T0	15	41	1,35	CTGCTGCTGAAGAGC
TERG_00890T0	15	12	0,30	CAGCAGCAACAG
TERG_00890T0	24	15	0,72	CAGCAGCAACCCGGACAGCCACTG
TERG_00890T0	45	7	0,69	CAGCAGCAACCTGGACAGGCATTGCAACAACCACCACCACCACAG
TERG_00890T0	48	7	0,81	CAGCAGCAACCTGGACAGGCATTGCAACAACCACCACCACCACCACAG
TERG_00908T0	27	11	0,66	CCAGCAGAAAATACTCTGGCAGAGATGA
TERG_00908T0	54	6	0,71	CTGCGGCAAAGATGACCAGCAGAAATTCTCTGGCAAGGATGACCAGAAATT
TERG_00962T0	12	8	0,18	AGAAGGAAAAGA
TERG_01015T0	15	10	0,83	AACCTCCCTCGCAGC
TERG_01459T0	18	6	0,07	AGCCAGAGGAGCCAACGG
TERG_01472T0	12	10	0,29	CAGCAGCAGCTC
TERG_02149T0	78	2	0,63	CATAAGTCTCTGATGGATACCGTTCACAGGTCTCTGATAGACACTAGCACAAAGCCTCTATCTCTAAAGGGATACTACT
TERG_02482T0	15	11	0,53	CAACAGCCACAACAG
TERG_02675T0	12	8	0,38	CTGGCGATGCTG
TERG_03122T0	13	6	0,15	TATACTACGAGAT
TERG_03164T0	12	12	0,47	GGAGTAGAAGAG
TERG_03355T0	15	7	0,33	AATCCCAGGCTACTA
TERG_03355T0	15	10	0,76	CTAAATCCCTGGCCG
TERG_03464T0	60	3	1,78	GGCTCCTCCGGCCGTGGCAATGATTCTATGGCTCCAAGTCCAGGGACAACGACTCCCAC
TERG_03723T0	54	3	1,93	CAAATGTGAGTGCCAATCGATCTGGCACAAACACCATTGCCAGTCGGCCAGCGG

ID da Sequência	Unidade	Repetições	Varscore	Repetição consenso
TERG_04269T0	12	11	0,61	ACCAGCAGCAGC
TERG_04269T0	15	9	0,89	CAGCATCACCAGCAG
TERG_04269T1	12	11	0,61	ACCAGCAGCAGC
TERG_04269T1	15	9	0,89	CAGCATCACCAGCAG
TERG_04287T0	12	8	0,33	ATCACCACCATC
TERG_05189T0	18	10	0,16	AGACAGAGATAGAGAAAG
TERG_05189T0	24	9	0,30	GAGACAGAGACAGAGACGAGATC
TERG_05189T0	72	4	0,09	GACAGGGAGAGAGACCGAGACAGAGAAAGAGATCGAGATAGAGACAGAGACAGGGACAGGTCCGAGATAGA
TERG_05393T0	15	6	0,13	GAGGAACCCAAAGCC
TERG_05479T0	12	10	0,28	CAGCAGCAGGGC
TERG_05480T0	21	12	0,71	ATGAAGAAGAGAGACGTCGTC
TERG_05480T0	63	4	1,72	GAAGAGGAGAGACGTCGTCACGAAGAAGAGAGACTCCGTTATGAAGAAGAGAAACGCCGTCTC
TERG_05576T0	18	5	0,00	AGGCCCTGAACCACCAA
TERG_05644T0	12	9	0,65	GCTCCCTACTTT
TERG_05644T0	12	11	0,53	ACTGGAAGTCC
TERG_05663T0	63	2	0,20	GGTCAACTTACTCAAGGGACTACTCAACCCACTCAAGGGGCCAGTCAACCAACCCAAGGGGGT
TERG_05732T0	18	6	0,24	CTGAGACCCAACTACCT
TERG_05733T0	15	16	1,11	AAGGACTATGGCCAC
TERG_05956T0	12	12	0,39	GAAGACGAAGAA
TERG_05956T0	15	10	0,54	AGGAGGATGAAGAAG
TERG_06432T0	12	24	1,11	CAGCAGCACCCA
TERG_06432T0	18	10	0,04	CACCCACAGCATCACCT
TERG_06432T0	81	2	0,03	CAGCATCAACAGCATCCGCACCACCCACAGCACCCACAGCAGCAACAACAACCACAGCCGCAACCTCAGCCGCAACAT CCC
TERG_06432T1	12	24	1,11	CAGCAGCACCCA
TERG_06432T1	18	10	0,04	CACCCACAGCATCACCT
TERG_06432T1	81	2	0,03	CAGCATCAACAGCATCCGCACCACCCACAGCACCCACAGCAGCAACAACAACCACAGCCGCAACCTCAGCCGCAACAT CCC
TERG_06691T0	63	2	0,11	GAACCAAAGAAGCTCGAGGAAATAAAGAACCACAAGAACTAGACGAACTGAAAGAATCGGTG
TERG_07042T0	33	18	1,37	GAGACTCCTATTCCAACCTCAGCCAACAGGAACGGGACT

ID da Sequência	Unidade	Repetições	Varscore	Repetição consenso
TERG_07042T0	33	16	1,20	GGAAGTACTGAGACTCCTATTCTACTCAACCAACA
TERG_07042T0	72	8	0,27	GGAACAGAGACTCCTATTCCAAGTCTAGCAACAGGAACTGAAATCCCTATTCTACTCAACCAACA
TERG_07180T0	12	7	0,20	CGATGATGAGGA
TERG_07370T0	18	33	2,54	CTCCTCTATCACTCAAG
TERG_07370T0	36	17	1,11	CCTTCTATTACCCAAGCTCCTCTATCACTCAAGCC
TERG_07482T0	60	2	0,92	GGTTCCTACGGAGACTCGCCTGACGATTGCTATAATGATTCTACGATGATTACTACGGA
TERG_07711T0	12	6	0,16	TACTACTACGA
TERG_08141T0	15	6	0,05	CAACAACAACAGCAA
TERG_08194T0	21	8	0,44	GAAGAAGAACGACAACAGGCA
TERG_08252T0	78	2	0,76	GACCCTCCCCGAAAACAAAACCACTTCCACCAGGCTTGTTCACGACTGGGGCAACAACCACGCTCAAAGCGCAAT
TERG_08421T0	15	16	1,03	AATGATCCGCCAAG
TERG_08569T0	27	9	0,40	CCAGTCGGCCTGGCACAACGCGGGTG
TERG_08569T0	81	2	0,22	CCAGTCGATCTGGCACAACGCGGGTGCCAATCGAGCCGGCACAACGCCAGCACCAATCGGCCTAGCACAACGCGGGTG
TERG_08724T0	12	14	0,83	CCAGCAACACCA
TERG_08771T0	12	11	0,92	CGGAGGAGGCAT
TERG_08771T0	18	57	1,02	GACAGCCAGGACAGCCAG
TERG_08771T0	45	23	4,97	GACAGCCAGGACAGCCAGGACAGCCAGGTCAGCCAGGACAGCCAG
TERG_11526T0	75	3	0,09	ACACCTGTGGAAATGACTGTAAAGAGTGCAAAGCACCTAAGCAATGCATCAATGGCCAATGTACTTGCCCATAGG
TERG_11718T0	12	7	0,23	ATCACCAGCAGT
TERG_11764T0	12	8	0,50	GTAGGGCTGTTG
TERG_11764T0	24	12	0,04	GGTAGGAGAGTTGGTAGGGGAATT
TERG_12443T0	10	7	0,10	ACACACCAT
TERG_12724T0	42	2	0,88	TTCGTAAGTGGTACATTGGCCGCCTAGGAGATGATGGCAATA

Além da categorização funcional através da tabela Funcat, também utilizou-se as categorias funcionais descritas no *Gene Ontology* (ASHBURNER et al., 2000), para isto realizou-se o mapeamento dos genes transcritos do *T. rubrum* CBS utilizando-se a ferramenta Blast2Go (CONESA et al., 2005). A Tabela 10 apresenta a lista de genes com repetições variáveis anotadas através da ferramenta Blast2Go.

Tabela 10- Anotação dos genes transcritos com repetições de variáveis de minissatélites.

ID da Sequência	Anotação	Bases	GOs
TERG_00153T0	actin associated protein wsp1	1935	F:cytoskeletal protein binding
TERG_00241T0	hypothetical protein	7065	-
TERG_00396T0	f-box domain containing protein	1839	-
TERG_00509T0	hypothetical protein	2763	-
TERG_00768T0	serine threonine protein kinase	7248	P:signal transduction; F:ion binding; P:cellular protein modification process; P:cellular amino acid metabolic process; P:cellular nitrogen compound metabolic process; P:biosynthetic process; C:protein complex; C:cellular_component; F:kinase activity; F:signal transducer activity
TERG_00789T0	hypothetical protein	1521	-
TERG_00819T0	pe repeat family partial	4356	F:DNA binding; P:chromosome organization; P:DNA metabolic process
TERG_00890T0	cell surface	2145	F:molecular_function
TERG_00908T0	hypothetical protein	2676	-
TERG_00962T0	nucleolar protein 58	1566	P:biological_process; F:kinase activity
TERG_01015T0	domain partial	927	P:catabolic process; P:cell wall organization or biogenesis
TERG_01459T0	conserved leucine-rich repeat protein	5265	F:molecular_function
TERG_01472T0	hypothetical protein	1374	-
TERG_02149T0	hypothetical protein	1155	-
TERG_02482T0	transcription initiation factor tfiid subunit 12	1965	F:translation factor activity
TERG_02675T0	oxysterol binding protein	1425	-
TERG_03122T0	integral membrane protein	558	F:nuclease activity; P:nucleobase-containing compound catabolic process; P:DNA metabolic process; C:cellular_component
TERG_03164T0	hypothetical protein	516	-
TERG_03355T0	hypothetical protein	1185	-
TERG_03464T0	stress protein ddr48	966	-
TERG_03723T0	uncharacterized kda protein in rps4-rps11 intergenic region- partial	669	-
TERG_04269T0	hypothetical protein	1794	-
TERG_04269T1	hypothetical protein	1641	-
TERG_04287T0	hypothetical protein	1935	-
TERG_05189T0	arginine rich protein	2205	-

ID da Sequência	Anotação	Bases	GOs
TERG_05393T0	hypothetical protein	615	-
TERG_05479T0	bzip transcription	1086	F:DNA binding; F:nucleic acid binding transcription factor activity; P:cellular nitrogen compound metabolic process; P:biosynthetic process; C:protein complex; C:intracellular
TERG_05480T0	hypothetical protein	2568	F:molecular_function
TERG_05576T0	cell wall glucanase	1569	P:carbohydrate metabolic process; F:molecular_function
TERG_05644T0	mucin family signaling protein	2934	-
TERG_05663T0	bacteriodes thetaiotaomicron symbiotic chitinase	4053	F:chitinase activity F:chitin binding P:carbohydrate metabolic process P:chitin catabolic process
TERG_05732T0	serine-threonine rich	891	-
TERG_05733T0	salivary glue protein sgs-3	627	-
TERG_05956T0	rrna processing protein ebp2p	1206	-
TERG_06432T0	c2h2 transcription factor	1857	F:DNA binding; F:ion binding
TERG_06432T1	c2h2 transcription factor	1464	F:DNA binding; F:ion binding
TERG_06691T0	plasmodium exported protein	1191	-
TERG_07042T0	hypothetical protein	2052	-
TERG_07180T0	dna binding protein ure-b1	12015	P:cellular protein modification process; F:ligase activity
TERG_07370T0	major outer envelope glyco protein	1146	P: metabolic process
TERG_07482T0	ribonuclease h1	780	F:nuclease activity; P:cellular nitrogen compound metabolic process; F:molecular_function
TERG_07711T0	ras guanine-nucleotide exchange protein	3549	P:cell division; P:small molecule metabolic process; P:signal transduction; P:nucleobase-containing compound catabolic process; F:enzyme regulator activity; C:intracellular
TERG_08141T0	bzip transcription factor	1947	F:DNA binding; F:nucleic acid binding transcription factor activity; P:cellular nitrogen compound metabolic process; P:biosynthetic process; C:protein complex; C:intracellular
TERG_08194T0	hypothetical protein	2322	P:carbohydrate metabolic process; P:lipid metabolic process; F:ion binding; P:biosynthetic process; C:cellular_component; F:molecular_function
TERG_08252T0	domain partial	591	P:catabolic process; P:cell wall organization or biogenesis
TERG_08421T0	ph signal transduction protein	1893	-
TERG_08569T0	hypothetical protein	1020	-
TERG_08724T0	hypothetical protein	750	-
TERG_08771T0	extracellular serine-threonine rich protein	2946	F:molecular_function; P:biological_process; F:hydrolase activity; P:metabolic process
TERG_11526T0	tenascin c	393	-

ID da Sequência	Anotação	Bases	GOs
TERG_11718T0	hypothetical protein	645	-
TERG_11764T0	hypothetical protein	480	-
TERG_12443T0	hypothetical protein	390	-
TERG_12724T0	nonribosomal peptide synthase	4761	P:biological_process; F:molecular_function

Dos 54 genes transcritos que possuem repetições variáveis de minissatélites, 22 destes são hipotéticos, enquanto outros foram anotados estando relacionados a diferentes processos, como o gene TERG_08771 anotado como *extracellular serine-threonine rich protein*, que de acordo com Hevia, et al. (2013) pode estar relacionado à adesão celular, o que corrobora com a categorização realizada através da tabela Funcat. Além disso, o gene citado contém o domínio repetitivo *Collagen triple helix repeat* conservado, que de acordo com estudos já realizados mostra-se relacionado à adesão celular (PARK et al., 2013).

5.7. DOMÍNIOS REPETITIVOS EM *T. rubrum*

Muitas proteínas, especialmente em eucariotos, contêm repetições adjacentes. Estas repetições possuem uma variedade de propriedades de ligação e estão envolvidas em interações proteína-proteína. Domínios repetitivos são frequentemente curtos, embora grandes repetições com mais de dez domínios em conjunto ocorra em eucariotos (BJORKLUND; EKMAN; ELOFSSON, 2006).

Muitos organismos parasitas possuem antígenos de superfície que são compostos de repetições de aminoácidos. Ao formar uma interface entre o hospedeiro e o agente patogênico, estas proteínas repetitivas podem atuar como fatores de virulência e se envolver na evasão imunológica e na aderência (NIKLAUS et al., 2007). Desta maneira, neste trabalho também se buscou identificar genes com domínios repetitivos em *T. rubrum* utilizando os domínios definidos na base de dados PFAM¹⁵(PUNTA et al., 2012).

Tabela 11- Genes com domínios repetitivos em *T. rubrum*

Domínio	Qtde	ID Sequência
Ankyrin repeat	12	TERG_00445, TERG_00609, TERG_00873, TERG_01424, TERG_02303, TERG_02743, TERG_03443, TERG_05213, TERG_05757, TERG_06067, TERG_07650, TERG_07788
Bacterial signalling protein N terminal repeat	1	TERG_04696
Bacterial transferase hexapeptide	1	TERG_04687
Chlamydial polymorphic outer membrane protein repeat	1	TERG_03207
Collagen triple helix repeat	1	TERG_08771
DnaJ central domain (4 repeats)	2	TERG_05280, TERG_06104
HAT (Half-A-TPR) repeat	2	TERG_06734, TERG_08441
HEAT repeat	8	TERG_00492, TERG_00652, TERG_01072, TERG_02923, TERG_02937, TERG_06010, TERG_07933, TERG_08083
Hr1 repeat	2	TERG_00630, TERG_06340
Leucine Rich Repeat	1	TERG_03048
PPR repeat	1	TERG_08841
PT repeat	1	TERG_01828
Paired amphipathic helix repeat	1	TERG_05578
Plexin repeat	1	TERG_07962
Prenyltransferase and squalene oxidase repeat	3	TERG_01630, TERG_02234, TERG_07136
Proteasome/cyclosome repeat	1	TERG_05485
Pumilio-family RNA binding repeat	3	TERG_01212, TERG_04136, TERG_04970
RNA polymerase Rpb1 C-terminal repeat	1	TERG_07396
Repeat in ubiquitin-activating (UBA) protein	1	TERG_08280
Sel1 repeat	4	TERG_01288, TERG_02519, TERG_05574, TERG_06928
Supt5 repeat	1	TERG_05483
Tetratricopeptide repeat	6	TERG_00518, TERG_02590, TERG_03736, TERG_04260, TERG_06870, TERG_08417
Transcription factor TFIIIB repeat	1	TERG_07478
WD domain. G-beta repeat (WD40)	31	TERG_00023, TERG_00220, TERG_00539, TERG_01098, TERG_01449, TERG_01588, TERG_01976, TERG_02326, TERG_02390, TERG_02452, TERG_02611, TERG_02853, TERG_03394, TERG_03587, TERG_03716, TERG_03993, TERG_04002, TERG_04321, TERG_04337, TERG_04435, TERG_04861, TERG_05501, TERG_05844, TERG_06007, TERG_06121, TERG_06214, TERG_06513, TERG_06952, TERG_07108, TERG_07253, TERG_07746
pentatricopeptide repeat domain	4	TERG_01814, TERG_02327, TERG_03717, TERG_08841

5.8. IDENTIFICAÇÃO DE REGIÕES REPETITIVAS EM GENES DE *Trichophyton rubrum* MODULADOS EM MEIO DE CULTURA QUE MIMETIZA A INFECÇÃO DA PELE HUMANA PELA TÉCNICA DE MICROARRAY

Para entender a complexa interação do fungo com o hospedeiro, identificou-se genes diferencialmente expressos por técnicas de expressão gênica em larga escala, simulando a infecção da pele humana através da utilização de meio de cultura mínimo incluindo queratina ou elastina como fonte de proteína. Nosso grupo de pesquisa (Grupo de Biotecnologia de microorganismos da Unidade de Biotecnologia, UNAERP) realizou um experimento de *microarray* em duas condições de cultivo, obtendo-se um total de 212 sondas moduladas em meio mínimo adicionado de elastina e 243 sondas moduladas em meio mínimo adicionado de queratina (resultados não publicados).

Com intuito de analisar as regiões repetitivas em sequências moduladas nas duas situações de cultivo, utilizou-se o programa local Blast com *e-value* $\leq 1e-05$ para mapear as mesmas com os genes transcritos de *T. rubrum* CBS 118892 resultantes do projeto comparativo de fungos dermatófitos. Como resultado do mapeamento obteve-se 94 genes relacionados às modulações em meio de elastina e 89 em meio com queratina.

Em seguida, para identificação de regiões repetitivas nos genes modulados de *Trichophyton rubrum* utilizou-se a ferramenta TRF. Através dos parâmetros definidos na metodologia deste trabalho, a ferramenta permitiu a identificação de repetições perfeitas e imperfeitas. Para identificação de regiões repetitivas variáveis foi utilizado o algoritmo SERV também descrito na metodologia. Foram consideradas repetições adjacentes variáveis com comprimento de dez pares de base levando-se em consideração os parâmetros mínimos definidos nos experimentos realizados por Kolpakov; Bana e Kucherov (2003); Leclercq; Rivals e Jarne, (2007); Lim et al. (2012). A Tabela 12 apresenta as repetições adjacentes variáveis (*varscore* > 0) encontradas em ambos os meios e sua modulação resultante do experimento de *microarray*.

Tabela 12- Regiões repetitivas variáveis em genes modulados

ID Sequência	Anotação	Varscore	Unidade	Repetições	Fold	Repetição Consenso	Meio
TERG_01949	hypothetical protein TERG_01949	0,32	6	13,5	-5,61	ACGACA	Elastina
TERG_02489	sister chromatid cohesion acetyltransferase eco1	0,44	3	7,3	5,83	ACT	Elastina
TERG_03130	c2h2 transcription factor	0,38	3	8	5,27	AAC	Elastina
TERG_04799	ras guanyl-nucleotide exchange factor	0,58	3	9	6,25	GCA	Elastina
TERG_05624	cutinase transcription factor 1 alpha	0,56	3	12,7	6,22	CAG	Elastina
TERG_05624	cutinase transcription factor 1 alpha	0,63	3	18,7	6,22	GCA	Elastina
TERG_06235	hypothetical protein TERG_06235	0,24	6	7,3	-7,49	GGAAAT	Elastina
TERG_06235	hypothetical protein TERG_06235	0,75	6	9,8	-7,49	AACGGC	Elastina
TERG_08771	extracellular serine-threonine rich protein	0,92	12	10,8	5,64	CGGAGGAGGCAT	Queratina
TERG_08771	extracellular serine-threonine rich protein	4,99	45	23,4	5,64	GGACAGCCAGGACAGCCAGGACAG CCAGGACAGCCAGGTCAGCCA	Queratina

A presença de queratina no meio de cultura induziu a expressão do gene TERG_08771 (Número de acesso: 327302703) de *T. rubrum* que contém um extenso padrão de repetição adjacente. Esse gene codifica uma *serine threonine rich protein* com extensão de 3030-pb. O padrão de repetição adjacente identificado é um minissatélite localizado no locus 1382-2425 contendo uma região consenso de 45-pb que se repete por 23,4

vezes com extensão total de 1044-pb. O padrão repetitivo codifica uma extensa sequência de prolina, glicina e glutamina (Figura 23A). Foi observado também que este gene foi induzido *in vitro* quando *T. rubrum* foi co-cultivado na presença de linhagem de queratinócitos humanos.

Através dos dados obtidos no projeto comparativo de fungos dermatófitos, identificou-se genes homólogos listados na Tabela 13, que ainda apresenta as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens.

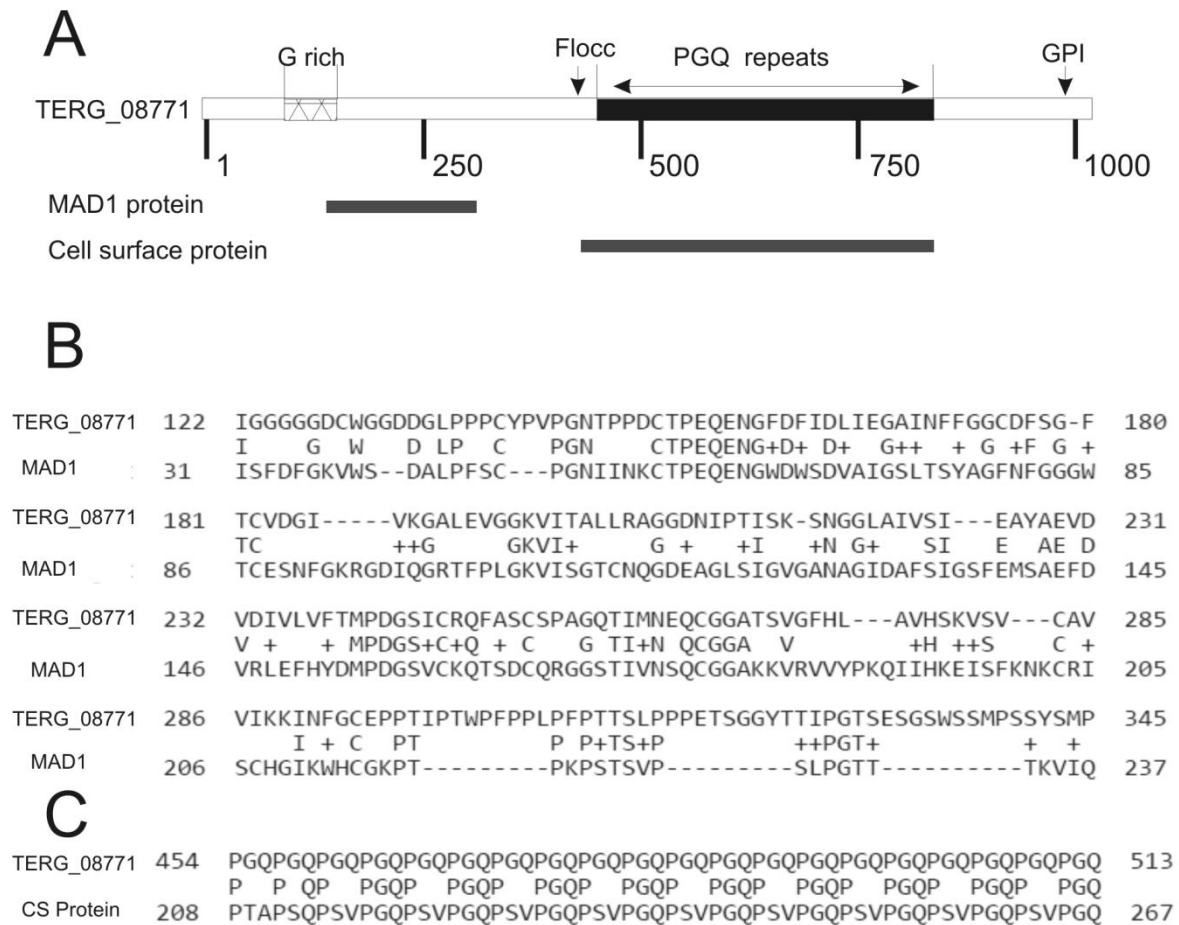
Tabela 13- Genes ortólogos de *T. rubrum*: TERG_08771

GENES	Repetições
TERG_08771: <i>T. rubrum</i> CBS 118892	23,4
ARB_05215: <i>A. benhamiae</i> CBS 112371	0
H100_05370: <i>T. rubrum</i> MR850	2,1
H100_05371: <i>T. rubrum</i> MR850	1,8
H101_05067: <i>T. interdigitale</i> H6	0
H102_05336: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 100081	1,8
H103_05361: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 288.86	3,6
H104_05352: <i>T. rubrum</i> (<i>morphotype kanei</i>) CBS 289.86	1,8
H105_05379: <i>T. rubrum</i> (<i>morphotype soudanense</i>) CBS 452.61	4,2
H106_05199: <i>T. rubrum</i> (<i>morphotype megninii</i>) CBS 735.88	2,8
H107_05492: <i>T. rubrum</i> (<i>morphotype raubitschekii</i>) CBS 202.88	4
H109_06729: <i>T. interdigitale</i> MR816	3,5
H110_05357: <i>T. rubrum</i> MR1448	4
H110_05358: <i>T. rubrum</i> MR1448	4,4
H112_05351: <i>T. rubrum</i> D6	7
H112_05352: <i>T. rubrum</i> D6	2,2
H113_05397: <i>T. rubrum</i> MR1459	2,2
MCYG_04810: <i>M. canis</i> CBS 113480	0
MGYG_08038: <i>M. gypseum</i> CBS 118893	0
TEQG_02877: <i>T. equinum</i> CBS 127.97	8,8
TESG_02648: <i>T. tonsurans</i> CBS 112818	5
TRV_02360: <i>T. verrucosum</i> HKI 0517	4,5

VITAGLIANO, 2012), sendo estas adesinas também requeridas para os processos iniciais da infecção em dermatófitos (BALDO et al., 2012). Enquanto que o domínio de *mucin-like glycoprotein* apresenta funções na aderência e patogenicidade de microrganismos, inclusive na formação de biofilme (KARUNANITHI et al., 2010). O grupo de proteínas com domínio *flocculin* é relacionado à parede celular (CWP). A presença de repetições adjacentes variáveis são responsáveis por mediar a interação patógeno hospedeiro através da adesão célula-célula e estar relacionado com a formação do biofilme como observado em *Aspergillus fumigatus* e *Saccharomyces cerevisiae* (VERSTREPEN et al. 2005; LEVDANSKY; SHARON; OSHEROV, 2008).

Além disso, como pode ser observado na Figura 25-B a proteína traduzida do gene TERG_08771 apresenta regiões homologas com a adesina que possui regiões repetitivas *Mad1* de *Metarhizium anisopliae* (WANG; LEGER, 2007). Outro ponto, é que a região repetitiva do gene TERG_08771 apresenta homologia com a sequência repetitiva de uma proteína de superfície celular de *Aspergillus fumigatus* (*cspA*) (Figura 25C), que de acordo com Levdansky et al. (2010) também possui um número variável de repetições adjacentes entre isolados. Segundo os autores, a supressão do gene *cspA* resultou em um fenótipo caracterizado pela rápida germinação de conídios e pela redução drástica da adesão à matriz extracelular, o que sugere que este gene esteja envolvido na definição de propriedades de superfície celular. Ainda, utilizando-se a ferramenta FaaPred que realiza a predição de adesinas em fungos, o gene TERG_08771 foi caracterizado como uma adesina, apresentando um alto *score* de confiança (0,997) (RAMANA; GUPTA, 2010).

Figura 25- Proteína resultante do gene TERG_08771. (A) Alinhamento com outras adesinas descritas. (B) Região homóloga com *Mad1* de *Metarhizium anisopliae*. (C) Região repetitiva homóloga com *cspA* de *Aspergillus fumigatus*



5.9. ADESÃO CELULAR E REPETIÇÕES ADJACENTES VARIÁVEIS EM *T. rubrum*

Adesinas são proteínas de superfície celular que conferem capacidade de ligação a células, tecidos e/ou superfícies abióticas. Estas proteínas são ferramentas do patógeno para invasão da célula hospedeira e são determinantes para sua virulência (RAMANA; GUPTA, 2010). Segundo os trabalhos de Verstrepen et al. (2005) e Levdansky et al. (2010) genes relacionados à adesão celular são ricos em repetições adjacentes variáveis. Assim utilizou-se a ferramenta FaaPred de predição de adesinas em fungos para identificar e quantificar genes que possuem características de adesinas.

Do total de 439 genes transcritos de *T. rubrum* CBS (isoformas foram desconsideradas) com repetições variáveis, 56% foram preditos com características estruturais de adesinas (*adhesin-like protein*) utilizando o parâmetro padrão de corte (-0,8). A partir de um critério estridente de corte utilizado pelos autores na validação da ferramenta (corte

$\geq 0,5$) 43 genes foram classificadas como adesinas (*adhesin protein*), correspondendo assim a ~10% dos genes com repetições variáveis (Figura 27).

A listagem completa dos genes contendo repetições variáveis, tentativa de anotação e pontuação da predição estão apresentadas no Apêndice B.

5.10. REPETIÇÕES ADJACENTES EM SISTEMAS DE TRANSDUÇÃO DE SINAL E REGULAÇÃO DE TRANSCRIÇÃO EM *T. rubrum*

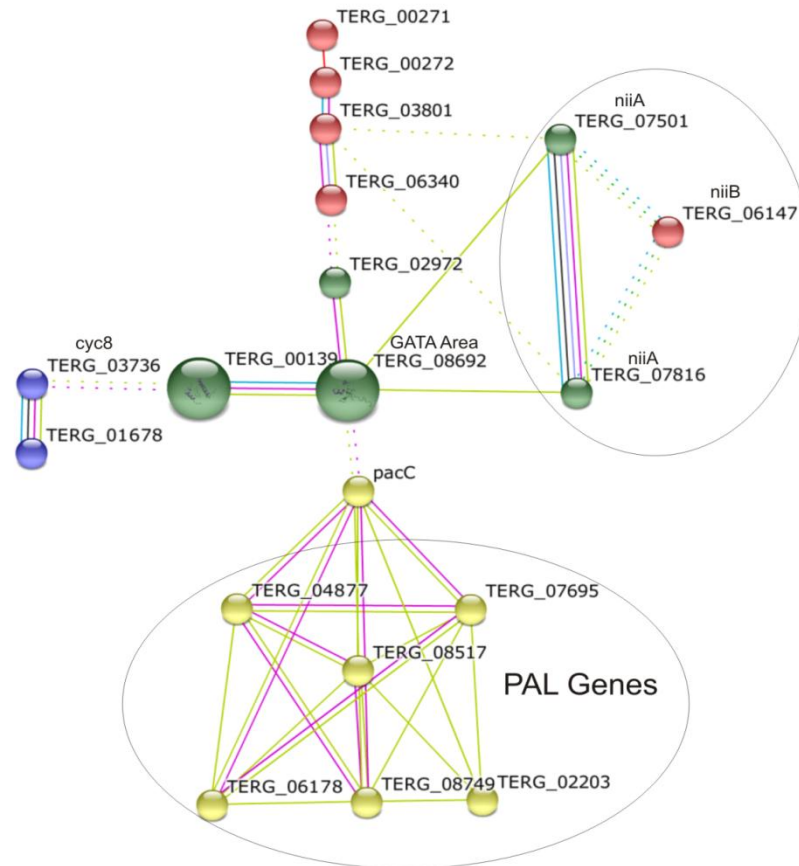
Estudos demonstram que a variação de repetições em *tandem* curtas (microsatélites) em regiões reguladoras pode impactar fortemente a expressão e / ou função do gene através de diferentes mecanismos. A alta prevalência dessas repetições em diferentes genomas sugere que este tipo de regulação da expressão gênica pode ser um mecanismo regulador comum e conservado (GEMAYEL et al., 2010).

Segundo Sawaya et al. (2013), repetições de microsatélites estão presentes em cerca de 3% do genoma humano, que embora representem uma pequena fração do genoma, quando encontradas em regiões funcionais, sua instabilidade pode modular fenótipos e regular a expressão gênica. Em leveduras, Vines et al. (2009) relatam que 25% de regiões reguladoras de *Saccharomyces cerevisiae* possuem repetições de microsatélites. De acordo com os autores a expansão e contração desses microsatélites alteram a estrutura da cromatina, resultando em alterações nos níveis de expressão do gene.

No dermatófito *T. rubrum*, ~10% dos genes analisados possuem repetições de microsatélites, no qual ~5% são variáveis. O *motif* CAG/CAA que codifica uma sequência de poliglutamina é a repetição de microsatélite mais abundante de *T. rubrum* sendo encontrado em um total de 110 genes. Este padrão é comumente encontrado em genes de transdução de sinal e regulação de transcrição de *T. rubrum*. Ainda, aproximadamente 21% dos genes com repetições variáveis de *T. rubrum* estão categorizados como relacionados à transdução de sinal e regulação de transcrição (Figura 27).

Em *T. rubrum*, uma via de interação proteica predita pela ferramenta STRING (SZKLARCZYK et al., 2015) é composta por genes que apresentam repetições de poliQ com um *score* de variabilidade acima de zero. Acredita-se que estas repetições sejam importantes, pois de acordo com Schaefer; Wanker e Andrade-Navarro (2012) as mesmas são responsáveis pela ligação proteína-proteína em uma via de interação.

Figura 26 - Via de interação de proteínas rica em poli-Q



Fonte: Adaptada pelo autor, gerada pela ferramenta STRING (SZKLARCZYK et al., 2015)

Genes de transdução de sinal (TERG_03801, TERG_06340) recebem estímulo externo podendo atuar na regulação da expressão dos fatores de transcrição *pacC* (TERG_00838) e *cyc8* (TERG_03736). Esta via é rica em repetições de poliQ, que de acordo com Atanesyan et al. (2012) está associada ao controle da expressão. Experimentos realizados pelos autores relatam que fatores de transcrição são mais ativos quando possuem repetições de poliQ, o que sugere um papel benéfico, pois além da contribuição para a atividade transcricional, sua instabilidade pode ajudar a espécie a se adaptar as mudanças ambientais de uma forma potencialmente reversível.

Ressalta-se que o gene TERG_08692 é um ativador de transcrição (*GATA transcriptional activator AreA*), envolvido na regulação de nitrogênio, o qual se mostra como elo entre os genes de transdução de sinal e os genes de transcrição *pacC* e *cyc8*. Acredita-se que a ativação da expressão gênica de *pacC* é dependente da interação com uma proteína reguladora de nitrogênio, de forma que o gene regulador de nitrogênio pode ser considerado como um potencial alvo antifúngico. O *pacC* em uma situação de infecção modula a

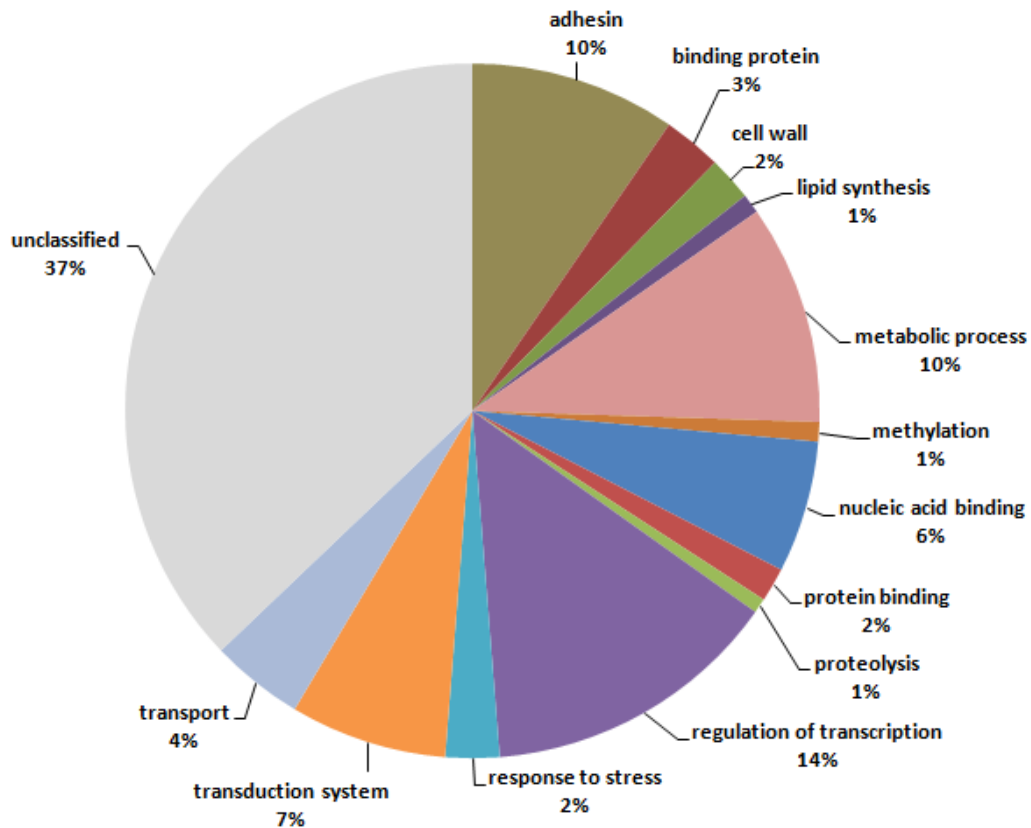
expressão de proteases de acordo com o pH do ambiente para obtenção de carbono e nitrogênio oriundos das proteínas degradadas, o mutante nulo gerado para esse gene apresentou menor capacidade de degradar queratina em fragmentos de unhas (FERREIRA-NOZAWA et al., 2006; MARTINEZ-ROSSI et al., 2012). De acordo com Limjindaporn; Khalaf e Fonzi (2003) reguladores de transcrição da família GATA são responsáveis por garantir a utilização eficiente de fontes de nitrogênio disponíveis para fungos. Já Tudzynski (2014) descreve estudos realizados em *Saccharomyces cerevisiae*, *Aspergillus nidulans*, e *Neurospora crassa*, os quais indicam que o ativador de transcrição *GATA transcription factor AreA* é um componente central em uma rede de regulação de nitrogênio, pois ele é importante para adaptação fúngica em ambientes em que a concentração de nitrogênio disponível é baixa.

A disrupção do gene *AcareA*, um homólogo do regulador de nitrogênio fúngico GATA de *Acremonium chrysogenum* revelou que o mesmo é necessário para o metabolismo de fontes alternativas de nitrogênio. O mutante nulo apresentou redução do crescimento em meio utilizando nitrato, ácido úrico e uma baixa concentração de amônio, glutamina ou ureia como única fonte de nitrogênio. Por outro lado, a análise transcricional revelou que a expressão dos genes de metabolismo de nitrato *niaD/niiA* foram fortemente induzidos em linhagens selvagens de *AcareA*, mas não na linhagem mutante (LI; PAN; LIU, 2013). Assim, o gene TERG_08692 mostra-se como um potencial alvo molecular, por mediar à interação proteica em uma importante via e, possivelmente, estar envolvido na regulação de nitrogênio.

5.11. COMPARAÇÃO DOS GENES DE *Trichophyton rubrum* CONTENDO REPETIÇÕES ADJACENTES VARIÁVEIS COM OUTROS DERMATÓFITOS

No item 5.8 descreveu-se a função e variabilidade do gene TERG_08771 modulado no experimento de *microarray*, o qual apresenta o maior *score* de variabilidade entre todos os genes de *T. rubrum*. Foi identificado um total de 522 genes transcritos incluindo isoformas (439 genes) com as regiões repetitivas variáveis em *T. rubrum* CBS os quais estão relacionados a diferentes processos metabólicos como transdutores de sinais, superfície celular e fatores de transcrição (Figura 27). Para categorização funcional dos genes contendo repetições utilizou-se o *Gene Ontology*, Funcat e predição de adesinas oriundas da ferramenta FaaPred (adesinas foram consideradas com corte acima de 0,5). A curadoria das categorias foi realizada pela doutoranda Tamires Aparecida Bitencourt.

Figura 27- Categorias funcionais de genes com repetições adjacentes variáveis



Alguns destes genes foram selecionados para comparar a variabilidade e provável função de suas repetições dentre as linhagens de *T. rubrum* e demais espécies/linhagens de dermatófitos.

5.12. GENE TERG_05576

Proteínas de parede celular são essenciais para quase todos os aspectos da biologia e patogenicidade de fungos, tendo relação com a adesão aos tecidos do hospedeiro, consistindo como uma barreira de permeabilidade que mantém a estrutura característica de fungos (CHAFFIN et al., 1998). O gene TERG_05576 anotado como *cell wall glucanase*, categorizado pela tabela Funcat como relacionado à parede celular e possuindo os termos do *Gene Ontology*: *C:extracellular region| C:cell wall| F:beta-glucosidase activity| P:cellulose catabolic process*, apresenta uma repetição variável (*varscore* 0,69) de minissatélite de 18-pb a unidade, se repetindo por 5,4 vezes no locus 638-731 do gene transcrito. Este gene foi modulado no trabalho Bitencourt et al. (resultados não publicados) no qual os autores realizaram um experimento de *microarray* em meio de elastina adicionado de transchalcona com objetivo de observar o perfil transcricional de *T. rubrum* em exposição a esta droga.

A repetição consenso identificada foi 'AGGCCCTGAACCACCAA'. Através dos dados obtidos no projeto comparativo de fungos dermatófitos identificou-se genes homólogos listados na Tabela 14, que ainda apresenta as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens.

Tabela 14- Genes ortólogos de *T. rubrum*: TERG_05576

GENE	Repetições
TERG_05576: <i>T. rubrum</i> CBS 118892	5,4
ARB_05864: <i>A. benhamiae</i> CBS 112371	4,4
H100_02957: <i>T. rubrum</i> MR850	5,4
H101_03613: <i>T. interdigitale</i> H6	0
H102_02950: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 100081	5,4
H103_02964: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 288.86	5,4
H104_02943: <i>T. rubrum</i> (<i>morphotype kanei</i>) CBS 289.86	5,4
H105_02969: <i>T. rubrum</i> (<i>morphotype soudanense</i>) CBS 452.61	5,4
H106_02792: <i>T. rubrum</i> (<i>morphotype megninii</i>) CBS 735.88	5,4
H107_03049: <i>T. rubrum</i> (<i>morphotype raubitschekii</i>) CBS 202.88	5,4
H109_07270: <i>T. interdigitale</i> MR816	0
H110_02965: <i>T. rubrum</i> MR1448	5,4
H112_02952: <i>T. rubrum</i> D6	5,4
H113_02963: <i>T. rubrum</i> MR1459	5,4
MCYG_02607: <i>M. canis</i> CBS 113480	0
MGYG_08325: <i>M. gypseum</i> CBS 118893	0
TEQG_08521: <i>T. equinum</i> CBS 127.97	0
TESG_01640: <i>T. tonsurans</i> CBS 112818	0
TRV_03664: <i>T. verrucosum</i> HKI 0517	0

Como se observa na listagem apresentada na Tabela 14, o gene TERG_05576 apresenta homólogos em todas as linhagens de *T. rubrum* e demais espécies de dermatófitos sequenciados pelo projeto comparativo. No que tange a repetição variável identificada, a mesma se mostra conservada entre todas as linhagens de *T. rubrum*, mantendo a repetição consenso, o mesmo tamanho de unidade e quantidade de repetições.

A repetição ainda aparece em *A. benhamiae* com a exclusão de uma unidade, pois a sequência consenso ocorre por 4,4 vezes. Nos demais organismos (*T. interdigitale* linhagem

H6, *T. interdigitale* linhagem MR816, *T. verrucosum*, *T. equinum*, *T. tonsurans*, *M. canis*, *M. gypseum*) a sequência repetitiva não ocorre.

Como muitas atividades enzimáticas estão associadas com a parede celular, sendo que as enzimas participam na degradação dos componentes extracelulares, identificou-se que a sequência repetitiva encontrada em todas as linhagens de *T. rubrum* se caracteriza pelos domínios *proline-rich* e *chaperone protein DnaJ*, sendo este último envolvido na degradação proteolítica em *Escherichia coli* (HUANG et al., 2001).

5.13. GENE TERG_00768

Proteínas de serina/treonina quinases são componentes cruciais de diversas vias de sinalização em eucariotos, incluindo fungos filamentosos. Essas proteínas são responsáveis pela regulação de diferentes aspectos do crescimento e desenvolvimento do organismo como a divisão celular, metabolismo e apoptose (GONZÁLEZ; BRITO; GONZÁLEZ, 2012).

O gene TERG_00768 anotado como *serine/threonine protein kinase*, categorizado pela tabela Funcat como relacionado à regulação do metabolismo, ciclo celular e destino celular, possuindo diversos termos do *Gene Ontology* relacionados à transferência de fosfato (*phosphorelay*), além de outros termos como *ATP binding*, *regulation of transcription*, *DNA-dependent*, *intracellular signal transduction*, apresenta uma repetição variável (*varscore* 0,33) de minissatélite de 12-pb a unidade, se repetindo por 6,8 vezes no locus 7043-7124 do gene transcrito. A repetição consenso identificada é 'CTGAGGAAAACA'.

A partir dos dados obtidos no projeto comparativo de fungos dermatófitos, identificou-se genes homólogos listados na Tabela 15, que ainda apresenta as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens.

Tabela 15- Genes ortólogos de *T. rubrum*: TERG_00768

GENE	Repetições
TERG_00768: <i>T. rubrum</i> CBS 118892	6,8
ARB_07756: <i>A. benhamiae</i> CBS 112371	5,8
H100_08222: <i>T. rubrum</i> MR850	6,8
H101_06505: <i>T. interdigitale</i> H6	3,8
H102_08179: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 100081	6,8
H103_08204: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 288.86	6,8
H104_08155: <i>T. rubrum</i> (<i>morphotype kanei</i>) CBS 289.86	6,8
H105_08207: <i>T. rubrum</i> (<i>morphotype soudanense</i>) CBS 452.61	6,8
H106_08076: <i>T. rubrum</i> (<i>morphotype megninii</i>) CBS 735.88	4,8
H107_08348: <i>T. rubrum</i> (<i>morphotype raubitschekii</i>) CBS 202.88	6,8
H109_05804: <i>T. interdigitale</i> MR816	3,8
H110_08201: <i>T. rubrum</i> MR1448	6,8
H112_01647: <i>T. rubrum</i> D6	0
H112_01648: <i>T. rubrum</i> D6	0
H113_08269: <i>T. rubrum</i> MR1459	6,8
MCYG_00336: <i>M. canis</i> CBS 113480	0
MGYG_00907: <i>M. gypseum</i> CBS 118893	0
TEQG_01117: <i>T. equinum</i> CBS 127.97	3,8
TESG_02286: <i>T. tonsurans</i> CBS 112818	0
TRV_02917: <i>T. verrucosum</i> HKI 0517	0

O gene TERG_00768 apresenta homólogos em todas as linhagens de *T. rubrum* e demais espécies de dermatófitos sequenciados pelo projeto comparativo. A repetição variável identificada se mantém conservada em oito das dez outras linhagens de *T. rubrum* (MR1448, *morphotype fischeri* CBS 288.86, *morphotype raubitschekii* CBS 202.88, *morphotype fischeri* CBS 100081, *morphotype kanei* CBS 289.86, MR1459, MR850, *morphotype soudanense* CBS 452.61) com a mesma unidade, quantidade de repetições e locos. A linhagem *T. rubrum* (*morphotype megninii*) CBS 735.88 apresenta a repetição, porém com variabilidade, se repetindo por 4,8 vezes no mesmo locos. A repetição ainda ocorre em *A. benhamiae* (5,8 vezes) e nas espécies *T. interdigitale* H6, *T. equinum* CBS 127.97, *T. interdigitale* MR816, com alterações na quantidade de repetições, sendo que a repetição ocorre por 3,8 vezes. A

linhagem de *T. rubrum* D6, e demais espécies (*T. verrucosum*, *T. tonsurans*, *M. canis* e *M. gypseum*) não apresentam a repetição.

5.14. GENE TERG_07370

Algumas glicoproteínas são sensores para alterações no ambiente extracelular, tendo propriedades de adesão e atividades hidrolíticas. Assim, uma atenção especial tem sido dada às glicoproteínas de parede celular em fungos patogênicos, pois alguns deles são considerados fatores de virulência (MORA-MONTES, 2012).

O gene TERG_07370 anotado como *major outer envelope glycoprotein-like protein*, possui o termo de processo metabólico do *Gene Ontology* relacionado à reparação/replicação de DNA e síntese de proteínas. O gene (EST DW709116) foi induzido no trabalho de Liu et al. (2014), no qual os autores realizaram um experimento de *microarray* observando que a exposição à pele humana resultou na modulação dos níveis de expressão deste gene em *T. rubrum*. O gene apresenta uma repetição variável (*varscore* 1,11) de 18-pb de unidade se repetindo por 33 vezes no locus 26-615 do gene transcrito. A repetição consenso identificada é 'CTCCTCCTATCACTCAAG'.

Através dos dados obtidos no projeto comparativo de fungos dermatófitos identificou-se genes homólogos listados na Tabela 16, que ainda apresentam as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens.

Tabela 16- Genes ortólogos de *T. rubrum*: TERG_07370

GENE	Repetições
TERG_07370: <i>T. rubrum</i> CBS 118892	33
ARB_00001: <i>A. benhamiae</i> CBS 112371	0
H100_00946: <i>T. rubrum</i> MR850	31
H101_06856: <i>T. interdigitale</i> H6	0
H102_00937: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 100081	16,9
H103_00945: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 288.86	0
H104_00929: <i>T. rubrum</i> (<i>morphotype kanei</i>) CBS 289.86	19,9
H105_00944: <i>T. rubrum</i> (<i>morphotype soudanense</i>) CBS 452.61	33
H106_00742: <i>T. rubrum</i> (<i>morphotype megninii</i>) CBS 735.88	15
H107_00995: <i>T. rubrum</i> (<i>morphotype raubitschekii</i>) CBS 202.88	31
H109_00592: <i>T. interdigitale</i> MR816	0
H110_00946: <i>T. rubrum</i> MR1448	20,9
H112_00947: <i>T. rubrum</i> D6	0
H113_00948: <i>T. rubrum</i> MR1459	19,9
MCYG_00456: <i>M. canis</i> CBS 113480	0
TEQG_03681: <i>T. equinum</i> CBS 127.97	0
TESG_04260: <i>T. tonsurans</i> CBS 112818	0
TRV_07964: <i>T. verrucosum</i> HKI 0517	0

A repetição variável identificada se manteve conservada em número de repetições apenas na linhagem *T. rubrum* (*morphotype soudanense*) CBS 452.61. Em outras sete linhagens a repetição ocorre com variabilidade no número de *repeats*. Em *T. rubrum* MR850 e *T. rubrum* (*morphotype raubitschekii*) CBS 202.88 a repetição ocorre por 31 vezes, em *T. rubrum* MR1448 a repetição ocorre por 20,9 vezes, em *T. rubrum* MR1459 e *T. rubrum* (*morphotype kanei*) CBS 289.86 a repetição ocorre por 19,9 vezes, em *T. rubrum* (*morphotype fischeri*) CBS 100081 por 16,9 vezes e em *T. rubrum* (*morphotype megninii*) CBS 735.88 por 15 vezes.

A repetição não ocorre nas outras duas linhagens de *T. rubrum* (D6 e *morphotype fischeri* CBS 288.86) nem nas demais espécies de dermatófitos que apresentam genes homólogos: *T. interdigitale* (H6 e MR816), *T. equinum*, *T. tonsurans*, *T. verrucosum*, *A. benhamiae* e *M. canis*.

5.15. GENE TERG_03736

Cyc8 é um regulador transcricional geral que controla a expressão de ~ 3% de todos os genes em *Saccharomyces cerevisiae* envolvidos em uma variedade de vias. Este gene é altamente conservado com homólogos em diversos organismos eucariotos, incluindo os seres humanos e o fungo patogênico *Candida albicans*. Existem evidências que as repetições adjacentes existentes neste gene de fato tem um papel na geração de variabilidade biológica funcional (BOEYNAEMS, 2012).

O gene TERG_03736 anotado como *transcriptional corepressor Cyc8*, possui o termo de *cellular response to drug* do *Gene Ontology*, relacionado a processos que resultam em uma mudança de estado ou atividade de uma célula como um resultado de um estímulo de drogas. Genes de regulação são enriquecidos de repetições de microssatélites, sendo que este gene apresenta duas repetições variáveis. Uma repetição de 6-pb que se repete por 54,7 vezes no locos 2885-3215 (*varscore* 1,77) e outra de 3-pb que se repete por 13,7 vezes no locos 1882-1922 (*varscore* 1,03) do referido gene transcrito. O referido gene (EST DW708254) foi induzido no trabalho de Zhao et al. (2009), no qual os autores realizam um experimento de *microarray* com objetivo de conhecer os perfis globais de transcrição de resposta ao fármaco a 5-flucitosina. Este gene também foi induzido no trabalho de Persinoti et al (2014), no qual os autores realizaram um experimento de *RNA-seq* comparando quatro bibliotecas de cDNA, sendo uma com o cultivo de *T. rubrum* em meio *sabouraud* e outras três em diferentes períodos de exposição à droga acriflavina: 3 horas, 12 horas e 24 horas.

A partir dos dados obtidos no projeto comparativo de fungos dermatófitos identificou-se genes homólogos listados na Tabela 17, que ainda apresenta as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens. Como o gene apresenta duas repetições variáveis, a Tabela 17, apresenta as quantidades das repetições de unidade 6-pb e 3-pb respectivamente.

Tabela 17- Genes ortólogos de *T. rubrum*: TERG_03736

GENE	Repetições
TERG_03736: <i>T. rubrum</i> CBS 118892	54,7 13,7
ARB_00627: <i>A. benhamiae</i> CBS 112371	0 9,7
H100_06045: <i>T. rubrum</i> MR850	54,7 13,7
H101_06231: <i>T. interdigitale</i> H6	45,7 15,7
H102_06014: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 100081	54,7 13,7
H103_06039: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 288.86	54,7 13,7
H104_06027: <i>T. rubrum</i> (<i>morphotype kanei</i>) CBS 289.86	54,7 13,7
H105_06053: <i>T. rubrum</i> (<i>morphotype soudanense</i>) CBS 452.61	41,7 13,7
H106_05876: <i>T. rubrum</i> (<i>morphotype megninii</i>) CBS 735.88	31,7 13,7
H107_06178: <i>T. rubrum</i> (<i>morphotype raubitschekii</i>) CBS 202.88	54,7 13,7
H109_07225: <i>T. interdigitale</i> MR816	48,7 15,7
H110_06035: <i>T. rubrum</i> MR1448	54,7 13,7
H112_06031: <i>T. rubrum</i> D6	54,7 13,7
H113_06082: <i>T. rubrum</i> MR1459	54,7 13,7
MCYG_04116: <i>M. canis</i> CBS 113480	0 0
MGYG_02496: <i>M. gypseum</i> CBS 118893	48,7 0
TEQG_07940: <i>T. equinum</i> CBS 127.97	44,7 14,7
TESG_07328: <i>T. tonsurans</i> CBS 112818	39,7 14,7
TRV_05603: <i>T. verrucosum</i> HKI 0517	28,7 9,7

A repetição de 6-pb que se repete por 54,7 vezes possui a sequência consenso 'GAGAGC' codificando uma extensa sequência do dipeptídeo arginina/ácido glutâmico (RE). Esta repetição se mantém conservada em número de repetições em oito outras linhagens de *T. rubrum* (MR850, *morphotype fischeri* CBS 100081, *morphotype fischeri* CBS 288.86, *morphotype kanei* CBS 289.86, *morphotype raubitschekii* CBS 202.88, MR1448, D6, MR1459) sequenciadas pelo projeto comparativo. Outras duas linhagens de *T. rubrum* - *morphotype megninii* CBS 735.88 e *morphotype soudanense* CBS 452.61- possuem a repetição com variabilidade, se repetindo por 31,7 vezes e 41,7 vezes respectivamente. A repetição também ocorre de forma variável nas espécies de *T. interdigitale* MR816 e H6, se repetindo por 48,7 e 45,7 vezes respectivamente. A repetição ainda ocorre em *A. benhamiae* (49,7 vezes), *M. gypseum* (48,7 vezes), *T. equinum* (44,7 vezes), *T. tonsurans* (39,7 vezes) e *T. verrucosum* (28,7 vezes). A repetição não ocorre em *M. canis*.

A outra repetição de 3-pb que se repete por 13,7 vezes no locos 1882-1922 possui a sequência consenso 'CAA' codificando uma cadeia de poliglutamina. Esta repetição se mantém perfeitamente conservada em número de repetições em todas outras linhagens de *T. rubrum*, já em *T. verrucosum* e *A. benhamiae* a mesma se repete por 9,7 vezes. Em *T. equinum* e *T. tonsurans* a repetição ocorre por 14,7 vezes, e as duas linhagens de *T. interdigitale* (H6 e MR816) apresentam a maior extensão da repetição, ocorrendo por 15,7 vezes. A repetição não ocorre em *M. canis* e *M. gypseum*.

Boeynaems (2012) realizou um experimento de *RNA-seq* para analisar genes diferencialmente expressos em linhagens mutantes de *Saccharomyces cerevisiae* com suas repetições alteradas em número no gene *cyc8*. As variações nas repetições existentes no gene se mostraram importantes, pois de acordo com o trabalho três principais classes de genes foram diferencialmente expressos nos mutantes *cyc8*: (1) os genes da parede celular, (2) reguladores de síntese da parede celular e (3) genes de estresse responsivo.

5.16. GENE TERG_05189

A expressão dos genes eucarióticos requer a coordenação de um complexo conjunto de proteínas de ligação de RNA mensageiro. Certo número de proteínas que contêm *motifs* ricos em arginina (*arginine-rich motifs* - ARMs) são conhecidas por se ligarem ao RNA e estarem envolvidas na regulação do seu processamento (MCBRIDE et al., 2009).

O gene TERG_05189 anotado como "*arginine rich protein*" contém uma extensa sequência repetitiva de arginina/ácido aspártico com resíduos de ácido glutâmico possuindo o domínio *COMPASS*, estando relacionado à regulação da expressão. Genes de regulação são enriquecidos de repetições de microssatélites, sendo que este gene apresenta uma repetição variável de unidade curta (*varscore* 1,43). A repetição possui 6-pb de unidade e se repete por 43,5 vezes no locos 1834-2094 do referido gene transcrito apresentando a sequência consenso 'GACAGA'. Este gene (ESTs DW697490 e DW686228) faz parte da biblioteca de cDNA criada por Wang et al. (2006).

Através dos dados obtidos no projeto comparativo de fungos dermatófitos identificou-se genes homólogos listados na Tabela 18, que ainda apresenta as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens.

Tabela 18- Genes ortólogos de *T. rubrum*: TERG_05189

GENE	Repetições
TERG_05189: <i>T. rubrum</i> CBS 118892	43,5
ARB_03132: <i>A. benhamiae</i> CBS 112371	0
H100_03865: <i>T. rubrum</i> MR850	43,5
H101_07120: <i>T. interdigitale</i> H6	38,5
H102_03851: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 100081	43,5
H103_03866: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 288.86	43,5
H104_03851: <i>T. rubrum</i> (<i>morphotype kanei</i>) CBS 289.86	43,5
H105_03879: <i>T. rubrum</i> (<i>morphotype soudanense</i>) CBS 452.61	43,5
H106_03675: <i>T. rubrum</i> (<i>morphotype megninii</i>) CBS 735.88	49,5
H107_03976: <i>T. rubrum</i> (<i>morphotype raubitschekii</i>) CBS 202.88	43,5
H109_02766: <i>T. interdigitale</i> MR816	43,5
H110_03857: <i>T. rubrum</i> MR1448	43,5
H112_03856: <i>T. rubrum</i> D6	43,5
H113_03889: <i>T. rubrum</i> MR1459	43,5
MCYG_03748: <i>M. canis</i> CBS 113480	12,5
MGYG_03350: <i>M. gypseum</i> CBS 118893	23,5
TEQG_03966: <i>T. equinum</i> CBS 127.97	42,5
TESG_05066: <i>T. tonsurans</i> CBS 112818	42,5
TRV_07051: <i>T. verrucosum</i> HKI 0517	0

A repetição variável se mantém conservada em número de repetições em 9 das 10 outras linhagens de *T. rubrum* (MR850, *morphotype fischeri* CBS 100081, *morphotype fischeri* CBS 288.86, *morphotype kanei* CBS 289.86, *morphotype soudanense* CBS 452.61, *morphotype raubitschekii* CBS 202.88, MR1448, D6, MR1459) e na espécie *T. interdigitale* MR816, sendo que em outra linhagem de *T. interdigitale* (H6) a repetição ocorre por 38,5 vezes. A maior quantidade de repetições ocorre na linhagem de *T. rubrum* (*morphotype megninii*) CBS 735.88 (49,5 vezes). Em *T. equinum* e *T. tonsurans* a repetição ocorre 42,5 vezes em ambas. A repetição ainda ocorre em *M. gypseum* (23,5 vezes) e *M. canis* (12,5 vezes.). A repetição não ocorre em *A. benhamiae* e *T. verrucosum*.

5.17. GENE TERG_08141

Genes de regulação são de grande importância para a fisiologia de todos os organismos e tem sido objeto de estudos em fungos. Neste contexto, os fatores de transcrição são elos fundamentais nas redes de regulação em organismos eucarióticos, incluindo fungos filamentosos (TIAN; LI; GLASS, 2001). O fator de transcrição bZIP (*basic leucine zipper*) é uma família de genes multifuncional em várias espécies, incluindo fungos patogênicos, no qual desempenham um papel na regulação da patogenicidade (YE et al., 2013). Através da tentativa de anotação realizada neste trabalho, identificou-se 14 genes pertencentes à família bZIP em *T. rubrum* CBS 118892 (TERG_00359, TERG_01028, TERG_02140, TERG_02940, TERG_03901, TERG_05396, TERG_05479, TERG_05593, TERG_05996, TERG_06815, TERG_07733, TERG_08022, TERG_08141, TERG_08825), mesma quantidade apresentada em *S. cerevisiae* de acordo com Tian; Li e Glass (2001).

O gene TERG_08141, anotado como *bzip transcription factor*, associado a termos do *Gene Ontology* relacionados à regulação da transcrição, possui uma repetição variável de microssatélite (*varscore* 0,99) de 3-pb de unidade que se repete por 27,3 vezes no locus 661-742 da região codificante do referido gene. A sequência consenso 'CAA' codifica uma sequência repetitiva de glutamina com resíduos de histidina. Este gene foi induzido no trabalho de Persinoti et al (2014) citado anteriormente, em condição de exposição a droga acriflavina no tempo de 24 horas.

Através dos dados obtidos no projeto comparativo de fungos dermatófitos identificou-se genes homólogos listados na Tabela 19, que ainda apresenta as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens.

Tabela 19- Genes ortólogos de *T. rubrum*: TERG_08141

GENE	Repetições
TERG_08141: <i>T. rubrum</i> CBS 118892	27,3
ARB_02172: <i>A. benhamiae</i> CBS 112371	18,3
H100_00527: <i>T. rubrum</i> MR850	27,3
H101_07555: <i>T. interdigitale</i> H6	27,3
H102_00527: <i>T. rubrum</i> (morphotype <i>fischeri</i>) CBS 100081	27,3
H103_00527: <i>T. rubrum</i> (morphotype <i>fischeri</i>) CBS 288.86	27,3
H104_00517: <i>T. rubrum</i> (morphotype <i>kanei</i>) CBS 289.86	27,3
H105_00515: <i>T. rubrum</i> (morphotype <i>soudanense</i>) CBS 452.61	27,3
H106_00410: <i>T. rubrum</i> (morphotype <i>megninii</i>) CBS 735.88	27,3
H107_00573: <i>T. rubrum</i> (morphotype <i>raubitschekii</i>) CBS 202.88	27,3
H109_00308: <i>T. interdigitale</i> MR816	27,3
H110_00531: <i>T. rubrum</i> MR1448	27,3
H112_00528: <i>T. rubrum</i> D6	27,3
H113_00532: <i>T. rubrum</i> MR1459	27,3
MCYG_00012: <i>M. canis</i> CBS 113480	13,3
MGYG_07213: <i>M. gypseum</i> CBS 118893	14,7
TEQG_08786: <i>T. equinum</i> CBS 127.97	0
TESG_07933: <i>T. tonsurans</i> CBS 112818	0
TESG_07997: <i>T. tonsurans</i> CBS 112818	0
TRV_00423: <i>T. verrucosum</i> HKI 0517	18,3

A repetição variável se mantém conservada em número de repetições em todas outras 10 linhagens de *T. rubrum* e nas duas linhagens da espécie *T. interdigitale* (MR816 e H6). Em *A. benhamiae* e *T. verrucosum* a repetição ocorre por 18,3 vezes. A repetição ainda ocorre em *M. gypseum* (14,7 vezes) e *M. canis* (13,3 vezes). A repetição não ocorre nas espécies *T. equinum* e *T. tonsurans*.

5.18. GENE TERG_01042

O fator de transcrição *Fungal-specific TF domain* é o mais abundante entre 77 genomas de fungos *ascomycete* de acordo com Todd et al. (2014). O gene TERG_01042 anotado como *C6 transcription factor, putative*, possui o domínio *Fungal-specific TF domain* e está associado a termos do *Gene Ontology* como *zinc ion binding* e *regulation of*

transcription from RNA polymerase II promoter, este gene ainda possui uma região de 460-pb homóloga ao gene *transcriptional regulatory protein GAL4*. Em *Candida Albicans* genes desta família estão relacionados com proteínas fúngicas reguladores da transcrição de genes envolvidos em uma ampla variedade de processos celulares, incluindo metabolismo de diferentes compostos, tais como açúcares, controle da resistência multi droga e arquitetura da parede celular (MAICAS et al., 2005).

Este gene apresenta uma sequência repetitiva variável (*varscore* 1,11) de microssatélite de 3-pb que se repete por 27 vezes no locus 1423-1502 caracterizada pelo *motif* "CAA/CAG". O padrão repetitivo codifica uma extensa sequência de poliglutamina (poliQ).

A sequência repetitiva se mantém conservada em nove das outras dez linhagens de *T. rubrum*, sendo que apenas na linhagem *T. rubrum (morphotype megninii)* CBS 735.88 a mesma não ocorre. A repetição ainda ocorre nas duas linhagens de *T. interdigitale* (17 vezes), em *T. equinum* e *T. tonsurans* (15 vezes), *A. benhamiae* (12 vezes) e *T. verrucosum* (20 vezes). A repetição não ocorre em *M. canis* e *M. gypseum*.

Os genes homólogos identificados estão listados na Tabela 20, que ainda apresenta as quantidades de ocorrências das unidades de repetições conservadas dentre as espécies e linhagens.

Tabela 20- Genes ortólogos de *T. rubrum*: TERG_01042

GENE	Repetições
TERG_01042: <i>T. rubrum</i> CBS 118892	28
ARB_03466: <i>A. benhamiae</i> CBS 112371	12
H100_08501: <i>T. rubrum</i> MR850	28
H101_03088: <i>T. interdigitale</i> H6	17
H102_08460: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 100081	28
H103_08483: <i>T. rubrum</i> (<i>morphotype fischeri</i>) CBS 288.86	28
H104_08436: <i>T. rubrum</i> (<i>morphotype kanei</i>) CBS 289.86	28
H105_08489: <i>T. rubrum</i> (<i>morphotype soudanense</i>) CBS 452.61	28
H106_08360: <i>T. rubrum</i> (<i>morphotype megninii</i>) CBS 735.88	0
H107_08638: <i>T. rubrum</i> (<i>morphotype raubitschekii</i>) CBS 202.88	28
H109_05990: <i>T. interdigitale</i> MR816	17
H110_08486: <i>T. rubrum</i> MR1448	28
H112_08479: <i>T. rubrum</i> D6	28
H113_08554: <i>T. rubrum</i> MR1459	28
MCYG_03330: <i>M. canis</i> CBS 113480	0
MGYG_00493: <i>M. gypseum</i> CBS 118893	0
TEQG_00844: <i>T. equinum</i> CBS 127.97	15
TESG_08566: <i>T. tonsurans</i> CBS 112818	15
TRV_06120: <i>T. verrucosum</i> HKI 0517	20

Repetições de poliglutamina em genes de regulação estão associados ao controle da expressão, pois de acordo com Atanesyan et al. (2012) modelos baseados no fator de transcrição *Gal-4* são mais ativos quando possuem repetições de poliQ, sendo que esta atividade depende do tamanho da repetição. Ainda segundo os autores, a ocorrência de repetições poliQ em proteínas reguladoras sugere um papel benéfico, pois além da contribuição para a atividade transcricional, sua instabilidade genética pode ajudar a espécie a se adaptar às mudanças nas condições ambientais de uma forma potencialmente reversível.

5.19. REPETIÇÕES ADJACENTES VARIÁVEIS EM REGIÕES *UPSTREAM* DE *Trichophyton rubrum*

Tendo em vista a evidência que repetições adjacentes variáveis modulam a expressão de certos genes (MARTIN et al., 2005), em especial repetições de microssatélites que podem alterar os níveis de expressão de genes em vários organismos, desde bactérias a humanos (SAWAYA et al., 2013), neste trabalho também se buscou identificar a ocorrência destas repetições em regiões *upstream* dos genes por serem responsáveis por conter as regiões promotoras e reguladoras da expressão gênica.

Foi submetido para identificação de repetições adjacentes um total de 8616 de sequências de regiões *upstream* de genes obtidos do projeto comparativo de dermatófitos de *Trichophyton rubrum* CBS 118892. O quadro representado na Figura 28 apresenta a quantidade de repetições encontradas e a quantidade de sequências que possuem estas repetições.

Figura 28- Repetições adjacentes em regiões upstream dos genes

Total de genes	Repetições identificadas	Sequências com repetições	Repetições variáveis	% de genes com repetições	% de genes com repetições variáveis
8616	7181	3228	1459	37,5%	16,9%

Como pode ser observado na Figura 28, das 8616 sequências de regiões *upstream*, 37,5% (3228) contêm ao menos uma repetição adjacente, sendo que 16,9% destas repetições foram classificadas como variáveis (*varscore* > 0), resultado que corrobora com o trabalho de Vines et al. (2009), no qual os autores realizaram a busca por repetições adjacentes em sequências *upstream* dos genes (região promotora) de *Saccharomyces cerevisiae*, revelando que estas regiões são enriquecidas em repetições, pois ao menos 25% delas possuem *tandem repeats*.

5.20. GENES COM REPETIÇÕES VARIÁVEIS EM OUTRAS ESPÉCIES DE FUNGOS DERMATÓFITOS

5.20.1. *Microsporum canis*

Assim como ocorreu em *T. rubrum* CBS 118892, o gene com a repetição adjacente com o maior *score* de variabilidade em *Microsporum canis* foi um homólogo do gene TERG_08771, também anotado como *extracellular serine-threonine rich protein*.

É importante salientar ainda que o gene *Adhesin-like protein (Mad1)* também possui repetições adjacentes variáveis em sua sequência, o que corrobora com Levdansky; Sharon e Oshero (2008) que genes de adesão e parede celular são ricos em *tandem repeats*. De acordo com Verstrep et al. (2005), a existência destas repetições em genes de adesão são responsáveis por uma adesão mais forte se comparado a genes sem repetições.

5.20.2. *T. tonsurans*

Em *Trichophyton tonsurans*, a repetição mais variável foi identificada no gene TESG_04591 pertencente à família de genes *PE repeat protein*. A repetição identificada possui 45-pb e se repete por 11 vezes (*varscore* 2,72) no locus 1924-2423 do referido gene. A repetição identificada codifica uma sequência rica em Alanina, Prolina e Ácido glutâmico como pode ser observado na Figura 30.

Figura 30- Sequência de aminoácidos proveniente da repetição adjacente do gene TESG_04591

```
PCNVKEEDDNASDIDGLKSSNDPVEVTTEEKEETSVAAPVPIEVTESVPLENATPETPAE
ESTVEQPTIEESVTEEPAAEFPVAEELAAEFPAAEFPAAQEPAAEELAAEFPAAQEPAAE
ELAAEFPAAQEPAAEELAAEESTAQEPAAEFPAAEFPVAEELAAEFPAAQEPAAEELAAE
EPAAQEPAAEFPAAEESTAQEPAAEFPAAEFPVAEELAAEFPAAQEPAAEFPAAEESTAQ
EPVEPKPIPEEPAAGDPVVDNPAEEQVEGPSSAEESKPEDPITEEPAANEPAREEQAAA
```

Genes da família *PE repeat protein* são de interesse para estudos, pois há indícios de que estas proteínas interagem com componentes do hospedeiro e, assim, subvertem vias de resposta imune e auxiliam o patógeno a se estabelecer dentro do hospedeiro (SAMPSON, 2011).

5.20.3. *T. equinum*

Em *Trichophyton equinum*, o gene TEQG_03078 apresenta uma repetição variável (*varscore* 2.26) de 42-pb que se repete por 7 vezes no locus 1182-1491 do gene transcrito. O gene codifica uma proteína hipotética a qual não apresentou homologia com genes já anotados. O gene apresenta o domínio OmpH-like (*Outer membrane protein*), o qual caracteriza a sequência repetitiva identificada. Este domínio foi caracterizado como uma chaperona molecular que interage com proteínas estando envolvido na maquinaria de translocação. *Tandem repeats* são importantes em chaperonas, pois de acordo com Edqvist et al. (2006) estas repetições estão envolvidas na manutenção da função e estabilidade destas proteínas. A sequência repetitiva identificada na proteína hipotética é apresentada em destaque na Figura 31.

VI. CONCLUSÃO

A partir dos levantamentos realizados acerca de ferramentas para identificação de repetições adjacentes, as ferramentas Tandem Repeat Finder e MREPS se mostraram as mais adequadas para utilização. Ambas mostraram-se eficientes ao objetivo proposto.

O *pipeline* desenvolvido mostrou-se importante para o presente trabalho, pois além de automatizar os processos de identificação, armazenamento e consulta de repetições adjacentes em um banco de dados, a mesma apresenta uma interface amigável. Com o uso da ferramenta foi possível traçar o perfil das repetições encontradas, as quais mostraram que não existe correlação entre o tamanho do genoma e a quantidade de repetições identificadas. Também foi possível traçar os padrões mais prevalentes em todas espécies, mostrando que o *repeat* CAG é o mais conservado dentre fungos dermatófitos. Observou-se ainda que o *repeat* CAG/CAA é o mais rico e altamente variável e conservado em genes relacionados a regulação.

Um gene de *T. rubrum* (TERG_08771) contendo sequências repetitivas, com função não conhecida foi induzido em meio de cultura contendo queratina em um experimento de *microarray* do nosso grupo de pesquisa. Através da análise *in silico* dessa sequência repetitiva no presente pode-se inferir que esse gene está relacionado ao processo de adesão celular por conter sequências características de adesinas. Foi observado também que este gene foi induzido *in vitro* quando *T. rubrum* foi co-cultivado na presença de uma linhagem de queratinócitos humanos. Dessa forma, esse gene se mostrou um importante alvo que vai ser investigado em experimentos futuros em nosso grupo de pesquisa a fim de comprovar a função do gene, relacionado à adesão e patogenicidade em *T. rubrum*.

Foram identificados e categorizados diversos genes com as regiões repetitivas variáveis em *T. rubrum*. Os genes ricos em repetições variáveis estão relacionados a diferentes processos metabólicos como fatores de transcrição, parede celular, adesão celular, entre outros, ressalta-se que genes selecionados foram comparados com o genoma de dermatófitos, sendo possível apontar diferenças dentre as espécies e linhagens da mesma espécie. Ainda observou-se que genes ricos em repetições variáveis foram preditos com características estruturais de adesinas.

Este estudo sobre repetições adjacentes em fungos recentemente sequenciados é um passo em direção a uma melhor compreensão da natureza dessas importantes sequências. A

partir deste trabalho, acredita-se que repetições adjacentes são importantes fontes de variação fenotípica, mostrando-se potencialmente relacionadas a fatores de patogenicidade em fungos dermatófitos. Apesar da disposição dos dados do sequenciamento do genoma dos principais fungos, ainda não há uma compreensão completa sobre as bases moleculares da função dessas sequências, sendo necessário o compartilhamento de todos dados gerados para uma melhor elucidação dos mecanismos relacionados a patogenicidade e outros processos metabólicos.

REFERÊNCIAS

- AQUINO, V.; CONSTANTE, C.; BAKOS, L. Frequência das dermatofitoses em exames micológicos no Hospital Geral de Porto Alegre, Brasil. **Anais Bras Dermatol.**, v. 82, p. 239-244, 2007.
- ASHBURNER, M. et al. *Gene Ontology*: tool for the unification of biology. The *Gene Ontology Consortium*. **Nat Genet**, v. 25, p. 25-29, 2000.
- ATANESYAN, L. et al. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. **Biological Chemistry**, v. 393, p. 63–70, 2012.
- BALDO, A. et al. Mechanisms of skin adherence and invasion by dermatophytes. **Mycoses**, v. 55, p. 218-223, 2012.
- BEAUSSART, A. et al. Single-molecule imaging and functional analysis of Als adhesins and mannans during *Candida albicans* morphogenesis. **ACS Nano**, v. 6, p. 10950-64, 2012.
- BENSON, G. Tandem repeats Finder: a program to analyze DNA sequences. **Nucleic Acids Res.**, v. 27, p. 573–580, 1999.
- BERISIO, R.; VITAGLIANO, L. Polyproline and triple helix motifs in host-pathogen recognition. **Current protein & peptide science**, v. 13, p. 855-65, 2012.
- BJORKLUND, A. K.; EKMAN, D.; ELOFSSON, A. Expansion of Protein Domain Repeats. **PLoS Comput Biol**, v. 2, n. 8, p. e114, 2006.
- BOEYNAEMS, S. **Functional Consequences of Variable Tandem Repeats within the Yeast Cyc8 Transcriptional Regulator**. 2012. 112 f. Dissertação (Mestrado em Bioengenharia) - Katholieke Universiteit Leuven, Leuven.
- BROAD INSTITUTE*. Dermatophyte Comparative Sequencing Project - *Broad Institute of Harvard and MIT*. **BROAD INSTITUTE**, 2014. Disponível em: <www.broadinstitute.org/annotation/genome/dermatophyte_comparative>. Acesso em mar. 2014.
- BROUTA, F. et al. Secreted metalloprotease gene family of *Microsporium canis*. **Infect Immun.**, n. 70, p. 5676-5683, 2002.
- BROWN, T. **Genomes - The Repetitive DNA Content of Genomes**. 2. ed. Manchester: Oxford: Wiley-Liss, 2002. 520 p.
- BURMESTER, A. et al. Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. **Genome Biol**, v. 12, p. R7, 2011.
- CHAFFIN, W. L. et al. Cell Wall and Secreted Proteins of *Candida albicans*: Identification, Function, and Expression. **Microbiology and Molecular Biology Reviews**, v. 62, p. 130–180, 1998.
- CONESA, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, p. 3674-3676, 2005.

- DE GROOT, P. et al. Adhesins in Human Fungal Pathogens: Glue with Plenty of Stick. **Eukaryot Cell**, v. 12, p. 470–481, 2013.
- DEGREEF, H. Clinical forms of dermatophytosis (ringworm infection). **Mycopathologia**, n. 166, p. 257–265, 2008.
- DENISON, S. H. pH regulation of gene expression in fungi. **Fungal Genet. Biol.**, v. 29, p. 61-71, 2000.
- DONG, X. et al. Roles of adherence and matrix metalloproteinases in growth patterns of fungal pathogens in cornea. **Curr Eye Res**, v. 30, p. 613-620, 2005.
- DUITAMA, J. et al. Large-scale analysis of tandem repeat variability in the human genome. **Nucleic Acids Research**, v. 42, p. 5728-41, 2014.
- EDQVIST, P. et al. Tetratricopeptide repeats in the type III secretion chaperone, LcrH: their role in substrate binding and secretion. **Mol Microbiol.**, v. 59, p. 31-44, 2006.
- ELLEGREN, H. Microsatellites: simple sequences with complex. **Nat Rev Genet**, v. 5, p. 435–445, 2004.
- FACHIN, A. L. et al. Role of the ABC transporter TruMDR2 in terbinafine, 4-nitroquinoline N-oxide and ethidium bromide susceptibility in *Trichophyton rubrum*. **J Med Microbiol**, v. 55, p. 1093-9, 2006.
- FERREIRA-NOZAWA, M. S. et al. The pH signaling transcription factor *PacC* mediates the growth of *Trichophyton rubrum* on human nail in vitro. **Med Mycol**, v. 44, p. 641-645, 2006.
- FONDON, J.; GARNER, H. Molecular origins of rapid and continuous morphological evolution. **Proc Natl Acad Sci U S A**, v. 101, p. 18058–18063, 2004.
- FRANK, A. T. et al. Structure and Function of Glycosylated Tandem Repeats from *Candida albicans* Als Adhesins. **Eukaryotic Cell**, v. 9, p. 405-414, 2010.
- GARBER, G. An overview of fungal infections. **Drugs**, v. 61, p. 1-12, 2011.
- GEMAYEL, R. et al. Variable tandem repeats accelerate evolution of coding and regulatory sequences. **Annu. Rev. Genet**, v. 44, p. 445–477, 2010.
- GEMAYEL, R. et al. Beyond Junk-Variable Tandem Repeats as Facilitators of Rapid Evolution of Regulatory and Coding Sequences. **Genes**, v. 3, p. 461-480, 2012.
- GIBBONS, J. G.; ROKAS, A. Comparative and Functional Characterization of Intragenic Tandem Repeats in 10 *Aspergillus* Genomes. **Mol. Biol. Evol.**, v. 26, p. 591–602, 2009.
- GIBBS, S. et al. Molecular characterization and evolution of the SPRR family of keratinocyte differentiation markers encoding small proline-rich proteins. **Genomics**, v. 16, p. 630-637, 1993.
- GONZÁLEZ, M.; BRITO, N.; GONZÁLEZ, C. High abundance of Serine/Threonine-rich regions predicted to be hyper-O-glycosylated in the secretory proteins coded by eight fungal genomes. **BMC Microbiology**, v. 12, 2012.

- GRUMBT, M.; MONOD, M.; STAIB, P. Genetic advances in dermatophytes. **FEMS Microbiology Letters**, v. 320, p. 79–86, 2011.
- HAY, R. J. The future of onychomycosis therapy may involve a combination of approaches. **The British journal of dermatology**, v. 145, n. 60, p. 3-8, 2001.
- HEVIA, A. et al. An extracellular Serine/Threonine-rich protein from *Lactobacillus plantarum* NCIMB 8826 is a novel aggregation-promoting factor with affinity to mucin. **Appl Environ Microbiol.**, v. 79, p. 6059-66, 2013.
- HOYER, L. L. et al. Discovering the Secrets of the *Candida albicans* Agglutinin-Like Sequence (ALS) Gene Family—a Sticky Pursuit. **Med Mycol.**, v. 46, p. 1-15, 2008.
- HUANG, H. et al. The molecular chaperone DnaJ is required for the degradation of a soluble abnormal protein in *Escherichia coli*. **J Biol Chem**, v. 276, p. 3920-3928, 2001.
- HUNTLEY, M. A.; CLARK, A. G. Evolutionary analysis of amino acid repeats across the genomes of 12 *Drosophila* species. **Mol. Biol. Evol.**, v. 24, p. 2598-2609, 2007.
- JOUSSON, O. et al. Secreted subtilisin gene family in *Trichophyton rubrum*. **Gene**, v. 339, p. 79-88, 2004.
- KANNAN, P.; JANAKI, C.; SELVI, G. Prevalence of Dermatophytes and Other Fungal Agents Isolated from Clinical Samples. **Indian Journal of Medical Microbiology**, v. 24, p. 212-215, 2006.
- KARAOGLU, H.; LE, C. M.; MEYER, W. Survey of simple sequence repeats in completed fungal genomes. **Mol. Biol**, v. 22, p. 639–649, 2005.
- KARUNANITHI, S. et al. Shedding of the Mucin-Like Flocculin Flo11p Reveals a New Aspect of Fungal Adhesion Regulation. **Current Biology**, v. 20, 2010.
- KAUFMAN, G. et al. Markers for host-induced gene expression in *Trichophyton* dermatophytosis. **Infect. Immun.**, v. 73, p. 6584-6590, 2005.
- KOLPAKOV, R.; BANA, B.; KUCHEROV, G. Mreps: efficient and flexible detection of tandem repeats in DNA. **Nucleic Acids Res**, v. 31, p. 3672–3678, 2003.
- LECLERCQ, S.; RIVALS, E.; JARNE, P. Detecting microsatellites within genomes: significant variation among algorithms. **BMC Bioinformatics**, v. 8, 2007.
- LEGENDRE, M. et al. Sequence-based estimation of minisatellite and microsatellite repeat variability. **Genome Research**, v. 17, p. 1787–1796, 2007.
- LENG, W. et al. Proteomic profile of dormant *Trichophyton rubrum* conidia. **BMC Genomics**, v. 9, 2008.
- LEVDANSKY, E. et al. Coding Tandem Repeats Generate Diversity in *Aspergillus fumigatus* Genes. **Eukaryotic Cell**, v. 6, p. 1380-1391, 2007.

- LEVDANSKY, E. et al. 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. **Eukaryot Cell**, v. 9, p. 1403–1415, 2010.
- LEVDANSKY, E.; SHARON, H.; OSHEROV, N. Coding fungal tandem repeats as generators of fungal diversity. **Fungal Biology Reviews**, v. 22, p. 85–96, 2008.
- LI, J.; PAN, Y.; LIU, G. Disruption of the nitrogen regulatory gene *AcareA* in *Acremonium chrysogenum* leads to reduction of cephalosporin production and repression of nitrogen metabolism. **Fungal Genet Biol**, v. 61, p. 69-79, 2013.
- LIM, K. G. et al. Review of tandem repeat search tools: a systematic approach to evaluating algorithmic performance. **Briefings in Bioinformatics**, v. 23, p. 1-15, 2012.
- LIMJINDAPORN, T.; KHALAF, R. A.; FONZI, W. A. Nitrogen metabolism and virulence of *Candida albicans* require the GATA-type transcriptional activator encoded by *GAT1*. **Mol Microbiol**, v. 50, p. 993-1004, 2003.
- LIU, T. et al. Analysis of gene expression changes in *Trichophyton rubrum* after skin interaction. **Journal of Medical Microbiology**, v. 63, p. 642–648, 2014.
- MAICAS, S. et al. *In silico* analysis for transcription factors with Zn(II)(2)C(6) binuclear cluster DNA-binding domains in *Candida albicans*. **Comp Funct Genomics**, v. 6, p. 345–356, 2005.
- MARANHÃO, F. C. et al. Membrane transporter proteins are involved in *Trichophyton rubrum* pathogenesis. **J Med Microbiol**, v. 58, n. 2, p. 163-8, 2009.
- MARCHLER-BAUER, A. et al. CDD: a Conserved Domain Database for the functional annotation of proteins. **Nucleic Acids Research**, v. 39, p. D225–D229, 2011.
- MARCONI, V. C. et al. Disseminated dermatophytosis in a patient with hereditary hemochromatosis and hepatic cirrhosis: case report and review of the literature. **Med Mycol**, v. 48, p. 518-27, 2010.
- MARTIN, P. et al. Microsatellite instability regulates transcription factor binding and gene expression. **Proc Natl Acad Sci USA**, v. 10, p. 3800–3804, 2005.
- MARTINEZ, D. A. et al. Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. **MBio**, v. 5, p. e00259–12, 2012.
- MARTINEZ-ROSSI, N. M. et al. Role of pH in the pathogenesis of dermatophytoses. **Mycoses**, v. 55, n. 5, p. 381-7, 2012.
- MARTINEZ-ROSSI, N. M. et al. Molecular aspects of dermatophyte-host interactions. **Fungi in Human and Animal Health**, p. 143-165, 2004.
- MAYER, C.; LEESE, F.; TOLLRIAN, R. Genome-wide analysis of tandem repeats in *Daphnia pulex* - a comparative approach. **BMC Genomics**, v. 11, p. 277, 2010.
- MCBRIDE, A. E. et al. Specific sequences within arginine–glycine-rich domains affect mRNA-binding protein function. **Nucl. Acids Res.**, v. 37, p. 4322-4330, 2009.

- MCMANUS, B.; COLEMAN, D. Molecular epidemiology, phylogeny and evolution of *Candida albicans*. **Infect Genet Evol.**, v. 21, p. 166–178, 2014.
- MERKEL, A.; GEMMELL, N. Detecting short tandem repeats from genome data: opening the software black box. **Brief Bioinformatics**, v. 9, p. 355–366., 2008.
- METZGAR, D.; BYTOF, J.; WILLS, C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. **Genome Res.**, v. 10, p. 72–80, 2000.
- MICHAEL, T. et al. Simple Sequence Repeats Provide a Substrate for Phenotypic Variation in the *Neurospora crassa* Circadian Clock. **PLoS ONE**, v. 8, p. e795, 2007.
- MICROSOFT BIOLOGY FOUNDATION..**Net Bio**, 2013. Disponível em: <<https://bio.codeplex.com/>>. Acesso em: Jan. 2014.
- MIRKIN, S. Expandable DNA repeats and human disease. **Nature**, v. 447, p. 932–940, 2007.
- MORA-MONTES, H. M. Fungal cell wall: biosynthesis of glycoproteins and their importance in fungal biology. In: **Glycans: Biochemistry, Characterization and Applications**. New York: Editorial Nova Press, 2012. Cap. 2, p. 189-206.
- MOXON, R.; BAYLISS, C.; HOOD, D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. **Annu Rev Genet**, v. 40, p. 307-333, 2006.
- NCBI. Nucleotide database. **National Center for Biotechnology Information**, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/nucleotide>>. Acesso em jan. 2012.
- NENOFF, P. et al. Mycology – an update. Part 1: Dermatomycoses: Causative agents, epidemiology and pathogenesis. **JDDG: Journal der Deutschen Dermatologischen Gesellschaft**, v. 12, 2014.
- NHGRI. NHGRI Announces Latest Sequencing Targets, 2006. Disponível em: <<http://www.genome.gov/19517271>>. Acesso em: fev 2012.
- NIKLAUS, F. et al. Surface antigens and potential virulence factors from parasites detected by comparative genomics of perfect amino acid repeat. **Proteome Science**, v. 5, 2007.
- NOBILE, C. et al. Complementary Adhesin Function in *C. albicans* Biofilm Formation. **Current Biology**, v. 18, p. 1017–1024, 2008.
- NUSSBAUM, R. L.; MCINNES, R. R.; WILLARD, H. F. **Thompson & Thompson Genética Médica**. 7ª. ed. Rio de Janeiro: Elsevier, 2008. 640 p.
- OH, S. et al. Functional specificity of *Candida albicans* Als3p proteins and clade specificity of ALS3 alleles discriminated by the number of copies of the tandem repeat sequence in the central domain. **Microbiology**, v. 151, p. 673-681, 2005.
- OHST, T. et al. Origins of microsatellite diversity in the *Trichophyton rubrum* – *T. violaceum* clade (Dermatophytes). **J Clin Microbiol**, v. 42, p. 4444–4448, 2004.

- OREJAS, M. et al. Activation of the *Aspergillus PacC* transcription factor in response to alkaline ambient pH requires proteolysis of the carboxy-terminal moiety. **Gene and Development**, v. 9, p. 1622-1632, 1995.
- PAIÃO, F. G. et al. Analysis of *Trichophyton rubrum* gene expression in response to cytotoxic drugs. **FEMS Microbiol Lett**, v. 271, p. 180-186, 2007.
- PARK, E. et al. Collagen triple helix repeat containing-1 promotes pancreatic cancer progression by regulating migration and adhesion of tumor cells. **Carcinogenesis**, v. 34, p. 694-702, 2013.
- PELLEGRINI, M.; RENDA, M. E.; VECCHIO, A. TRStalker: an efficient heuristic for finding fuzzy tandem repeats. **Bioinformatics**, v. 26, n. 12, p. i358-66, 2010.
- PERES, N. T. A. et al. Dermatofitos: interação patógeno-hospedeiro e resistência a antifúngicos. **An. Bras. Dermatol.**, v. 85, p. 657-667, 2010.
- PERSINOTI, G. et al. RNA-sequencing analysis of *Trichophyton rubrum* transcriptome in response to sublethal doses of acriflavine. **BMC Genomics**, v. 15, 2014.
- PUNTA, M. et al. The Pfam protein families database. **Nucleic Acids Research**, v. 40, p. 290-301, 2012.
- RAMANA, J.; GUPTA, D. FaaPred: A SVM-Based Prediction Method for Fungal Adhesins and Adhesin-Like Proteins. **Plos One**, v. 5, p. e9695, 2010.
- RANDO, O.; VERSTREPEN, K. Timescales of Genetic and Epigenetic Inheritance. **Cell**, v. 128, p. 655-668, 2007.
- RAUCEO, J. M. et al. Threonine-rich repeats increase fibronectin binding in the *Candida albicans* adhesin Als5p. **Eukaryot Cell**, v. 5, p. 1664-73, 2006.
- RICHARD, G.-F.; DUJON, B. Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. **Gene**, v. 174, p. 165-174, 1996.
- RICHARD, G.-F.; DUJON, B. Molecular evolution of minisatellites in *hemiascomycetous* yeasts. **Mol. Biol. Evol.**, v. 23, p. 189-202, 2006.
- RICHARD, G.-F.; KERREST, A.; DUJON, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. **Microbiol Mol Biol Rev**, v. 72, p. 686-727, 2008.
- RUEPP, A. et al. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. **Nucleic Acids Res.**, v. 32, p. 5539-5545, 2004.
- SAMPSON, S. L. Mycobacterial PE/PPE Proteins at the Host-Pathogen Interface. **Clinical and Developmental Immunology**, v. 2011, 11 p., 2011.
- SAWAYA, S. et al. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. **PLoS ONE**, v. 8, p. e54710, 2013.

- SCHAEFER, M. H.; WANKER, E. E.; ANDRADE-NAVARRO, M. A. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. **Nucleic Acids Res**, v. 40, p. 4273-87, 2012.
- SCHOELER, A. P. et al. Prevalência de dermatófitos na rotina de micologia em hospital particular de médio porte na cidade de Chapecó, estado de Santa Catarina, Brasil. **Rev Ciênc Farm Básica Apl.**, v. 31, p. 103-106, 2010.
- SHARON, I. et al. Correcting BLAST e-values for low-complexity segments. **J Comput Biol**, v. 12, n. 7, p. 980-1003, 2005.
- SILVEIRA, H. et al. Transcriptional profiling reveals genes in the human pathogen *Trichophyton rubrum* that are expressed in response to pH signaling. **Microbial Pathogenesis**, v. 48, p. 91-96, 2010.
- SILVEIRA-GOMES, F. et al. Dermatophytosis diagnosed at the Evandro Chagas Institute, Pará, Brazil. **Braz J Microbiol.**, v. 44, p. 443-446, 2013.
- SINGH, R. et al. Mining and survey of simple sequence repeats in wheat rust *Puccinia* sp. **Bioinformatics**, v. 7, p. 291-295, 2011.
- SIQUEIRA, E. R. et al. Ocorrência de dermatófitos em amostras de unhas, pés e mãos coletadas de estudantes universitários. **Rev. Soc. Bras. Med. Trop.**, v. 39, n. 3, p. 269-271, 2006.
- SMITH, E. S.; FLEISCHER, J. A. B.; FELDMAN, S. R. Nondermatologists are more likely than dermatologists to prescribe antifungal/corticosteroid products: an analysis of office visits for cutaneous fungal infections. **J Am Acad Dermatol**, v. 39, p. 43-47, 1998.
- SMITH, T. F.; WATERMAN, M. S. Identification of Common Molecular Subsequences. **Journal of Molecular Biology**, n. 171, p. 195-197, 1981.
- SPATZ, S. J.; SILVA, R. F. Sequence determination of variable regions within the genomes of *Gallid herpesvirus-2* pathotypes. **Archives of Virology**, v. 152, p. 1665-1678, 2007.
- STAIB, P. et al. Host versus in vitro signals and intrastrain allelic differences in the expression of a *Candida albicans* virulence gene. **Mol. Microbiol.**, v. 44, p. 1351-1366, 2002.
- SZKLARCZYK, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. **Nucleic Acids Res**, v. 43, p. D447-52, 2015.
- THIERRY, A. et al. Megasatellites: a peculiar class of giant minisatellites in genes involved in cell adhesion and pathogenicity in *Candida glabrata*. **Nucleic Acids Res**, v. 36, p. 5970-5982, 2008.
- TIAN, C.; LI, J.; GLASS, L. Exploring the bZIP transcription factor regulatory network in *Neurospora crassa*. **Microbiology**, v. 157, p. 747-759, 2001.
- TODD, R. B. et al. Prevalence of transcription factors in *ascomycete* and *basidiomycete* fungi. **BMC Genomics**, v. 15, 2014.

TRED. *Trichophyton rubrum* Expression Database. Disponível em: <www.mgc.ac.cn/TrED>. Acesso em: fev. 2012.

TUDZYNSKI, B. Nitrogen regulation of fungal secondary metabolism in fungi. **Front Microbiol**, v. 5, p. 656, 2014.

USDIN, K. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. **Genome Res**, v. 18, p. 1011–9, 2008.

VERMEULEN, T.; SCHOONBEEK, H.; DE WAARD, M. A. The ABC transporter BcatrB from *Botrytis cinereais* a determinant of the activity of the phenylpyrrole fungicide fludioxonil. **Pest Manag Sci**, v. 57, p. 393-402, 2001.

VERSTREPEN, K. et al. Intragenic tandem repeats generate functional variability. **Nat Genet.**, v. 37, p. 986–990, 2005.

VERSTREPEN, K.; KLIS, F. Flocculation, adhesion and biofilm formation in yeasts. **Mol Microbiol.**, v. 60, p. 5-15, 2006.

VINCES, M. D. et al. Unstable tandem repeats in promoters confer transcriptional evolvability. **Science**, v. 324, p. 1213-1216, 2009.

WANG, C.; LEGER, R. J. S. The MAD1 Adhesin of *Metarhizium anisopliae* Links Adhesion with Blastospore Production and Virulence to Insects, and the MAD2 Adhesin Enables Attachment to Plants. **EUKARYOTIC CELL**, v. 6, p. 808–816, 2007.

WANG, L. et al. Analysis of the dermatophyte *Trichophyton rubrum* expressed sequence tags. **BMC Genomics**, v. 11, 2006.

WEITZMAN, I.; SUMMERBELL, R. C. The dermatophytes. **Clinical microbiology reviews**, v. 8, p. 240–259, 1995.

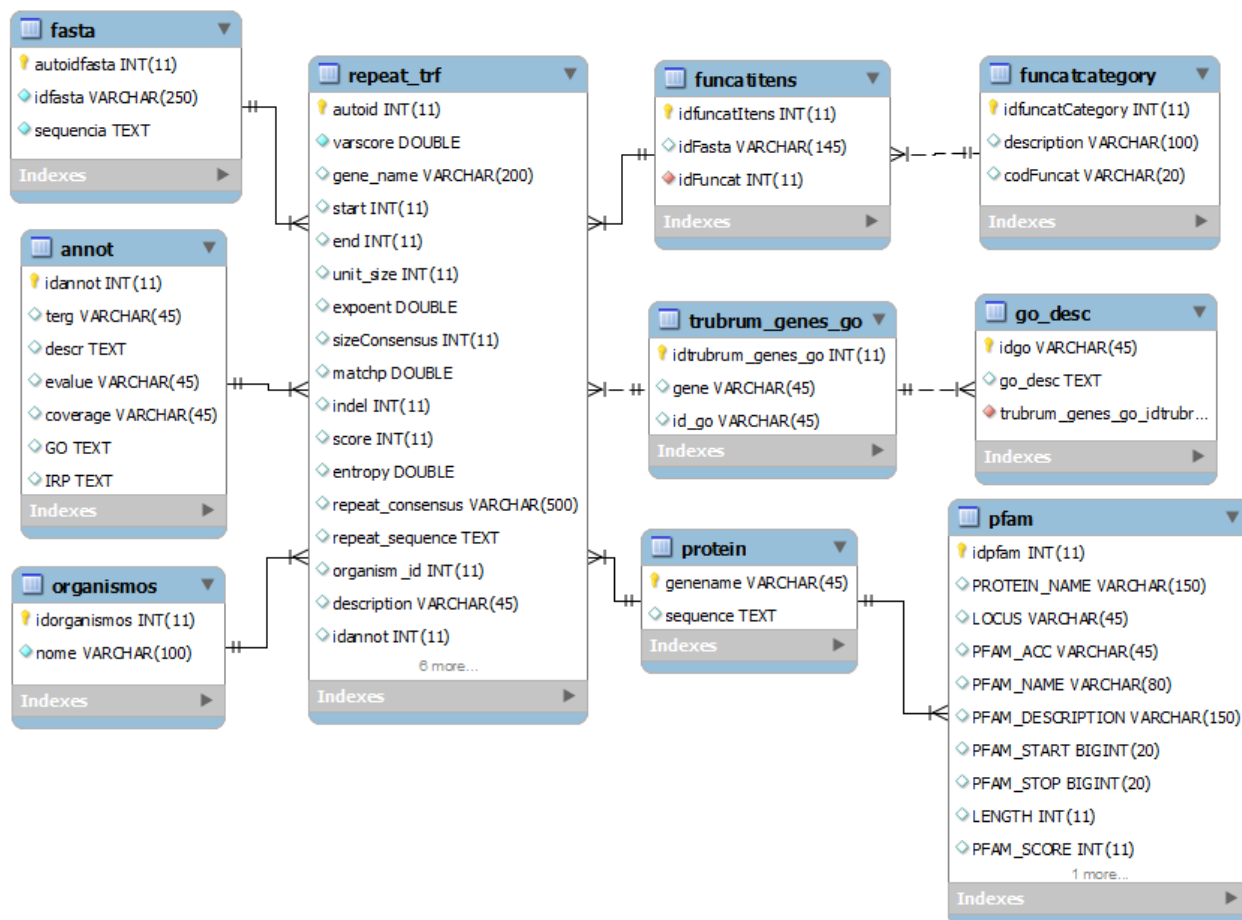
WHITE, T. C. et al. Generating and testing molecular hypotheses in the dermatophytes. **Eukaryotic Cell**, v. 7, p. 1238-1245, 2008.

YANG, J. et al. TrED: the *Trichophyton rubrum* Expression Database. **BMC Genomics**, v. 8, p. 1-6, 2007.

YE, W. et al. Phylogenetic and transcriptional analysis of an expanded bZIP transcription factor family in *Phytophthora sojae*. **BMC Genomics**, v. 14, 2013.

ZHAO, R. et al. Global transcriptional profiles of *Trichophyton rubrum* in response to Flucytosine. **Science in China Series C: Life Sciences**, v. 52, p. 1173-1185, 2009.

APÊNDICE A - Diagrama Entidade Relacionamento do Banco de Dados



DESCRIÇÃO DAS TABELAS

repeat_trf: Armazena as repetições identificadas.

organismos: Armazena a descrição dos organismos armazenados.

annot: Armazena a tentativa de anotação das sequências utilizando a ferramenta Blast2GO

fasta: Armazena as sequências fasta dos organismos.

funcatitens: Armazena a descrição dos itens da tabela funcat.

funcatcategory: Armazena as categorias da tabela funcat.

trubrum_genes_go: Relação entre termos GO e genes de *T. rubrum*

go_desc: Descrição dos itens do *Gene Ontology*.

protein: Sequência de aminiácidos relativos a tradução de um gene.

pfam: Itens da tabela pfam relacionados com proteínas de *T. rubrum*

**APÊNDICE B - Lista de genes com repetições adjacentes variáveis (varscore >0) com
predição de adesão celular**

Sequence IDs	Putative Annotation	SVM Score	Decision
TERG_07042	hypothetical protein	2,0523744	Adhesin
TERG_08199	hypothetical protein	1,9231888	Adhesin
TERG_05644	mucin family signaling protein	1,8640077	Adhesin
TERG_01834	hypothetical protein	1,4091427	Adhesin
TERG_08882	c2h2 transcription factor	1,3841832	Adhesin
TERG_02143	protein	1,3545475	Adhesin
TERG_08369	serine arginine repetitive matrix protein 1	1,3172438	Adhesin
TERG_05946	hypothetical protein	1,1555167	Adhesin
TERG_11802	mucin	1,1377928	Adhesin
TERG_00743	sh3 domain protein	1,1090786	Adhesin
TERG_05576	cell wall glucanase	1,0605741	Adhesin
TERG_11911	hypothetical protein TERG_11911	1,0598959	Adhesin
TERG_08033	hypothetical protein TERG_08033	1,0519579	Adhesin
TERG_08771	extracellular serine-threonine rich protein	0,99703817	Adhesin
TERG_12213	hypothetical protein H100_03617	0,99236463	Adhesin
TERG_06875	hypothetical protein TERG_06875	0,97894446	Adhesin
TERG_03333	transcription factor	0,97843649	Adhesin
TERG_00294	hypothetical protein H100_07720	0,95410456	Adhesin
TERG_02079	hypothetical protein H100_06753	0,93610651	Adhesin
TERG_03536	hydroxyproline-rich glycoprotein dz-hrgp	0,93526011	Adhesin
TERG_11718	hypothetical protein H100_08930	0,89714494	Adhesin
TERG_00791	developmental regulatory protein	0,8450012	Adhesin
TERG_04645	pre-mrna splicing factor	0,80199043	Adhesin
TERG_02123	hypothetical protein TERG_02123	0,79417634	Adhesin
TERG_05067	hypothetical protein TERG_05067	0,76636899	Adhesin
TERG_00721	hypothetical protein TERG_00721	0,74350085	Adhesin
TERG_00388	domain protein	0,73199659	Adhesin
TERG_04451	forkhead transcription factor	0,70603195	Adhesin
TERG_02163	hlh transcription factor	0,70176279	Adhesin
TERG_04630	hypothetical protein H106_08920	0,67875565	Adhesin
TERG_08421	ph signal transduction protein	0,62242223	Adhesin
TERG_08724	hypothetical protein H101_00512	0,61372983	Adhesin
TERG_00395	bag domain protein	0,60951655	Adhesin
TERG_12175	hypothetical protein H113_04970	0,60214402	Adhesin
TERG_07083	hypothetical protein H105_01451	0,5923917	Adhesin
TERG_00798	zinc knuckle domain protein	0,58916654	Adhesin
TERG_04914	spo7-like protein	0,58602171	Adhesin
TERG_02674	hypothetical pathogenicity factor	0,5547985	Adhesin
TERG_03659	hypothetical protein H105_05975	0,55276849	Adhesin
TERG_07146	3-demethylubiquinone-9 3-o-methyltransferase	0,54511646	Adhesin
TERG_06537	cyclin-dependent protein kinase regulator pho80	0,53794418	Adhesin
TERG_05755	hypothetical protein TERG_05755	0,51025802	Adhesin
TERG_00153	actin associated protein wsp1	0,50594786	Adhesin

Sequence IDs	Putative Annotation	SVM Score	Decision
TERG_00223	hypothetical protein TERG_00223	0,49318487	Adhesin-like
TERG_00754	hypothetical protein H100_08207	0,48014305	Adhesin-like
TERG_07681	sexual development activator	0,46843153	Adhesin-like
TERG_12443	hypothetical protein H100_01955	0,46695214	Adhesin-like
TERG_02972	srf-type transcription factor	0,37469145	Adhesin-like
TERG_01667	domain-containing protein	0,33246213	Adhesin-like
TERG_03735	hypothetical protein H106_05875	0,30783702	Adhesin-like
TERG_05501	wd repeat protein	0,30619461	Adhesin-like
TERG_07370	major outer envelope glyco protein	0,299171	Adhesin-like
TERG_11820	cyclin-dependent protein kinase complex	0,2931186	Adhesin-like
TERG_02551	hypothetical protein TERG_02551 bacteriodes thetaiotaomicron symbiotic	0,28572523	Adhesin-like
TERG_05663	chitinase	0,24496949	Adhesin-like
TERG_01441	hypothetical protein H105_08897	0,23663791	Adhesin-like
TERG_12201	hypothetical protein H113_03589	0,22667972	Adhesin-like
TERG_02312	sister chromatid cohesion protein	0,20263882	Adhesin-like
TERG_08430	hypothetical protein	0,20006094	Adhesin-like
TERG_08825	bzip transcription factor idi- partial	0,19634861	Adhesin-like
TERG_00981	hypothetical protein H107_08575	0,18687024	Adhesin-like
TERG_02224	rho guanyl nucleotide exchange	0,18262623	Adhesin-like
TERG_02893	hydrolase	0,1666496	Adhesin-like
TERG_00334	hypothetical protein TERG_00334	0,15787624	Adhesin-like
TERG_00802	hypothetical protein TERG_00802	0,15531283	Adhesin-like
TERG_02801	hypothetical protein TERG_02801	0,14909767	Adhesin-like
TERG_08692	nitrogen regulatory protein area	0,14637631	Adhesin-like
TERG_04424	myb dna-binding domain-containing protein	0,13329505	Adhesin-like
TERG_01458	inositol hexaphosphate kinase	0,13093007	Adhesin-like
TERG_06432	c2h2 transcription factor	0,10868884	Adhesin-like
TERG_03110	gtpase activating protein	0,1056002	Adhesin-like
TERG_06295	c2h2 conidiation transcription factor	0,082481229	Adhesin-like
TERG_07604	kh domain-containing protein	0,08010368	Adhesin-like
TERG_05733	salivary glue protein sgs-3	0,07122848	Adhesin-like
TERG_08164	protein bni4	0,066311765	Adhesin-like
TERG_05751	leucine rich repeat domain-containing protein	0,063155583	Adhesin-like
TERG_06989	c6 finger domain	0,063129034	Adhesin-like
TERG_04287	hypothetical protein TERG_04287	0,051426217	Adhesin-like
TERG_04158	ars binding protein	0,049945732	Adhesin-like
TERG_07081	protein ssh4	0,043033944	Adhesin-like
TERG_04109	c3hc4 finger protein	0,032305373	Adhesin-like
TERG_04005	hypothetical protein H100_04232	0,024018161	Adhesin-like
TERG_01442	hypothetical protein H105_08898	0,021446676	Adhesin-like
TERG_05812	hypothetical protein TERG_05812	0,006513126	Adhesin-like
TERG_05021	c6 transcription factor	8,81E-05	Adhesin-like
		-	
TERG_11879	protein	0,019041971	Adhesin-like
TERG_04777	exocyst complex component	-	Adhesin-like

Sequence IDs	Putative Annotation	SVM Score	Decision
		0,019134688	
		-	
TERG_05082	hypothetical protein TERG_05082	0,022052918	Adhesin-like
		-	
TERG_06080	serine threonine protein kinase	0,030313871	Adhesin-like
		-	
TERG_02629	lyr family protein	0,036569463	Adhesin-like
TERG_05645	vacuolar protein sorting protein	-0,03847953	Adhesin-like
		-	
TERG_05083	adaptin ear-binding coat-associated protein 1	0,044665754	Adhesin-like
TERG_00447	hypothetical protein H106_07703	-0,04574321	Adhesin-like
		-	
TERG_05340	swi-snf complex subunit	0,053568427	Adhesin-like
		-	
TERG_02489	sister chromatid cohesion acetyltransferase eco1	0,062644907	Adhesin-like
		-	
TERG_12713	hypothetical protein H100_00118	0,062811868	Adhesin-like
		-	
TERG_08238	copper-activated transcription factor	0,066195993	Adhesin-like
		-	
TERG_02902	hypothetical protein TERG_02902	0,067811034	Adhesin-like
		-	
TERG_03801	camk protein kinase	0,081337583	Adhesin-like
		-	
TERG_05465	ras-like gtp-binding	0,094281169	Adhesin-like
		-	
TERG_08760	hypothetical protein TRV_04385	0,096881747	Adhesin-like
TERG_08831	gnat family	-0,1026438	Adhesin-like
TERG_08766	hypothetical protein TERG_08766	-0,10826487	Adhesin-like
TERG_08596	vegetative cell wall protein gp1	-0,12305606	Adhesin-like
TERG_06698	aspartic acid-rich protein aspolin1	-0,13956564	Adhesin-like
TERG_06002	nima-interacting protein	-0,1428421	Adhesin-like
TERG_05396	bzip transcription factor	-0,15047859	Adhesin-like
TERG_05189	arginine rich protein	-0,15520991	Adhesin-like
TERG_02871	cue domain-containing protein	-0,16291058	Adhesin-like
TERG_04877	ph-response regulator protein palf rim8	-0,16426932	Adhesin-like
TERG_00846	efr3_cocim ame: full=protein efr3	-0,16696428	Adhesin-like
TERG_04095	hypothetical protein H100_04326	-0,16979893	Adhesin-like
TERG_08099	protein phosphatase 2c	-0,17438562	Adhesin-like
TERG_07257	hypothetical protein TERG_07257	-0,17623974	Adhesin-like
TERG_12581	hypothetical protein H100_01281	-0,1782169	Adhesin-like
TERG_03130	c2h2 transcription factor	-0,18263381	Adhesin-like
TERG_12309	transcriptional repressor sin3p	-0,19540502	Adhesin-like
TERG_11731	zinc cadmium resistance protein	-0,19647502	Adhesin-like
	fungus zn binuclear cluster domain containing		
TERG_08173	protein	-0,19865198	Adhesin-like
TERG_07841	rna binding domain-containing protein	-0,22087921	Adhesin-like
TERG_03177	ring finger domain	-0,25474107	Adhesin-like

Sequence IDs	Putative Annotation	SVM Score	Decision
TERG_03967	tho2 protein	-0,25530817	Adhesin-like
TERG_04360	rna binding protein	-0,2610044	Adhesin-like
TERG_00226	hypothetical protein TERG_00226	-0,26107236	Adhesin-like
TERG_02675	oxysterol binding protein	-0,26507176	Adhesin-like
TERG_02390	wd repeat protein	-0,26709023	Adhesin-like
TERG_03740	beta- glucan synthetase	-0,26806227	Adhesin-like
TERG_08247	hypothetical protein TERG_08247	-0,26919947	Adhesin-like
TERG_02118	gpi anchored	-0,27846379	Adhesin-like
TERG_01459	conserved leucine-rich repeat protein	-0,28521627	Adhesin-like
TERG_05820	gamma-aminobutyric acid a receptor	-0,2892675	Adhesin-like
TERG_01567	ino80 chromatin remodeling complex	-0,29210739	Adhesin-like
TERG_07105	transcription factor tfiia complex subunit	-0,30095618	Adhesin-like
TERG_00509	protein	-0,30401575	Adhesin-like
TERG_07513	pheromone processing endoprotease kex2	-0,30917521	Adhesin-like
TERG_03810	pre-mrna-splicing factor 38b	-0,31084101	Adhesin-like
TERG_11565	hypothetical protein H100_07843	-0,31362782	Adhesin-like
TERG_01002	chaperone dnak	-0,31491334	Adhesin-like
TERG_06745	hypothetical protein H107_02044	-0,31700218	Adhesin-like
TERG_07719	hypothetical protein TRV_05460	-0,3274473	Adhesin-like
TERG_06545	hypothetical protein H100_01769	-0,33139	Adhesin-like
TERG_08792	hypothetical protein TERG_08792	-0,33905686	Adhesin-like
TERG_02691	hypothetical protein H107_07545	-0,34566949	Adhesin-like
TERG_03042	ste ste20 paka protein kinase	-0,34986614	Adhesin-like
TERG_04032	dna repair and transcription factor ada universal stress protein family domain- containing protein	-0,36021286	Adhesin-like
TERG_03196	cell cycle inhibitor	-0,36572876	Adhesin-like
TERG_06928	integral membrane protein	-0,37142146	Adhesin-like
TERG_06886	duf636 domain-containing protein	-0,37214344	Adhesin-like
TERG_06208	trna-specific adenosine-34 deaminase subunit	-0,37236337	Adhesin-like
TERG_06536	tad3	-0,37601525	Adhesin-like
TERG_12038	protein phosphatase regulatory subunit	-0,38069693	Adhesin-like
TERG_05790	hmg box protein	-0,38084866	Adhesin-like
TERG_00938	rna-binding protein	-0,39081737	Adhesin-like
TERG_03043	rna binding effector protein	-0,40029527	Adhesin-like
TERG_04844	gtpase activating protein for	-0,4013188	Adhesin-like
TERG_05556	hypothetical protein TERG_05556	-0,40383589	Adhesin-like
TERG_01015	domain partial	-0,40683592	Adhesin-like
TERG_06561	sex determining protein	-0,41706087	Adhesin-like
TERG_06667	hypothetical protein H100_01895	-0,43314121	Adhesin-like
TERG_06089	ph domain-containing protein	-0,43691283	Adhesin-like
TERG_08311	pre-mrna polyadenylation factor fip1	-0,44340762	Adhesin-like
TERG_03116	zf-parp-type zinc finger protein	-0,44527755	Adhesin-like
TERG_08231	mitochondrial chaperone atpase	-0,44588344	Adhesin-like
TERG_08209	btb poz domain-containing protein	-0,45208021	Adhesin-like
TERG_04778	ribosome assembly and transport protein	-0,45269723	Adhesin-like

Sequence IDs	Putative Annotation	SVM Score	Decision
TERG_08579	hypothetical protein TERG_08579	-0,45903546	Adhesin-like
TERG_02222	tbc domain-containing protein	-0,46354245	Adhesin-like
TERG_06692	integral membrane protein	-0,46390268	Adhesin-like
TERG_07407	c6 transcription factor	-0,46540235	Adhesin-like
TERG_02970	hypothetical protein TERG_02970	-0,47235432	Adhesin-like
TERG_00789	hypothetical protein H100_08242	-0,47298492	Adhesin-like
TERG_07499	ccch zinc finger and smr domain-containing protein	-0,48031361	Adhesin-like
TERG_07535	sam binding domain-containing protein	-0,4847218	Adhesin-like
TERG_04590	containing protein	-0,48875837	Adhesin-like
TERG_01067	dihydrofolate synthetase fol3	-0,49289179	Adhesin-like
TERG_11722	gtpase activating protein	-0,49326094	Adhesin-like
TERG_04743	hypothetical protein H100_08954	-0,50533627	Adhesin-like
TERG_07711	c6 finger domain	-0,5152798	Adhesin-like
TERG_08826	ras guanine-nucleotide exchange protein	-0,51896538	Adhesin-like
TERG_00908	hypothetical protein TERG_08826	-0,51986735	Adhesin-like
TERG_01250	carbohydrate-binding module family 48 protein	-0,52036296	Adhesin-like
TERG_08965	hypothetical protein H100_08721	-0,52880842	Adhesin-like
TERG_05101	hypothetical protein TERG_08965	-0,53146902	Adhesin-like
TERG_02055	myb dna binding protein	-0,53384009	Adhesin-like
TERG_03736	methylated-dna--protein-cysteine methyltransferase	-0,53490182	Adhesin-like
TERG_01971	transcriptional corepressor cyc8	-0,53581481	Adhesin-like
TERG_07937	hypothetical protein TERG_01971	-0,54442976	Adhesin-like
TERG_01472	g-patch domain-containing protein	-0,54553276	Adhesin-like
TERG_08488	glutamine repeat protein-1	-0,54987965	Adhesin-like
TERG_06235	fha domain-containing protein	-0,55036119	Adhesin-like
TERG_11699	hypothetical protein TERG_06235	-0,55231001	Adhesin-like
TERG_05557	outer membrane beta-barrel domain protein	-0,55873025	Adhesin-like
TERG_06972	cation efflux family protein	-0,56711225	Adhesin-like
TERG_00838	ph-response transcription factor <i>pacC</i> rim101 fungal zn binuclear cluster domain containing protein	-0,57114401	Adhesin-like
TERG_08171	protein	-0,57528784	Adhesin-like
TERG_06737	rnapii degradation factor def1	-0,5780014	Adhesin-like
TERG_08194	hypothetical protein TERG_08194	-0,57802021	Adhesin-like
TERG_01605	sensor histidine kinase response regulator	-0,57958634	Adhesin-like
TERG_00265	meiotically up-regulated gene 80 protein	-0,580939	Adhesin-like
TERG_08327	ser thr protein phosphatase	-0,58141265	Adhesin-like
TERG_02973	morphogenesis protein	-0,59200905	Adhesin-like
TERG_07931	vesicle-mediated transport protein	-0,59300943	Adhesin-like
TERG_12695	c6 transcription factor	-0,59492009	Adhesin-like
TERG_05539	tam domain methyltransferase protein	-0,60182703	Adhesin-like
TERG_02059	actin binding protein	-0,60404402	Adhesin-like
TERG_12563	hypothetical protein TRV_01248	-0,60972315	Adhesin-like
TERG_07820	1-phosphatidylinositol- -bisphosphate	-0,61461836	Adhesin-like

Sequence IDs	Putative Annotation	SVM Score	Decision
	phosphodiesterase		
TERG_04099	hypothetical protein TERG_04099	-0,61666075	Adhesin-like
TERG_02617	dna primase small subunit	-0,62121373	Adhesin-like
TERG_12269	dna-binding 42 kda	-0,62570143	Adhesin-like
TERG_06603	hypothetical protein H100_01829	-0,62693588	Adhesin-like
TERG_08253	hypothetical protein TERG_08253	-0,65078416	Adhesin-like
TERG_06192	phd finger and bah domain protein	-0,65134974	Adhesin-like
TERG_11764	hypothetical protein	-0,65682114	Adhesin-like
TERG_01953	60s acidic ribosomal protein p2	-0,65839268	Adhesin-like
TERG_02144	arsenite resistance protein ars2	-0,66051376	Adhesin-like
TERG_04196	ste ste20 paka protein kinase	-0,66256042	Adhesin-like
TERG_01949	hypothetical protein TERG_01949	-0,66628558	Adhesin-like
TERG_12717	conserved fungal protein	-0,66635051	Adhesin-like
TERG_01985	hypothetical protein TERG_01985	-0,66657634	Adhesin-like
TERG_00811	c-x8-c-x5-c-x3-h type zinc finger protein	-0,67717804	Adhesin-like
TERG_03634	hypothetical protein TERG_03634	-0,68612039	Adhesin-like
TERG_06952	rrna processing protein pwp1	-0,68718754	Adhesin-like
TERG_01825	dna repair protein	-0,68904448	Adhesin-like
TERG_04208	hal protein kinase	-0,69914719	Adhesin-like
TERG_07192	far upstream element-binding protein 3	-0,70065429	Adhesin-like
TERG_03795	sorting nexin-41	-0,70163938	Adhesin-like
TERG_11518	glucokinase	-0,70185793	Adhesin-like
TERG_06709	hypothetical protein TERG_06709	-0,70932949	Adhesin-like
TERG_05480	hypothetical protein H105_02865	-0,71661702	Adhesin-like
TERG_12603	hypothetical protein H100_00717	-0,71841579	Adhesin-like
TERG_03618	endochitinase 1	-0,72262078	Adhesin-like
TERG_12102	hypothetical protein TERG_08805	-0,73200332	Adhesin-like
TERG_12103	hypothetical protein H105_04372	-0,73624201	Adhesin-like
TERG_03596	hlh transcription	-0,73656273	Adhesin-like
TERG_12397	hypothetical protein TERG_12397	-0,73669009	Adhesin-like
TERG_04107	hypothetical protein TERG_04107	-0,74622806	Adhesin-like
TERG_00149	hypothetical protein TERG_00149	-0,75124686	Adhesin-like
TERG_07958	myosin class ii heavy chain	-0,75742588	Adhesin-like
TERG_06919	hypothetical protein TERG_06919	-0,76161724	Adhesin-like
TERG_05273	chromosome condensation protein	-0,76911811	Adhesin-like
TERG_12191	ras guanyl-nucleotide exchange factor	-0,76993997	Adhesin-like
TERG_12544	hypothetical protein H100_00997	-0,7704301	Adhesin-like
TERG_03217	hypothetical protein TERG_03217	-0,77583884	Adhesin-like
TERG_05732	serine-threonine rich	-0,77784735	Adhesin-like
TERG_08991	atp synthase regulation protein nca2	-0,77943161	Adhesin-like
TERG_11860	hypothetical protein H100_07201	-0,78154176	Adhesin-like
TERG_00926	exosome complex exonuclease rrp6	-0,78589411	Adhesin-like
TERG_05200	c6 finger domain	-0,78958087	Adhesin-like
TERG_05372	hypothetical protein	-0,79912339	Adhesin-like