

PESQUISA DOS INTEINS VMA E THRRS EM LEVEDURAS DO GÊNERO *Candida* ISOLADAS DE HEMOCULTURA.

TÂMARA HELOÍSA ROCHA PRANDINI

Dissertação apresentada ao Instituto de Biociências,
Câmpus de Botucatu, UNESP, para obtenção do título de
Mestre no Programa de Pós-Graduação em Biologia Geral
e Aplicada, Área de concentração: Biologia de Parasitas
e Micro-organismos.

Orientador: Prof. Dr. Eduardo Bagagli

UNIVERSIDADE ESTADUAL PAULISTA

“Julio de Mesquita Filho”

INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

PESQUISA DOS INTEINS VMA E THRRS EM LEVEDURAS DO
GÊNERO *Candida* ISOLADAS DE HEMOCULTURA

TÂMARA HELOÍSA ROCHA PRANDINI

RAQUEL CORDEIRO THEODORO

EDUARDO BAGAGLI

Dissertação apresentada ao Instituto de Biociências,
Campus de Botucatu, UNESP, para obtenção do título
de Mestre no Programa de Pós-Graduação em Biologia
Geral e Aplicada, Área de concentração: Biologia de
Parasitas e Micro-organismos.

Orientador: Prof. Dr. Eduardo Bagagli

Agradecimentos

Ao Prof. Dr. Eduardo Bagagli, pela oportunidade oferecida, pela orientação deste estudo e, principalmente por enriquecer meu conhecimento com sua sabedoria e paciência.

À Raquel Cordeiro Theodoro, pela orientação, por sempre estar disposta a me ajudar em qualquer situação, pela amizade e pelo exemplo de profissional.

Aos amigos de bancada, Sandra, Assís, Gabriel, Juliana Rizzo, Juliana Giacobino, Ariane, Tarsila, Thales, Mariana, Raquel, pela colaboração, amizade, e por fazer do ambiente de trabalho um local agradável e extremamente divertido.

Aos funcionários Sônia, Nice e Lula, pela atenção e carinho.

À professora Terue Sadatsune, pela disposição e suas valiosas sugestões.

À toda minha família, em especial à minha mãe, Cesarina Prandini, pela educação, amor e incentivo, à meu pai, Luiz Prandini, que mesmo de outro plano, ilumina meu caminho, à minha irmã, Sílvia, por todo apoio e amor, sendo minha segunda mãe.

À meu namorado, Ricardo Ferraz, pelo apoio e amor a mim oferecidos durante a elaboração deste trabalho.

Aos amigos que acompanharam de longe, sempre me ajudando e incentivando.

À Deus por me amparar nos momentos difíceis, me dar força para superar as dificuldades e me suprir em todas as minhas necessidades.

Às agências financiadoras

À FAPESP, pela concessão da bolsa de mestrado e pelas correções e sugestões do assessor.

À CAPES, pela concessão da bolsa de mestrado durante seis meses.

"Tu te tornas eternamente responsável por aquilo que cativas"
(O Pequeno Príncipe - Antonie de Saint-Exupéry)

SUMÁRIO

Revisão de Literatura – *Intein*: Um elemento genético parasita

1. Resumo.....	1
2. Introdução.....	2
3. Domínio <i>Splicing</i>	5
4. Domínio <i>Homing</i> Endonuclease e o ciclo homing.....	8
5. Inteins em leveduras do gênero <i>Candida</i>	15
6. Uso de <i>inteins</i> em análises filogenéticas.....	18
7. Potencial alvo terapêutico.....	20
8. Justificativa.....	23
9. Objetivos.....	24
10. Referências bibliográficas.....	25

Artigo submetido: Phylogenetic analysis of VMA and ThrRS inteins in <i>Candida</i> species.....	31
--	----

*Revisão de Literatura: Inteín: Um
elemento genético parasita*

1. Resumo

Leveduras do gênero *Candida* são consideradas os principais agentes de micoses sistêmicas oportunistas, que aumentaram nas duas últimas décadas, devido ao número cada vez maior de hospedeiros com algum grau de comprometimento do sistema imune. Diferentes espécies podem apresentar diferentes perfis clínico-epidemiológicos, assim, um maior conhecimento do perfil molecular e uma correta identificação dos isolados faz-se necessária. Com o sequenciamento de genomas completos de algumas espécies de *Candida*, tornou-se possível a descoberta e melhor compreensão de algumas regiões gênicas peculiares com importantes implicações evolutivas, biotecnológicas e de desenvolvimento de novas opções terapêuticas. Dentre estas regiões gênicas, destacam-se alguns elementos genéticos parasitas, dentre eles os *inteins* ou *internal proteins*, descobertos recentemente e ainda pouco estudados em populações de *Candida*. Este trabalho avaliou a ocorrência dos *inteins* nos genes VMA (ATPase vacuolar) e ThrRS (treonil RNAt-sintetase) na população de *Candida* de interesse médico (*C. albicans*, *C. parapsilosis*, *C. tropicalis* e *C. glabrata*), possibilitando um maior conhecimento dos perfis moleculares e filogenéticos dos isolados causadores de infecção em nossa região. Foi comprovado também que ambos os *inteins* aqui pesquisados apresentam um potencial como marcador molecular para a identificação destas espécies, bem como as espécies crípticas do complexo *psilosis*.

2. Introdução

Descritos há pouco mais de vinte anos, os *inteins* ou *internal proteins*, também já chamados de *protein introns* são conhecidos como elementos genéticos parasitas, pois utilizam o maquinário celular de replicação para se manter e, em alguns casos, se propagar na população. São sequências de nucleotídeos que se encontram dentro de genes extremamente conservados e importantes, porém, diferente dos introns, são transcritos e traduzidos juntamente com o fragmento hospedeiro para então serem excisados autocataliticamente, permitindo a funcionalidade normal da proteína hospedeira ou *extein* (Gogarten et al., 2002).

Estes elementos apresentam importantes implicações evolutivas e possibilitam um maior conhecimento de perfis moleculares e filogenéticos, podendo contribuir para a eficácia dos tratamentos já existentes para algumas infecções por leveduras do gênero *Candida* causadoras de doenças, bem como na busca de novas alternativas terapêuticas. Outro fato importante é que o *intein* VMA, o mais estudado em fungos, está presente em grande parte do gênero *Saccharomyces*, sendo a *Saccharomyces cerevisiae* um organismo modelo para pesquisa e importante na indústria e na biotecnologia.

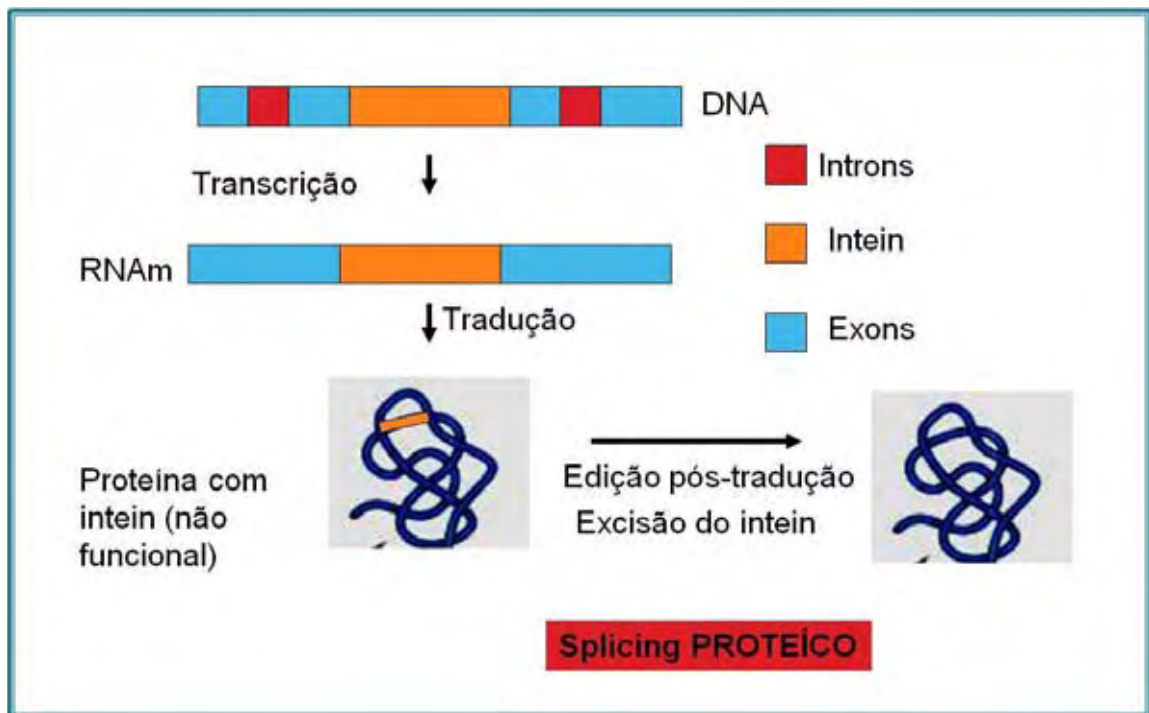


Figura 1: Desenho esquemático da excisão do *intein* da proteína hospedeira após o processo da tradução (Theodoro et al., 2009).

Existem três tipos de *inteins* na natureza: os *mini-inteins*, os *inteins* bi-funcionais (também conhecidos como *large-inteins* ou *full-length inteins*), e os *split inteins* (Figura 2). Os *mini-inteins* são constituídos de um domínio responsável pelo *splicing* (excisão) protéico autocatalítico (domínio Spl) constituídos pelas porções N e C terminais. Os *inteins* bi-funcionais apresentam o domínio *splicing*, sendo este separado por um domínio central codificador de uma *homing* endonuclease (domínio HE) que confere ao *intein* mobilidade, resultando na ocupação de alelos vazios e duplicação do elemento genético parasita (Liu, 2000). Os *split-inteins* constituem *mini-inteins* cujas partes N e C terminais estão separadas, flanqueando genes distintos. Após a tradução, as porções N e C terminais se unem, sofrem o *splicing* e ligam seus *exteins* em uma reação de *trans-splicing* protéico (Wu et al., 1998) (Figura 2).

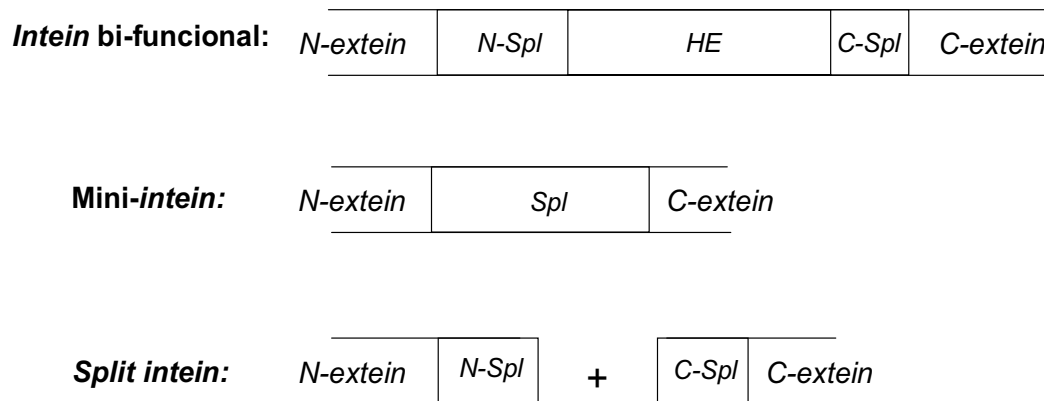


Figura 2: Desenho esquemático dos três tipos de *inteins* encontrados na natureza.

Segundo a base de dados disponível na internet, o *InBase* (base de dados de *inteins*) (<http://www.neb.com/neb/inteins.html>), a ocorrência deste elemento genético é observada nos três grandes domínios de vida: Eucariotos, Bactéria e Archaea (Perler, 2000). Geralmente os *inteins* são encontrados em sítios conservados de genes constitutivos e importantes para a sobrevivência ou reprodução do organismo hospedeiro, tais como genes codificadores de enzimas metabólicas, DNA e RNA polimerases, proteases e outros. Isso acontece, pois a ocorrência dos *inteins* preferencialmente em regiões conservadas diminui a chance de eliminação destes elementos, uma vez que a remoção do *intein* deve ser precisa a ponto de não alterar a função da proteína hospedeira (Liu, 2000). Quando *inteins* ocupam exatamente o mesmo sítio em um mesmo gene eles são referidos como *inteins* alélicos, mesmo que presentes em organismos de diferentes espécies (Butler et al., 2006). Mais de 130 *inteins* já foram descritos e estão distribuídos dentre aproximadamente 34 proteínas hospedeiras com funções distintas (Perler, 2000). Alguns *inteins* estão presentes em organismos filogeneticamente distantes, como é o caso do *intein* do gene DnaB que ocorre tanto em uma cianobactéria fotossintética como no termófilo marinho heterotrófico *Rhodothermus marinus* (Liu & Hu, 1997). Devido aos inúmeros projetos genomas, novos *inteins* vêm sendo

continuamente descobertos, sendo de extrema importância o uso de uma nomenclatura prática para designá-los. Assim, os *inteins* são nomeados segundo o organismo e a proteína hospedeira, por exemplo, *Ctr* ThrRS representa o *intein* que ocorre na proteína ThrRS da espécie *Candida tropicalis*.

Os HEs também possuem uma nomenclatura apropriada, sempre começando com a sigla I (*Insertion*) para HE que ocorrem em introns e PI (*Protein Insertion*) para relacioná-las ao *intein*, seguidas pelas três letras que designam a espécie e um numeral romano que especifica diferentes endonucleases presente no organismo (Perler, 2000). Assim, por exemplo, PI-*Ctr*I designa o HE presente no *intein* ThrRS de *C. tropicalis*.

3. Domínio *Splicing*

O termo *Splicing* protéico está associado aos *inteins* desde 1994. O domínio Spl é dividido em blocos de aminoácidos conservados: A e B para a porção N-Spl (a letra N refere-se ao amino-terminal) e F e G, da parte C-Spl, na região carboxi-terminal. O bloco A inicia a porção N-terminal, geralmente com um resíduo de cisteína ou serina, o B está a 60-90 aminoácidos do N-terminal e contém geralmente um *motif* Thr-x-x-His. Esta histidina é o resíduo mais conservado nos *inteins*. O bloco F precede o G, o qual contém o dipeptídeo C-terminal His-Asn e o nucleófilo +1 do C-*extein*. A serina, cisteína e treonina no C-*extein* são aminoácidos envolvidos na reação de *splicing* protéico do *intein*, juntamente com a cisteína ou serina do bloco A, asparagina da porção C-terminal do *intein* e as duas histidinas conservadas (uma no N-Spl e outra no C-Spl)-terminal (Figura 3). De forma geral, observa-se um padrão de *motifs* de aminoácidos com funções similares (Perler, 2008).



Figura 3: Desenho esquemático do domínio *splicing*, modificado do *Inbase* (Perler, 2002). Em letra maiúscula estão os aminoácidos conservados na maioria dos *inteins*, em minúscula estão aminoácidos presentes em *inteins* polimórficos que devem possuir mecanismos alternativos de *splicing* protéico.

Basicamente o *splicing* protéico envolve quatro passos (Gogarten et al., 2002) (Figura 4):

1) a junção N-terminal do domínio Spl com o N-terminal do *extein* é ativado por uma troca N-O ou N-S resultando em um intermediário ester ou tioester. Consequentemente a porção N-*extein* se liga ao oxigênio da serina ou ao enxofre da cisteína do N-Spl.

2) Transesterificação: a clivagem do ester na junção N-terminal ocorre através de um “ataque” do resíduo nucleofílico localizado na junção C-terminal resultando em uma proteína intermediária ramificada.

3) A clivagem procede através da circularização da asparagina da porção C-terminal causando a excisão do *intein* e *splicing* dos dois *exteins* por uma ligação ester. Muitos *inteins* têm uma glutamina ao invés de asparagina no C-Spl, sugerindo que a clivagem pode ocorrer via um intermediário aminoglutarimida e não aminosuccinimida.

4) Ligação peptídica espontânea ligando as porções N e C terminais do *extein*, constituindo uma proteína funcional.

Observa-se que o domínio Spl sofre uma seleção conservadora, ou seja, a razão entre substituições sinônimas e não-sinônimas (dS/dN) é maior que 1, uma vez que mutações não sinônimas podem prejudicar a função *splicing* do *intein* que, mantendo-se na proteína hospedeira, inibe o funcionamento normal da mesma podendo levar o organismo hospedeiro à morte ou inibindo seu crescimento ou reprodução (Butler et al., 2006).

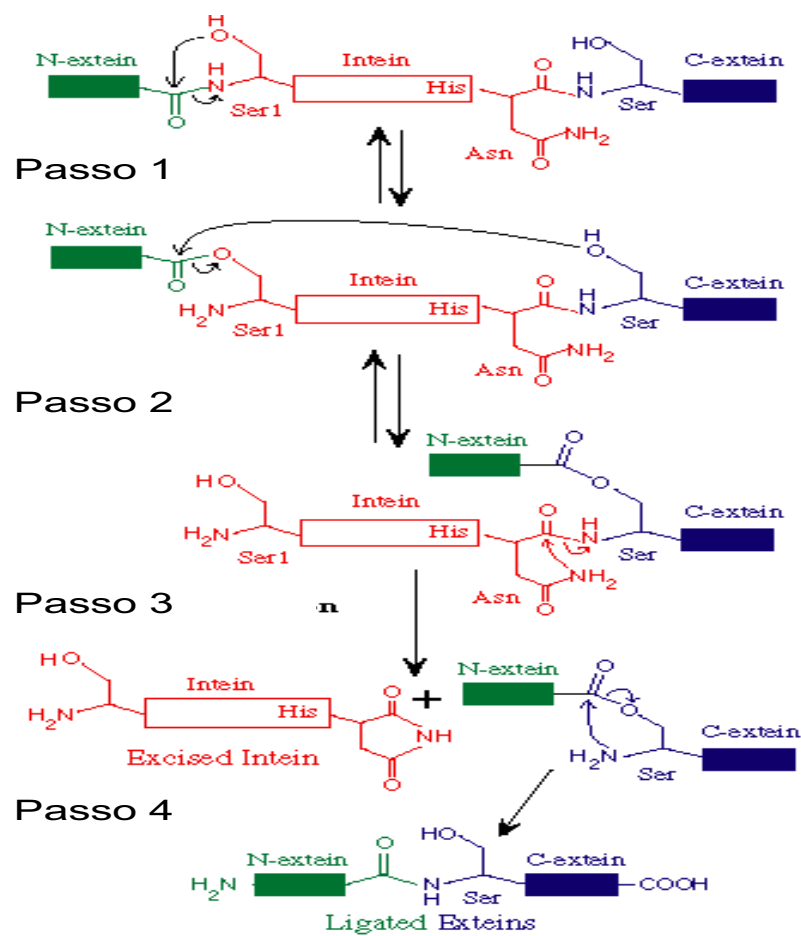


Figura 4: Passos envolvidos no *splicing* proteico do *intein*, segundo *Inbase* (Perler 2002).

4. Domínio *Homing* Endonuclease e o ciclo *homing*

Alguns *inteins* podem conter o domínio *Homing* endonuclease adicionados ao domínio Spl, dando ao *intein* mobilidade, resultando na ocupação de alelos vazios e duplicação do elemento genético parasita (Liu, 2000). Ou seja, o domínio HE é capaz de reconhecer um sítio específico na fita de DNA correspondente a um alelo sem o *intein*, este sítio é clivado e o alelo que contém o *intein* serve de molde para o reparo, transformando o organismo heterozigoto para o *intein* em homozigoto, possibilitando uma herança Super-Mendeliana (Burt & Koufopanou, 2004). No caso do *intein* VMA de *S. cerevisiae*, de um cruzamento $HE^+ \times HE^-$, ao invés de 50% da F1 ser HE^+ , como esperado em uma herança Mendeliana clássica, esta porcentagem chega aos 70, 90 e até 100% de HE^+ (Burt & Koufopanou, 2004), sendo, portanto notável a rapidez com a qual os “large-*inteins*” fixam-se nas populações.

Em geral o sítio de reconhecimento da *homing* endonuclease é específico e único, porém, Posey et al (2004) ao estudarem a atividade de endonucleases provenientes do *intein* presente no gene VMA (codificador da ATPase vacuolar) de 12 espécies de leveduras do gênero *Saccharomyces*, observaram que somente três endonucleases eram ativas e destas, a PI-ScaI (proveniente da espécie *S. cariocanus*) é mais eficaz no corte do sítio de reconhecimento do VMA de *S. cerevisiae* do que de seu próprio VMA. Os autores reforçaram, com estes dados, a adaptação do *intein* VMA para a transferência horizontal.

Estas enzimas raras estão presentes nos três grandes domínios de vida, em diferentes compartimentos celulares, e podem ocorrer de forma livre no genoma, porém a maioria delas está presente em introns ou *inteins*. Endonucleases presentes em introns do grupo I e em *inteins* criam uma quebra dupla na fita de DNA (*double-strand break*) induzindo um processo de recombinação dependente de DNA no qual o alelo contendo o intron ou *intein* serve de molde para o reparo do alelo vazio (Lambowitz & Zimmerly, 2004).

As *homing* endonucleases são classificadas em quatro famílias distintas, sendo caracterizadas pelos seus *motifs* conservados: LAGLIDADG, GIY-YIG, His-Cys box e HNH. (Belfort et al., 2008). Porém as endonucleases da família LAGLIDADG são as mais comuns e estão divididas nos blocos C, D, E e H, geralmente conservados nos *inteins* bi-funcionais (Liu, 2000) (Figura 3). Seus hospedeiros incluem genomas de plantas, cloroplastos de algas, fungos, mitocôndrias de protozoários, bactérias e *Archaea* (Dalgaard et al., 1997). Uma razão para a ampla distribuição filogenética destas endonucleases é a sua capacidade de invadir diferentes tipos de sequências intercalantes, como introns do grupo I, introns de *Archaea* e *inteins* (Belfort & Roberts, 1997). Em geral, o sítio de reconhecimento destas endonucleases varia de 18 a 22 pares de base. Elas clivam as duas fitas do DNA gerando pontas coesivas de quatro nucleotídeos na porção 3'. Como todas as nucleases, as endonucleases da família LAGLIDADG necessitam de cátions divalentes para sua atividade (Belfort et al., 2008). Estudos estruturais também demonstraram que existem dois resíduos de aspartatos considerados essenciais para o funcionamento das *homing* endonucleases da família LAGLIDADG, correspondentes às posições Asp218 e Asp326 do *intein* *Sce* VMA. Estes aspartatos parecem estar diretamente envolvidos na quebra da dupla fita (Gimble & Stephens, 1995; Posey et al., 2004; Koufopanou & Burt, 2005).

Inteins são chamados de elementos genéticos parasitas, pois, ao que tudo indica, não trazem nenhum aumento do valor adaptativo aos organismos que o possuem e sua fixação na população não está relacionada a um aumento do *fitness* do organismo, mas sim ao processo de *homing* e de transferência horizontal. Um gene egoísta é aquele que aumenta a sua chance de perpetuação em consequência do aumento do valor adaptativo do seu hospedeiro e, nesse caso isso não ocorre. Desta forma podemos considerar os *inteins* como genes “mais egoístas” que os demais, pois se proliferam na população sem trazer benefício algum e, por não diminuir o *fitness* do organismo hospedeiro, já que se retiram da proteína hospedeira pelo

processo de *splicing* protéico, driblando a Seleção Natural atuante no fenótipo. (Gogarten et al., 2002, Theodoro et al., 2009).

Por ocorrerem em regiões conservadas estas proteínas hospedeiras possuem função vital na célula, sendo que qualquer mudança, como uma substituição ou uma deleção/inserção poderia inviabilizar sua função. Em uma perspectiva evolutiva é esperado o questionamento a respeito das forças seletivas que atuaram e continuam atuando para o atual cenário da distribuição dos *inteins* em suas proteínas hospedeiras. Muito já foi especulado a este respeito (Liu, 2000; Pietrokovski, 2001).

Uma ideia é a de que os *inteins*, nestas regiões, foram selecionados ao longo do tempo, pois como a sequência hospedeira é altamente conservada eles não poderiam ser facilmente deletados, pois como já mencionado qualquer mudança nas sequências hospedeiras sofreria uma seleção negativa, não se disseminando na população, já que se tratam de proteínas vitais. Outro ponto bastante relevante é o fato de que a função *homing* dos *inteins* bi-funcionais funciona por reconhecimento de sítio específico para a ação da endonuclease e somente regiões conservadas manteriam intactos estes sítios ao longo das gerações. Sequências de maior taxa de mutação teriam maiores chances de se tornarem “imunes” aos *inteins* (Perler et al., 1997; Gogarten et al., 2002; Theodoro et al., 2009).

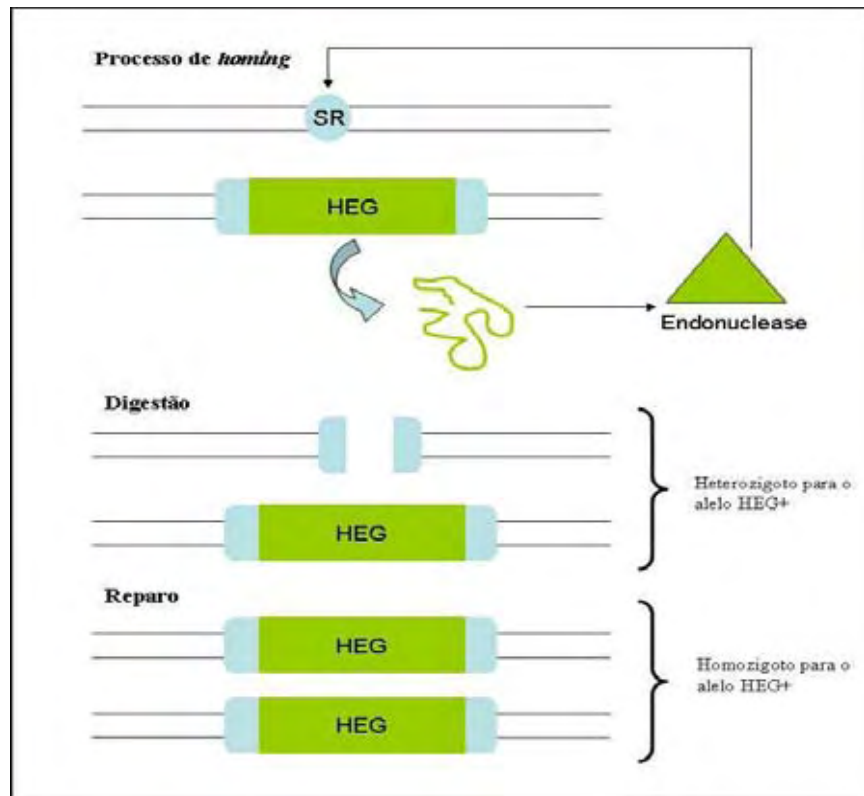


Figura 5: Processo de *homing*. SR: sítio de reconhecimento específico da endonuclease, HEG: *homing endonuclease gene*.

Alguns *inteins* se encontram dispersos de forma horizontal (ou seja, entre espécies distintas ao invés de se propagar verticalmente pela reprodução sexuada e herança dos caracteres) entre as diferentes espécies e entre genes hospedeiros das mais variadas funções, garantindo uma ampla distribuição filogenética (Perler, 2000). Os *inteins* hoje conhecidos estão distribuídos em aproximadamente 30 proteínas hospedeiras de funções distintas, muitos se repetem em organismos diferentes e, de certa forma, distantes, filogeneticamente (Liu, 2000). Evidências de transferência horizontal foram obtidas especificamente a partir de alguns *inteins*. Por exemplo, o *intein* DnaB pode ser encontrado em duas espécies remotamente relacionadas: uma cianobactéria fotossintética e um termófilo heterotrófico marinho (*Rhodothermus marinus*) (Liu & Hu, 1997). Quanto à transferência vertical, alelos sem e alelos com o *intein* se encontram na reprodução sexuada, o que propicia o processo de

homing. Isto já foi constatado experimentalmente pela re-introdução de um alelo livre do *intein* VMA em *Sacharomyces cerevisiae*. Após as meioses todos os alelos VMA continham o *intein* (Gimble & Thorner, 1992).

A frequência dos *inteins* HE, parece aumentar e diminuir de forma cíclica na população. A partir do momento que o *intein* ocupa um “alelo vazio” ele se propaga rapidamente na população podendo ser fixado na mesma. Porém, uma vez que este alelo está fixado, não existe mais uma seleção para o funcionamento da endonuclease, este elemento pode então se degenerar na população por substituições e deleções que podem se fixar na população por simples deriva genética, deixando assim novos sítios vazios que voltarão a ser ocupados, restabelecendo o ciclo de invasão, perda e re-invasão (Figura 6) (Burt & Koufopanou, 2004).

Entretanto, uma recente análise de *inteins* bi-funcionais sugerem que o domínio HE pode, em alguns casos, persistir ao longo do tempo evolutivo nas populações devido a várias razões, dentre as quais estariam estruturas populacionais complexas que previnem a fixação do HE, um balanço entre a dispersão do HE e uma possível diminuição do *fitness* do organismo hospedeiro.

Pode ocorrer até mesmo a manutenção desses elementos devido à aquisição de novas funções que acabam aumentando o *fitness* do organismo (Gogarten & Hilario, 2006). Aliás, acredita-se que alguns *inteins* tenham evoluído de uma condição parasítica para uma condição funcional, sendo positivamente selecionados devido a um eventual benefício trazido pelo HE.

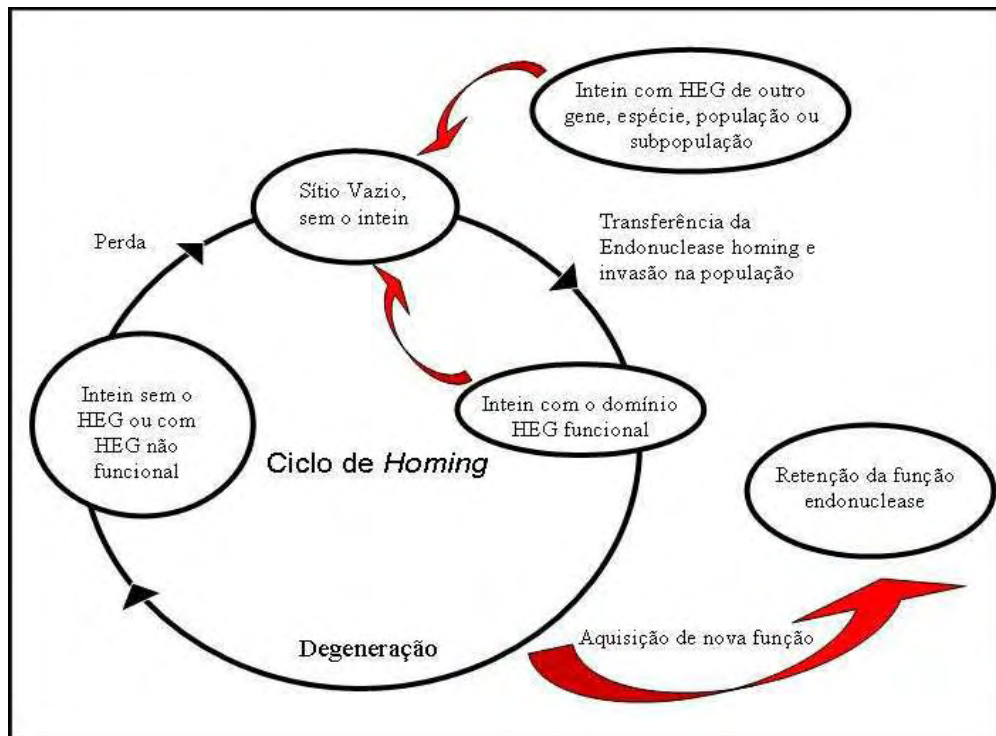


Figura 6: Esquema de invasão, fixação, degeneração e re-invasão do *intein* na população.

Um exemplo disso é a endonuclease HO de leveduras, responsável por uma conversão de gene que resulta em uma troca do *mating type* (tipo sexual) de leveduras (Liu, 2000). Esta endonuclease possui grande similaridade com o *intein* VMA de *S. cerevisiae*, incluindo regiões conservadas relacionadas à atividade de *splicing*, embora a HO não possua de fato a capacidade de se auto-remover. A presença de um domínio Spl inativo é visto como um forte indicativo de que esta endonuclease responsável pela troca de *mating type* tenha evoluído a partir de um ancestral *intein* (Gogarten et al., 2002).

A endonuclease HO inicia, durante a meiose, um evento de conversão gênica do *mating type*, catalisando a quebra da dupla fita de DNA no locus MAT do cromossomo da levedura, este locus tem aproximadamente 2400 pares de bases. As sequências que de fato definem o *mating type* Ya e Yα possuem cerca de 650 e 750 pares de base, respectivamente e são flanqueadas pelas sequências W, X, Z1 e Z2. Quando ocorre quebra na fita de DNA, pela

endonuclease HO, o maquinário de reparo da dupla fita usa como sequência molde ou “doadora” as sequências *HML α* e *HMR α* . Estas sequências não são expressas, elas estão mantidas silenciadas, em um estado heterocromático (lembrando que, em geral, sequências presentes na heterocromatina se encontram mais condensadas, não estando disponíveis para transcrição) pela ação de histonas e proteínas acessórias recrutadas por sequências regulatórias silenciadoras (E e I) próximas às sequências doadoras. A clivagem é feita no sítio Z1. Os loci doadores não podem ser clivados. A preferência por sequência doadora é governada pela sequência RE que controla a eficiência de recombinação ao longo de todo braço esquerdo do cromossomo III de *S. cerevisiae* (Figura 7) (Haber & Wolfe, 2005).

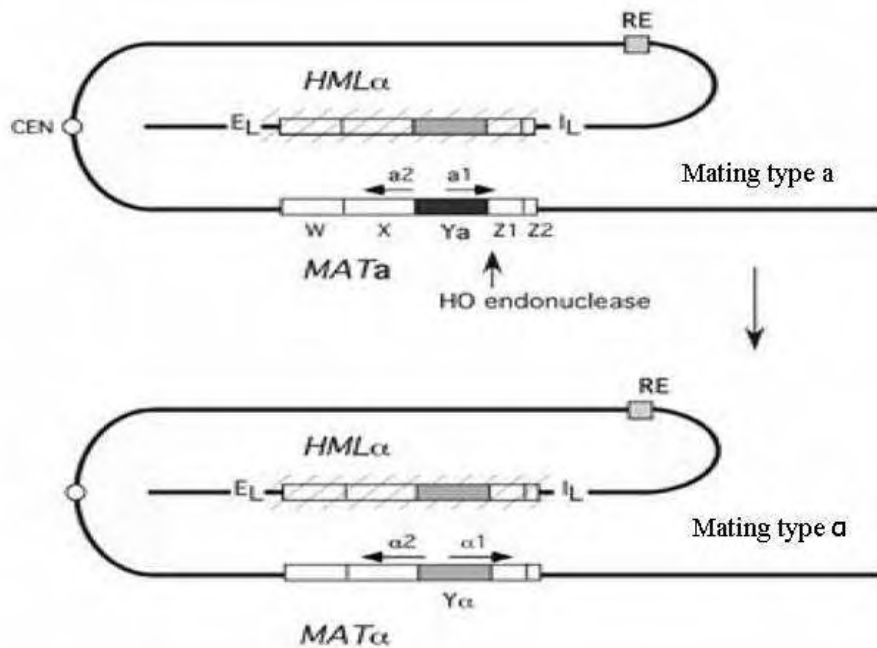


Figura 7: Troca do mating type em *S. cerevisiae*, catalisada pela endonuclease HO.

5. *Inteins* em leveduras do gênero *Candida*

As leveduras deste gênero pertencem a ordem Saccharomycetales (filo Ascomycota: subfilo Saccharomycotina: classe Saccharomycetes) que corresponde a um grupo monofilético, no qual aproximadamente 1000 espécies foram descritas. As leveduras desta ordem podem viver associadas a plantas e animais, e na maioria das vezes se reproduzem assexuadamente, por brotamento. São de extrema importância na indústria e em processos biotecnológicos, atuando no processo fermentativo de produção de álcool e na panificação (Suh et al., 2006).

A espécie *Saccharomyces cerevisiae* é responsável pelos processos acima citados e são cultivadas em destilarias para a produção de etanol a partir do açúcar da cana e também na fabricação de pães. Assim, é de extrema importância para a produção de álcool (álcool combustível e bebidas alcoólicas) e outros produtos de grande interesse industrial. Além disso, é tida como organismo modelo para pesquisa, pois é de fácil manutenção em laboratório e seu conhecimento biológico é bem desenvolvido, pois seu genoma já foi sequenciado e encontra-se numa base de dados para espécies de *Saccharomyces* (*Saccharomyces* Genome Database, <http://www.yeastgenome.org/>; Suh et al., 2006). Outro aspecto interessante das leveduras Saccharomycetales é sua história evolutiva, na qual ocorreram dois eventos importantes, um deles foi a duplicação do genoma de espécies do clado WGD (*whole-genome duplication*) que ocorreu há mais de 100 milhões de anos, e o outro é que algumas espécies traduzem o códon CUG como serina ao invés de leucina, formando um outro ramo chamado de clado CTG (Wang, H. et al., 2009).

Segundo Perler (2002), o *intein* VMA, um dos mais estudados em fungos está presente na maioria das leveduras do gênero *Saccharomyces* (tais como: *Saccharomyces castellii*, *Saccharomyces cariocanus*, *Saccharomyces cerevisiae*, *Saccharomyces dairenensis*,

dentre outras), *Zygosaccharomyces* e em algumas espécies do gênero *Candida*. Este *intein* é bi-funcional, (seu domínio *homing* corresponde a família LAGLIDADG) porém, pode-se encontrar degenerado em algumas espécies. O *intein* VMA ocorre no gene da ATPase vacuolar, codificando uma proteína que transporta íons H⁺ para organelas como lisossomos e endossomos, acidificando o interior dos mesmos. Uma vez silenciado o gene VMA, foi demonstrado que *S. cerevisiae* pode sobreviver em pH 5.0, porém não em pH 7,5 (Yamashiro et al., 1990). Tanto espécies do gênero *Saccharomyces*, quanto *C. glabrata*, que são pertencentes ao clado WGD, possuem o *intein* VMA, porém não se sabe se isso está diretamente relacionado com o evento de duplicação do genoma destas espécies.

A ordem *Saccharomycetales* também é composta por fungos oportunistas, que podem causar infecção no homem e em animais, sendo as espécies do gênero *Candida*, que têm emergido como importantes agentes de infecções oportunistas prevalentes, podendo disseminar-se na corrente sanguínea, principalmente, em indivíduos com a imunidade comprometida, incluindo portadores do HIV, pacientes com câncer submetidos à quimioterapia, transplantados em terapia imunossupressora e pacientes com diabetes avançada (Richardson, 2005; Aperis et al., 2006).

Nas últimas décadas, a incidência das candidemias aumentou substancialmente em todo o mundo (Saha et al., 2008), e atualmente essas infecções têm emergido como os maiores responsáveis pela morbidade (taxa de portadores da doença em relação à população total estudada) e mortalidade em pacientes imunocomprometidos (Pfaller et al., 2000, 2008; Girao et al. 2008,) e pacientes submetidos a hemodiálise (Pyrgos et al., 2008).

Por estarem associadas a altas taxas de morbidade e mortalidade, as candidemias podem prolongar o tempo de internação desses pacientes, tornando-os mais expostos a novas infecções (Girao et al., 2008). Pacientes que necessitam de procedimentos hospitalares invasivos estão em constante risco para o desenvolvimento dessas infecções. As unidades de

terapia intensiva apresentam predileção para este tipo de infecção, devido à maior concentração de pacientes expostos aos fatores de risco, (Rotstein, 2008; Girao et al., 2008).

A maioria dessas infecções é adquirida por via endógena, precedida pela colonização e deslocamento do microrganismo através do trato gastrointestinal do hospedeiro, mas ocasionalmente podem ocorrer infecções exógenas (Cole et al., 1996; Nucci & Anaisse, 2001; Rotstein, 2008), através do contato das mãos de profissionais da saúde em pacientes portadores de cateter venoso central, implante de próteses contaminadas, bem como pela administração parenteral de soluções contaminadas (Pfaller et al., 1995; Wenzel, 1995).

Atualmente, várias espécies de *Candida* estão implicadas em micoses superficiais ou invasivas de seres humanos (Dignanni et al., 2003). As principais espécies de interesse clínico são: *C. albicans*, *C. parapsilosis* (complexo de três espécies crípticas (*C. parapsilosis*, *C. metapsilosis* e *C. ortopsilosis*)) *C. tropicalis*, *C. glabrata* e *C. guilliermondii* (Coleman et al., 1998).

Algumas espécies de *Candida* apresentam o *intein* VMA, como *C. glabrata* e *C. tropicalis*, sendo estes bi-funcionais, porém, o domínio HE (pertencente à família LAGLIDADG) encontra-se degenerado em ambas.

A espécie *C. tropicalis* carrega adicionalmente outro *intein*, não alélico, no gene threonil-tRNA sintetase (CtThrRS), também presente na espécie *C. parapsilosis* (CpaThrRS). O domínio HE parece estar altamente degenerado (provável fase de degeneração e perda da função *homing*) em *C. tropicalis* e não existir em *C. parapsilosis*, sendo este um mini-*intein* (CZ284364, InBase). Este *intein* ocorre no gene thrRS (threonil-RNAt-sintase), que codifica uma aminoacil-RNAt-sintetase, responsável por acoplar a treonina ao RNAt com o anticódon correspondente, sendo necessário para a síntese protéica e, portanto, para a sobrevivência do microrganismo (Alberts et al., 2002).

Segundo Perler (2002), a espécie *Debaryomyces hansenii*, (*C. famata* em sua fase anamórfica) apresenta o *intein* VMA, carregando adicionalmente outro *intein*, não alélico, no gene GLT1, responsável pela síntese do glutamato, sendo portando, assim como o theonil-RNAt-sintetase, necessário para a síntese protéica (Alberts et al., 2002). O *intein* GLT1 também está presente em *Pichia guilliermondi* (*C. guilliermondi* em sua fase anamórfica). Os *inteins* DhanGLT1 e PguGLT1 são bi-funcionais e pertencentes a família LAGLIDADG (Perler, 2002).

As sequências destes *inteins* estão depositadas no banco de dados *Inbase* (<http://www.neb.com/neb/inteins.html>) porém, estes elementos genéticos foram pouco estudados no que se refere a uma avaliação populacional mais ampla, uma vez que em alguns casos pode-se encontrar indivíduos da mesma espécie com e sem o *intein*.

6. Uso de *inteins* em análises filogenéticas

Nos últimos anos, técnicas de sequenciamento de DNA e de bioinformática vêm sendo aprimoradas e com isso, sequências gênicas e genomas completos são rapidamente gerados e depositados em bancos de dados como GenBank, Broad Institute, Sanger Institute e outros, facilitando assim a correta identificação de micro-organismos, possibilitando a realização de métodos de tipagem molecular como Pulse-field, RAPD, RFLP, entre outras. Mesmo nesta fase, pós-genômica, a definição e o reconhecimento de “espécie” são tidos como grandes desafios.

Inteins, em especial os bi-funcionais, podem constituir uma fonte promissora de informação filogenética, pois apresentam um maior polimorfismo de sequência no domínio

HE do que no domínio Spl devido a uma seleção mais “frouxa”, especialmente no caso do HE não estar mais ativa (Gogarten & Hilário, 2006).

Estudos filogenéticos usando *inteins* são facilitados devido à localização deste elemento parasita em genes altamente conservados, como por exemplo, o thrRS, possibilitando assim o desenho de *primers* para sua amplificação por PCR seguido de seu sequenciamento. Desta forma, Butler & Poulter (2005) distinguiram duas variedades de *Cryptococcus*: *neoformans* e *gattii*, além disso, o sequenciamento destes mini-*inteins* evidenciaram a divisão do *C. gattii* em quatro subtipos, sugerindo que o *C. gattii* poderia se tratar de um grupo em especiação. Estas leveduras encapsuladas são capazes de causar sérias infecções tanto em pacientes imuno comprometidos como também em imuno competentes. As variedades *grubii* (sorotipo A), *neoformans* (sorotipo D) e *gattii* (sorotipos B e C) foram inicialmente designadas segundo diferenças antigênicas do polissacarídeo capsular e isto foi posteriormente confirmado por análises moleculares (Ellis et al. 2000, Franzot et al. 1999, Meyer et al. 1999). A variedade *gattii* recebeu, recentemente, o status de espécie, como *Cryptococcus gattii* (Boekhout et al. 2001, Kwon-Chung et al. 2002).

Em uma abordagem semelhante, Theodoro et al (2008) realizaram a análise filogenética do *intein* PRP8 (presente em um gene codificador de uma proteína essencial para a formação do spliceossomo) em diferentes isolados de *P. brasiliensis*, pertencentes às quatro diferentes espécies: S1, PS2, PS3 e *P. lutzii*. Constataram assim uma diferença significativa entre o *P. lutzii* e as demais espécies, corroborando com as análises filogenéticas realizadas por Carrero et al (2008) e Teixeira et al (2009) e comprovando que este elemento parasita contém informação genética suficiente para separar as espécies de *P. brasiliensis*, sendo, portanto um marcador genético adicional para este patógeno. *P. brasiliensis* é o agente etiológico da paracoccidioidomicose (PCM), doença caracterizada pelo desenvolvimento de

lesões granulomatosas com frequente evolução crônica, envolvendo pulmão, pele, membranas mucosas e outros tecidos (Franco, 1987).

Apesar da ocorrência comum de transferência horizontal, alguns *inteins*, cuja distribuição parece indicar um sistema de herança vertical, são usados em análises filogenéticas. Portanto, os *inteins* podem, em alguns casos, ser de grande utilidade em análises filogenéticas que visam esclarecer a história evolutiva e relações de parentesco entre espécies ou variedades distintas. Nos casos acima descritos, tratam-se de fungos patogênicos, cujas medidas de tratamento podem depender significativamente de uma melhor definição sobre a evolução e taxonomia dos mesmos.

7. Potencial alvo terapêutico

Os principais antifúngicos disponíveis para o tratamento de micoses sistêmicas, como os derivados poliênicos e imidazólicos, podem causar graves efeitos colaterais ao hospedeiro, o que é compreensível pelo fato de que fungos são filogeneticamente mais próximos de animais do que de vegetais. O fato de não haver *inteins* no genoma humano os torna alvos terapêuticos potencialmente seguros.

Os *inteins* ThrRS, GLT1 e VMA das leveduras *Saccharomycetales* são, obviamente, um potencial alvo terapêutico já que se encontram em genes extremamente conservados dentre os eucariotos, sendo que os dois primeiros são essenciais para a sobrevivência da célula hospedeira, e o VMA extremamente importante, mas não vital, pois uma vez silenciado o gene VMA, foi demonstrado que *S. cerevisiae* pode sobreviver em pH 5.0, porém não em pH 7,5 (Yamashiro et al., 1990). Uma vez inibido o *splicing* do *intein*, a proteína resultante não seria mais funcional, interrompendo assim a sobrevivência e proliferação da célula fúngica. Como os sítios de inserção destes *inteins* são conservados, a descoberta de uma droga que

iniba o *splicing* protéico poderia ter um amplo espectro de ação (Liu & Yang, 2004). Sendo assim, o estudo do funcionamento do domínio Spl deste *intein* é de extrema importância.

Theodoro et al (2011) avaliaram a atividade *splicing* dos *inteins* Bde PRP8, Epa PRP8 e Pbr PRP8 (incluindo *inteins* das quatro espécies crípticas do complexo P. *brasiliensis*, S1, PS2, PS3 e P. *lutzi*) inserindo o *intein* PRP8 no plasmídeo de expressão pMST entre os sítios *Xho*I e *Age*I, sendo flanqueado por uma proteína ligadora de maltose (*maltose binding protein*, MBP) no N-terminal e por uma tireodoxina no C-terminal. Western blotting com anticorpo anti-tireodoxina identificou a proteína precursora e a excisada pelos seus tamanhos deduzidos. Foi, portanto, possível observar o *splicing* protéico em todos os isolados fúngicos.

A fim de facilitar a observação do *splicing*, muitos pesquisadores vêm propondo sistemas nos quais o funcionamento do *intein* é ligado a um fenótipo facilmente selecionável. Por exemplo, o Mtu RecA (*intein* RecA de *Mycobacterium tuberculosis*) foi inserido no plasmídeo pUC19 interrompendo o gene *lacza* de forma que, após a indução da expressão gênica, o *splicing* era avaliado pelo simples sistema de colônias brancas/ azuis (Davis et al., 1992). Em alguns casos esta avaliação pode ser feita segundo o crescimento do microrganismo. Dizemos que a seleção é positiva quando o crescimento é dependente do *splicing*. Assim, Cooper et al (1993) inseriram o *intein* Sce VMA no gene *vat2* o qual, uma vez silenciado, permite crescimento da levedura em pH 5,0, mas não em pH 7,5, assim, leveduras transformadas com o gene interrompido pelo *intein* eram semeadas e mantidas em pH 7,5, de forma que o crescimento positivo indicava ocorrência de *splicing*. Esta foi a primeira vez que o funcionamento de um *intein* foi observado, em um *extein* não nativo em levedura.

Derbyshire et al (1997) criaram o sistema *Thymidylate Synthase* (TS) no qual é possível se utilizar duas abordagens: a seleção positiva e a seleção negativa (crescimento

associado à ausência de *splicing*). A enzima TS, cujo co-fator é o tetrahydrofolato, é responsável pela conversão de dUMP em dTMP, este será posteriormente convertido em dTTP que será incorporado nas novas cadeias de DNA sintetizadas. Ao inserirem o *intein Mtu* RecA no geneT4 *td*, codificador da TS, os autores desenvolveram um sistema de seleção positiva, cultivando as bactérias transformantes em meio sem timina. Ou seja, somente as bactérias nas quais o *splicing* ocorria poderiam sintetizar timina (endógena) e replicar seu DNA. Para adaptar um sistema de seleção negativa, os autores fizeram uso do Trimetoprim que é um inibidor da dihydrofolato redutase, diminuindo assim a síntese de tetrahydrofolato, necessário para outras reações de metilação. As bactérias transformantes foram então semeadas em meio contendo Timina + Trimetoprim. A timina no meio tornava possível o crescimento mesmo em caso de ausência de *splicing*. Mas se o *splicing* ocorresse, a produção endógena de TS consumiria todo o tetrahydrofolato e, por causa do Trimethoprim, ele não era mais sintetizado, levando a célula à morte.

Estes e muitos outros sistemas usam o crescimento bacteriano ou fúngico para determinar a atividade *splicing* dos *inteins* e, embora todos sejam úteis para uma rápida seleção, uma possível desvantagem destes sistemas reside no fato de que substâncias ou drogas usadas para avaliar a inibição do *splicing* poderiam na verdade inibir ou atenuar o crescimento do microrganismo por razões outras, não envolvidas com o funcionamento do *intein*. A fim de se excluir esta possibilidade, Gangopadhyay et al (2003) criaram um sistema de *screening* reporter *in vitro*. Neste sistema o *Mtu* RecA foi inserido em uma GFP (*green fluorescent protein*) modificada. O acúmulo da proteína precursora, ou seja, com o *intein*, forma corpos de inclusão não solúveis em *E. coli*, estes são dissolvidos em solução de urea, purificados e restituídos em um precursor estável na presença de zinco, elemento químico conhecido por inibir o *splicing* (Nichols et al., 2003). O *splicing* pode ser induzido pela adição de EDTA. Este sistema permitiu a avaliação de alguns compostos químicos para a inibição do

splicing do *Mtu* RecA, sendo que o zinco e o cobre inibem o *splicing in vitro*, já a cisplatina, recentemente testada por Zhang et al (2010) inibe o *splicing* tanto *in vitro* quanto *in vivo*. Estes compostos se mostraram eficientes devido a sua ligação com o primeiro resíduo de cisteína do domínio Spl, indispensável ao *splicing* protéico (Paulus, 2007; Zhang et al., 2010).

8. Justificativa

Uma vez que as diversas espécies do gênero *Candida* tendem a apresentar diferentes perfis epidemiológicos, clínicos e de susceptibilidade a drogas, uma correta identificação dos isolados faz-se necessária. Tendo em vista a existência de uma importante coleção de amostras de leveduras do gênero *Candida* (*C. albicans*, *C. parapsilosis*, *C. tropicalis* e *C. glabrata*) obtidas de pacientes do HC/UNESP-Botucatu, a proposta deste trabalho foi caracterizar estes isolados quanto a presença de *inteins* nos genes VMA e ThrRS, bem como realizar análises filogenéticas destes *inteins* a fim de se avaliar o potencial dos mesmos como marcadores moleculares. Além disso, o fato destes elementos estarem presentes em genes codificadores de proteínas vitais para a célula fúngica, faz deles alvos terapêuticos para desenho de drogas inibidoras de seu *splicing* protéico, porém, até então não tinha sido realizado nenhum estudo populacional a respeito da presença desses elementos parasitas em um número significativo de isolados do gênero *Candida*. Deste modo, os dados aqui obtidos puderam contribuir para um melhor conhecimento da biologia destes elementos genéticos parasitas neste grupo de leveduras, seu possível uso como auxiliar na identificação e caracterização molecular, bem como servir de base para futuros estudos no que diz respeito ao potencial alvo terapêutico destes genes.

9. Objetivos

- Pesquisar os *inteins* VMA e ThrRS na população de *C. albicans*, *C. parapsilosis*, *C. tropicalis* e *C. glabrata*.
- Avaliar a possível presença destes dois *inteins* (VMA e ThrRS) em uma população de *C. albicans* para confirmar a ausência e/ou possível presença destes elementos nesta espécie;
- Amplificar, sequenciar e fazer análise filogenética do *intein* VMA nos isolados clínicos de *C. tropicalis* e *C. glabrata*;
- Amplificar, sequenciar e fazer análise filogenética do *intein* ThrRS nos isolados clínicos de *C. parapsilosis* e *C. tropicalis*;
- Avaliar o potencial destes elementos genéticos parasitas como marcadores moleculares para a separação das espécies do gênero *Candida*, incluindo as espécies crípticas de *C. parapsilosis*.

10. Referências bibliográficas

- ALBERTS, B. et al.. Molecular Biology of the Cell. 4th edition. New York: Garland Science Publishing. 1400p. 2002.
- APERIS, G., MYRIOUNIS N., SPANAKIS E.K., Mylonakis E. Developments in the treatment of candidiasis: more choices and new challenges. Expert Opin Investig Drugs. 15, 1319-36, 2006.
- BELFORT, M., DERBYSHIRE, V., STODDARD, B.L., WOOD, D.W. Homing Endonucleases and Inteins. 2ed Springer – Verlag Berlin Heidelberg, 2008.
- BELFORT, M., ROBERTS, R.J. Homing endonucleases – keeping the house in order. Nucleic. acids Res. 25, 3379-3388, 1997.
- BOEKHOUT, T., THEELEN, B., DIAZ, M., FELL, J.W., HOP, W.C., ABELN, E.C., DROMER, F., MEYER, W. Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. Microbiol. 147, 891-907, 2001.
- BURT, A., KOUFOPANOU, V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr. Opin. Gen. & Develop. 14, 609-615, 2004.
- BUTLER, M.L., POULTER, R.T.M. The PRP8 inteins in *Cryptococcus* are a source of phylogenetic and epidemiological information. Fungal Genet. Biol. 42, 452-463, 2005.
- BUTLER, M.L., GRAY, J., GOODWIN, T.J.D., POULTER, R.T.M. The distribution and evolutionary history of the PRP8 intein. BMC Evol. Biol. 6, 1-26, 2006.
- CARRERO, L.L., NIÑO-VEJA, G., TEIXEIRA, M.M., CARVALHO, M.J.A., SOARES, C.M.A., PEREIRA, M., JESUÍNO, R.S.A., MCEWEEN, J.G., MENDOZA, L., TAYLOR, J.W., FELIPE, M.S., SAN-BLAS, G. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. Fungal Genet. Biol. 45, 605-612, 2008.
- COLE, G.T., HALAWA, A.A., ANAISSE, E.J. The role of the gastrointestinal tract in hematogenous candidiasis: from the laboratory to the bedside. Clinical Infectious Diseases 22, 73-88, 1996.
- COLEMAN, D.C., RINALDI, M. G., HAYNES, K. A., REX, J. H., SUMMERBELL, R. C., ANAISSE, E. J., LI, A., SULLIVAN, D. J. Importance of *Candida* species other than *Candida albicans* as opportunistic pathogens. Med Mycol. 36, 156-65, 1998.

- COOPER, A.A., CHEN, Y.J., LINDORFER, M.A., STEVENS, T.H. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J.* 12, 2575-2583, 1993.
- DALGAARD, J.Z., KLAR, A.J., MOSER, M.J., HOLLEY, W.R., CHATTERJEE, A., MIAN, I.S. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of HNH family. *Nucleic acids Res.* 25, 4626-4638, 1997.
- DAVIS, E.O., JENNER, P.J., BROOKS, P.C., COLSTON, M.J., SEDGWICK, S.G. Protein splicing in the maturation of *M. tuberculosis* recA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell*, 71, 201-210, 1992.
- DERBYSHIRE, V., WOOD, D.W., WU, W., DANSEREAU, J., DALGAARD, J.Z., BELFORT, M. Genetic definition of a protein splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution. *Proc. Natl. Acad. Sci. USA.* 94, 11466-11471, 1997.
- DIGNANNI, M. C., SOLOMKIN, J. S., ANAISSIE, E. *Candida*. In ANAISSIE, E., MCGINNIS, M. R., PFALLER, M. A. (eds) *Medical Mycology*. 1, 195-239, 2003.
- ELLIS, D., MARRIOTT, D., HAJJEH, R.A., WARNOCK, D., MEYER, W., BARTON, R. Epidemiology, surveillance of fungal infections. *Med. Mycol.* 38, 173–182, 2000.
- FRANCO, M. Host-parasite relationships in paracoccidioidomycosis. *J. Med. Vet. Mycol.* 25, 5-18, 1987.
- FRANZOT, S.P., SALKIN, I.F., CASADEVALL, A. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. *J. Clin. Microbiol.* 37, 833–840, 1999.
- GANGOPADHYAY, J.P., JIANG, S.Q., PAULUS, H. An in vitro screening system for protein splicing inhibitors based on green fluorescent protein as an indicator. *Anal. Chem.* 75, 2456-2462, 2003.
- GIMBLE, F.S., STEPHENS, B.W. Substitutions in conserved dodecapeptide motifs that uncouple the DNA binding and DNA cleavage activities of PI-SceI endonuclease. *J. Biol. Chem.* 270, 5849-5856, 1995.
- GIMBLE, F.S., THORNER, J. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, 357, 301-306, 1992.

- GIRAO, E., Levin, A. S., BASSO, M., GOBARA, S., GOMES, L. B., MEDEIROS, E. A., COSTA, S. F. Seven-year trend analysis of nosocomial candidemia and antifungal (fluconazole and caspofungin) use in Intensive Care Units at a Brazilian University Hospital. *Med Mycol.* 10, 1-8, 2008.
- GOGARTEN, J.P., HILARIO, E. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* 6, 1-5, 2006.
- GOGARTEN, J. P., SENEJANI, A. G., ZHAXYBAYEVA, O., OLENDZENSKI, L., HILARIO, E. Inteins: Structure, Function and Evolution. *Ann. Rev. Microbiol.* 56, 263-287, 2002.
- HABER, J.E., WOLFE, K.H. Function and Evolution of HO and VDE Endonucleases in Fungi. *Nucleic Acids Mol. Biol.* 16, 161-175, 2005.
- KWON-CHUNG, K.J, BOEKHOUT, T., FELL, J.W., DIAZ, M. Proposal to conserve the name *Cryptococcus gattii* against *C. hondurianus* and *C. bacillisporus*, Basidiomycota, Hymenomycetes, Tremellomycetidae. *Taxon* 51, 804-806, 2002.
- LAMBOWITZ, A.M., ZIMMERLY, S. Mobile group II introns. *Annu. Rev. Genet.* 38, 1-35, 2004.
- LIU, X. Protein-Splicing Intein: Genetic Mobility, Origin, and Evolution. *Ann. Rev. Genet.* 34, 61-76, 2000.
- LIU, X.Q., HU, Z. A DnaB intein in *Rhodothermus marinus*: indication of recent intein homing across remotely related organisms. *Proc. Natl. Acad. Sci. USA.* 94, 7851-7856, 1997.
- LIU, X. Q., YANG, J. Prp8 intein in fungal pathogens: target for potential antifungal drugs. *FEBS Letters.* 572, 46–50, 2004.
- MEYER, W., MARSZEWSKA, K., AMIRMOSTOFIAN, M., IGREJA, R.P., HARDTKE, C., METHLING, K., VIVIANI, M.A., CHINDAMPORN, A., SUKROONGREUNG, S., JOHN, M.A., ELLIS, D.H., SORRELL, T.C. Molecular typing of global isolates of *Cryptococcus neoformans* var. *neoformans* by polymerase chain reaction Fingerprinting and randomly amplified polymorphic DNA-a pilot study to standardize techniques on which to base a detailed epidemiological survey. *Electrophoresis.* 20, 1790–1799, 1999.
- NUCCI M, ANAÏSSE, E. Revisiting the source of candidemia: skin or gut? *Clin. Infect. Dis.* 33,1959-1967, 2001.

- NICHOLS, N.M., BENNER, J.S., MARTIN, D.D., EVANS, T.C. JR. Zinc Ion Effects on Individual *Ssp* DnaE Intein Splicing Steps: Regulating Pathway Progression. *Biochemistry* 42, 5301-5311, 2003.
- PAULUS, H., Protein splicing inhibitors as a new class of antimycobacterial agents. *Drugs of the Future* 32, 973-984, 2007.
- PERLER, F. B. InBase, the intein database. *Nucleic Acids Res.*, v.22, p.344-345, 2002.
- PERLER, F.B. Inteins- a historical perspective. In: Belfort, M., Derbyshire, V., Stoddard, B.L., Wood, D.W. (eds). *Homing Endonucleases and Inteins*. 2ed Springer – Verlag Berlin Heidelberg, 2008.
- PERLER, F.B., Adam, E. Protein splicing and its applications. *Current Opinion in Biotechnology* 11, 377-383, 2000.
- PERLER, F.B., XU, M.Q., PAULUS, H. Protein splicing and autoproteolysis mechanisms. *Curr. Opin. Chem. Biol.* 1, 292-299, 1997.
- PFALLER, M. A., DIEKEMA, D. J., GIBBS, D. L., NEWELL, V. A., NG, K. P., COLOMBO, A., FINQUELIEVICH, J., BARNES, R., WADULA, J. and the Global Antifungal Surveillance Group. Geographic and Temporal Trends in Isolation and Antifungal Susceptibility of *Candida parapsilosis*: a Global Assessment from the ARTEMIS DISK Antifungal Surveillance Program, 2001 to 2005. *J. Clin. Microbiol.* 46, 842–849, 2008.
- PFALLER, M. A., JONES, R. N., DOERN, G. V., SADER, H. S., MESSER, S. A., HOUSTON, A., COFFMAN, S., HOLLIS, R. J. Bloodstream infections due to *Candida* species: SENTRY antimicrobial surveillance program in North America and Latin America, 1997-1998. *Antimicrob. Agents Chemother.* 44, 747-51, 2000.
- PFALLER, M. A., MESSER, S. A., HOLLIS, R. J. Variations in DNA subtype, antifungal susceptibility, and slime production among clinical isolates of *Candida parapsilosis*. *Diagn. Microbiol. Infect. Dis.* 21, 9-14, 1995.
- PIETROKOVSKI, S. Intein spread and extinction in evolution. *Trends Genet.* 17, 465-472, 2001
- POSEY, K.L., KOUFOPANOU, V., BURT, A., GIMBLE, F.S. Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. *Nucleic acids res.* 32, 3947-3956, 2004.
- PYRGOS, V., RATANAVANICH, K., DONEGAN, N., VEIS, J., WALSH, T. J., SHOHAM, S. *Candida* bloodstream infections in hemodialysis recipients. *Med Mycol.* 16, 1-5, 2008.

- RICHARDSON, M. D. Changing patterns and trends in systemic fungal infections. *J. Antimicrob. Chemother.* 56, 5-11, 2005.
- ROTSTEIN, C. Invasive Candidiasis in the ICU: Prophylaxis Versus Preemptive Treatment. *Curr. Infect. Dis.Rep.* 10, 454-8, 2008.
- SAHA, R., DAS DAS, S., KUMAR, A., KAUR, I. R. Pattern of Candida isolates in hospitalized children. *Indian J. Pediatr.* 75, 858-60, 2008.
- SUH, S.O., KURTZMAN, C.O., LACHANCE, M.A. Phylogenetics of Saccharomycetales, the ascomycete yeasts. *The Mycol. Soc. Of America. Mycologia.* 98, 1006-1017, 2006.
- TEIXEIRA, M.M., THEODORO, R.C., DE CARVALHO, M.J., FERNANDES, L., PAES, H.C., HAHN, R.C., MENDOZA, L., BAGAGLI, E., SAN-BLAS, G., FELIPE, M.S. Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides* genus. *Mol. Phylogenet. Evol.* 52, 273-283, 2009.
- THEODORO, R. C., BAGAGLI, E. Inteins in pathogenic fungi: a phylogenetic tool and perspectives for therapeutic applications. *Mem. I. Oswaldo Cruz* 104, 497-504, 2009.
- THEODORO, R. C., BAGAGLI, E., OLIVEIRA, C. Phylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complex. *Fungal Genet. Biol.* 45, 1284-1291, 2008.
- THEODORO, R. C., VOLKMANN, G., LIU, X.Q., BAGAGLI, E. PRP8 intein in Ajellomycetaceae family pathogens: sequence analysis, splicing evaluation and homing endonuclease activity. *Fungal.Genet. Biol.* 48, 80-91, 2011.
- YAMASHIRO, C.T., KANE, P. M., WOLCZYK, D. F. Role of vacuolar acidification in protein sorting and zymogen activation: a genetic analysis of the yeast vacuolar proton-translocating ATPase. *Mol. Cell. Biol.* 10, 3737-3749, 1990.
- WANG, H., XU, Z., GAO, L., HAO, L. A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol. Biol.* 10, 9-195, 2009.
- WENZEL, R.P. Nosocomial candidemia: risk factors and attributable mortality. *Clin. Infect. Dis.* 2, 1531-1534, 1995.
- WU, H., HU, Z., LIU, X. Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Nat. Acad. Sci. USA* 95, 9226-9231, 1998.
- ZHANG, L., ZHENG, Y., CALLAHAN, B., BELFORT, M., LIU, Y. Cisplatin inhibits protein splicing, suggesting inteins as targets in *Micobacteria*. *J. Biol. Chem.* 286, 1277-1282, 2010.

*Article: Phylogenetic analysis of
VMA and ThrRS inteins in Candida
species*

Title: Phylogenetic analysis of VMA and ThrRS inteins in *Candida* species.

Authors: Tâmara Heloísa Rocha Prandini¹, Raquel Cordeiro Theodoro¹, Ariane Cristina Mendes de Oliveira Bruder Nascimento¹, Eduardo Bagagli^{1*}

1- Dept of Microbiology and Immunology, Institute of Biosciences, Unesp

***Corresponding author:**

Dept of Microbiology and Immunology

Institute of Biosciences, Unesp

Distrito de Rubião Júnior, Botucatu, SP, Brazil

Zip Code: 18618000

Email address: bagagli@ibb.unesp.br

Abstract

This work aimed to evaluate the presence of VMA and ThrRS inteins, their sequence variability and related phylogenetic aspects in the most relevant clinical *Candida* species, including the cryptic species of the *psilosis* complex. Both inteins are absent in *C. albicans*. The full-length CgVMA and CtrVMA inteins are phylogenetically related, with the HE domains degenerated, in contrast to the same intein of *Saccharomyces cerevisiae*. The ThrRS intein proved to be more informative for distinguishing *C. tropicalis*, *C. parapsilosis* and its cryptic species, *C. orthopsilosis* and *C. metapsilosis*. While CpaThrRS and CmeThrRS are mini-inteins, CtrThrRS and CorThrRS are full-length inteins with the HE domain highly degenerated. The inteins, by having simultaneously the two well preserved splicing domains flanking the HE domains that may present higher variability, and being located in highly conserved genes, are particularly attractive for their potential in correctly identifying clinical yeasts.

Key-words: *Candida* spp, Intein, vacuolar ATPase (VMA), Theonil-tRNA synthetase (ThrRS), molecular identification.

Introduction

Candida species have emerged as important pathogens over the past few decades. These yeasts are important agents of opportunistic infections in humans, especially in individuals with compromised immunity, such as HIV+ and diabetic patients, or those that undergo chemotherapy, prolonged antibiotic therapy, transplant and other immunosuppressive therapies (Richardson et al., 2005; Aperis et al., 2006).

Although *C. albicans* remains the most frequent yeast species causing human infections, other *Candida* species such as *C. parapsilosis*, *C. tropicalis* and *C. glabrata* have been increasingly recognized by clinical sources (Pfaller et al., 1998; Colombo et al., 2007). The frequency of these non-albicans species depends on the patient population, the therapy, specifically the use of broad-spectrum antibiotics (Nucci and Colombo, 2007; Pfaller et al., 2008). For example, in nosocomial bloodstream infections from most tropical and/or Latin American countries, *C. parapsilosis* occurs at the same or even higher frequency than *C. albicans* (Colombo et al., 1999; Passos et al., 2007; Trofa et al., 2008; Bruder-Nascimento et al., 2010).

Phylogenetically, *Candida* species are members of the Order Saccharomycetales. While *C. albicans*, *C. parapsilosis* and *C. tropicalis* form a branch called CTG, that translate CUG codon into serine instead of leucine, *C. glabrata* belongs to the WGD clade, which underwent a whole-genome duplication more than 100 million years ago together with several species of *Saccharomyces* genus (Wang et al., 2009).

Several genetic regions have been used for phylogenetic and/or molecular identification studies of these species, such as the small-subunit (18S) rRNA gene (Mannarelli & Kurtzman, 1998), the 5.8S rRNA gene (Lott et al., 1993), the hypervariable D1/D2 region of the 26S rRNA gene (Linton et al., 2007), the internal transcribed spacer regions (ITS1 and

ITS2) of rRNA genes (Chen et al., 2000; Chen et al., 2001; Ciardo et al., 2006; Leaw et al., 2006), and mitochondrial cytochrome *b* gene sequences (Yokoyama et al., 2000).

Inteins, known as parasitic genetic elements, seem to constitute a promising source of phylogenetic information (Butler and Poulter, 2005). Inteins are intervening sequences that are transcribed and translated with flanking host protein sequences and then self-excised by protein splicing; and the flanking protein sequence (exteins) are joined by a normal peptide bond to form the functional protein (Perler et al., 1994; Xu et al., 1993; Chong et al., 1996).

There are three types of inteins, the bi-functional inteins, the mini-inteins, and the split inteins. The bi-functional inteins consist of a protein splicing domain that is divided into two halves by a centrally located homing endonuclease domain, which renders the intein a mobile genetic element, resulting in occupation of empty cognate alleles and duplication of the parasitic genetic element (Liu 2000). Mini-inteins lack the endonuclease domain, and thus have a continuous protein splicing domain. Split inteins constitute mini-inteins split into two separately encoded parts, which associate and ligate their exteins in a protein *trans*-splicing reaction (Wu et al., 1998).

The bi-functional inteins are expected to have more sequence variation in the endonuclease domain than in the splicing domain due to a more relaxed selection, mainly when the homing endonuclease is no longer active (Butler et al., 2006). Therefore, they constitute a promising source of phylogenetic information.

A survey of InBase shows that the inteins are between 134 and 608 amino acids long and occur in all three domains of life (eukaryotes, bacteria and archaea). They are usually found at conserved sites of housekeeping proteins that have vital functions in the cell, such as DNA and RNA polymerases, proteases and others (Liu, 2000). For this reason, any mutation of the splicing domain can be fatal to the cell.

The first evidence of inteins arose from structural studies and expression analyses of the vacuolar ATPase gene and its encoded protein (VMA), from *S. cerevisiae* (Hirata et al., 1990; Kane et al., 1990), which presents the most studied example of this phenomenon, the VMA intein, which also occurs in other Hemiascomycetes including *C. glabrata* and *C. tropicalis* (Kane et al., 1990; Koufopanou et al., 2002). All these VMA inteins are bi-functional, but the HE domain, whose endonuclease belongs to the LAGLIDADG family, appears to be degenerated in the *Candida* species (Poulter et al., 2007).

While *inteins* have not been observed in *C. albicans* until now, the species *C. tropicalis* and *C. parapsilosis* have another *intein*, in the gene threonil-tRNA synthetase (intein CtrThrRS and CpaThrRS, respectively). The HE domain appears to be highly degenerated and probably not functional in CtrThrRS and does not exist at all in CpaThrRS, which is a mini-intein (Poulter et al., 2007). On the other hand, the splicing domains of these inteins are well preserved and functional, as expected, since the host gene thrRS (threonil-tRNA synthetase), encodes an aminoacyl-tRNA synthetase, responsible for engaging the threonine tRNA with the corresponding anticodon, being essential for protein synthesis and thus for the survival of the microorganism (Alberts et al., 2002).

By simultaneously presenting the two well preserved splicing domains flanking the HE domains that may present higher variability, and on account of being located in highly conserved genes that facilitate the design of primers for its amplification, the inteins are particularly attractive for carrying out phylogenetic studies. Using this approach, while studying another fungal intein, the PRP8 intein, Butler and Poulter (2005) distinguished strains of the two varieties of *C. neoformans* (*neoformans* and *grubii*) from each other, and from strains of *C. gattii*, recently designated as a new species, whereas Theodoro et al. (2008) distinguished strains of the four species of *Paracoccidioides brasiliensis* (S1, PS2, PS3 and P. *lutzi*).

This work aimed to evaluate the presence of VMA and ThrRS inteins, their sequence variability and some phylogenetic aspects in the most important clinical group of *Candida* species, including the cryptic species of the *psilosis* complex.

Materials and methods

Origin of the isolates

Forty-seven *Candida* isolates, previously identified, were used: 30 *C. albicans* strains, 6 *C. tropicalis*, 3 *C. glabrata*, 6 *C. parapsilosis*, 1 *C. metapsilosis* and 1 *C. orthopsilosis*. All the yeast cultures were obtained from patients of the Hospital das Clínicas, UNESP, Botucatu, São Paulo state, Brazil. Species identification was based on the colony morphology on chromogenic agar (CHROmagar *Candida*, Difco), microscopy features on cornmeal agar slide culture, assimilation and fermentation tests, according to Kurtzman & Fell (1998), and also by molecular sequencing of the rDNA region ITS1-5.8S-ITS2, according to Chen et al., 2000 and Bruder-Nascimento et al., 2010. Also applied to the *psilosis* group (composed of *C. parapsilosis*, *C. orthopsilosis* and *C. metapsilosis*) was the SADH restriction protocol, as described by Tavanti et al., 2005 and Bruder-Nascimento et al., 2010.

Culture and DNA extraction

The isolates were cultured on Sabouraud Dextrose Agar slants at 35°C and DNA was extracted according to McCullough et al. (2000) by initial cell disruption with glass beads (425-600 µm, acid washed, Sigma, St. Louis, MO, USA) in a solution of 1 M sorbitol and 125

mM EDTA. The DNA was quantified by agarose gel electrophoresis and diluted in ultrapure water to 10ng/μl.

Amplification and sequencing of the VMA and ThrRS inteins

The primer design initially involved a preliminary search in the Broad Institute, Sanger and Inbase databases for the intein and host gene sequences (VMA and ThrRS). The presence or absence of the inteins was evaluated by using the primers indicated in Table 1, which were also used for sequencing.

The PCRs were performed in total volumes of 25 μL (2 μL DNA), 1X PCR buffer (200 mM Tris-HCl pH 8.4, 1.5 mM MgCl₂ 50 mM, 500 mM KCl), 1.5 mM MgCl₂, 0.2 mM dNTP, 0.4 mM of each primer, and 1 unit of Taq Polymerase (GE Healthcare). The thermal cycling conditions were: 94°C for 4 min followed by 40 cycles at 94°C for 1 min, 55°C for 1 min and 72°C for 2 min. The PCR products were identified by 1% agarose gel electrophoresis stained with ethidium bromide.

After the PCR reactions, 10 μL of each amplified product was purified enzymatically with 4μL of ExoSAP-IT (GE Healthcare) for 1 h at 37°C followed by 20 min at 80°C to inactivate the enzyme. The samples were subsequently submitted to sequencing reaction and capillary electrophoresis in the ABI3500 DNA Analyzer (Applied Biosystems) at the Laboratory of Molecular Diagnosis in the Department of Microbiology and Immunology, IBB-UNESP.

Phylogenetic analysis

The sequences obtained were aligned in Mega version 5.0, together with sequences from Inbase (<http://www.neb.com/neb/inteins.html>). (GenBank accession n° BAC66648.1, CZ284364, A46080, AAY89365, Q874G3).

Phylogenetic analysis by Neighbor Joining (NJ) and Maximum Likelihood (ML) were performed separately for each intein in the program Mega 5.0 (Tamura et al., 2011, Inpress). For the ThrRS intein sequences, only the splicing domain was used for phylogeny because of the existence of mini-inteins among the species studied while the intein *CreRPB2* (RNA polymerase II of the *Chlamydomonas reinhardtii*) was used as the outgroup.

The nucleotide sequences of the VMA and ThrRS inteins were aligned using the ClustalW algorithm and the best distance model, calculated by the program Mega 5.0 (Tamura et al., 2011), was the Tamura-3-parameters (Tamura, 1992), for both alignments.

The translated sequences of both inteins were compared with the VMA intein of *Saccharomyces cerevisiae* (GenBank accession number: Q874G3) to observe the presence or absence of the two essential aspartic acid residues (Asp-218 and Asp-326) which are involved in the activity of homing endonucleases (Posey et al., 2004; Koufopanou and Burt, 2005). For both NJ and ML analysis, pairwise deletion was used to treat gaps.

Table 1: Primers used for PCR and sequencing of VMA and ThrRS inteins in *Candida* species

Specie	Gene	Primer	Sequence (5'-3')	Fragment length
<i>C. glabrata</i>	VMA (intein)	P57 P58	GTGAAATCTGCTTTGTCAAG ACAATTTCAACAAATTGGATG	~ 430
<i>C. glabrata</i>	VMA (extein)	P81 P82	AGAAACTATCAGCAGATTAC TGTAGATAGAAGCTTCTCTG	~ 1.570
<i>C. tropicalis</i>	VMA (intein)	P96 P97	TGGTATTGAATTAGTTAAAGC GCAAATTCCTTGTCTAACACC	~ 870
<i>C. tropicalis</i>	VMA (extein)	P83 P84	CTGCTGATCATCCATTGTTG ACCAGTATAAATAGAAGCTTC	~ 1.730
<i>C. tropicalis</i>	ThrRS (intein)	P59 P60	TTGAAGCTTCTCAACTTGTC TGATCTAAGTTCATCTTTGG	~ 360
<i>C. tropicalis</i>	ThrRS (intein)	P61 P62	ATTACTCCAAACATGTACTCC GTTAAACCACTAAGAGCACC	~ 1.300
<i>C. parapsilosis</i>	ThrRS (intein)	P78 P79	GTCATCTTGTTGAAACGAC AAGAATTGAATACAGAAAGAG	~ 900
<i>C. albicans</i>	VMA (extein)	P100 P101	TCCATGATGCACACCTGG CCAACGTGATGAAGAATCG	~ 400 *
<i>C. albicans</i>	ThrRS (extein)	P98 P99	CCATGTGGTTACCACATG GTCCCAATTTGATCAGTGG	~ 350 *

* Fragments expected in the absence of the intein

Results

Amplification and sequencing of the VMA and ThrRS inteins

Using the primers indicated in Table 1, we amplified and sequenced the fragments containing the inteins VMA of *C. glabrata* and *C. tropicalis* (CglVMA and CtrVMA, respectively), the inteins ThrRS of *C. tropicalis* (CtrThrRS), *C. parapsilosis* (CpaThrRS), *C. orthopsilosis* (CorThrRS) and *C. metapsilosis* (CmeThrRS). The last two *inteins* (CorThrRS and CmeThrRS) were described in this study.

Concerning the 30 *C. albicans* isolates herein evaluated, only the expected fragments of 400 and 350 bp were observed in the VMA and ThrRS genes, respectively, showing that all of them do not present any of these inteins.

We obtained a good quality consensus sequence for three CglVMA inteins (*C. glabrata* isolates S-1, AC-33 and S-130), four CtrVMA inteins (*C. tropicalis* isolates S-35, S-36, S-39 and S-158), three CtrThrRS inteins (*C. tropicalis* isolates S-5, S-36 and S-48), for six CpaThrRS (*C. parapsilosis* isolates S-18, S-33, S-41, S-45, S-46 and S-50), as well as for CmeThrRS and CorThrR inteins of *C. metapsilosis*, isolate ATCC-96143) and *C. orthopsilosis*, isolate ATCC-96141, respectively.

VMA intein

The VMA inteins CglVMA and CtrVMA are bi-functional, which means they have both splicing and HE domains. The translated sequences of the VMA inteins herein evaluated were compared to the SceVMA (GenBank Accession No Q874G3) in order to observe the presence or absence of the two aspartic acid residues (Asp-218 and Asp-326) that are involved in the activity of the homing endonuclease (Posey et al., 2004; Koufopanou and Burt, 2005). In all the CglVMA inteins, the first and the second critical aspartates were changed to serine and glutamine, while in all the CtrVMA they were changed to isoleucine and an alanine, respectively (Figure 1). The splicing domain (Blocks A, B, F and G) are highly conserved, mainly among isolates from the same species, with only a few substitutions among the three species evaluated in the present study.

The genetic distance and the phylogenetic analysis of VMA inteins indicated that CglVMA and CtrVMA share more similarities with each other than they do with SceVMA (Table Supplementary Material 1, Figure 2). This observation contradicts the close

relationship between *C. glabrata* and *S. cerevisiae* previously documented (Suh et al., 2006; Wang et al., 2009)

ThrRS intein

CpaThrRS and CmeThrRS are mini-inteins, while CtrThrRS and CorThrRS are bi-functional. In all the CtrThrRS inteins, the first and the second critical aspartate were replaced with asparagine and serine, respectively; while CorThrRS has the first aspartate changed to an arginine, and the second was deleted, when compared to the SceVMA (GenBank Accession No Q874G3).

The splicing domain is highly conserved in CtrThrRS, CpaThrRS and CmeThrRS inteins (Figure 3). The full-length intein described in *C. orthopsilosis* (CorThrRS) has a significant sequence polymorphism in splicing and especially in HE domain, than the other ThrRS inteins. It is also the longest ThrRS intein (the PCR product has ~ 1600 bp).

CpaThrRS has ~800bp, while the PCR for the CmeThrRS presented two bands in the agarose gel (~ 800 and ~850bp) (Figure 4). The size difference is probably due to some variation in the extein (C and N-terminal) sequences, because even in a partial sequencing we noted some indels in the extein portions, while no indel was observed in the intein. Curiously, despite the usual conservation of the intein insertion sites, the nucleophile +1 of the C-extein of the CmeThrRS is Trp in the larger allele and Cys in the smaller one (data not shown). As to the sequence variation in the intein sequences of these alleles, there are four amino acid substitutions (one in block B, one in block F and two in block G. All the ThrRS inteins herein evaluated present a glycine residue instead of the usual histidine in the block G of the splicing domain.

The genetic distance of ThrRS inteins are represented in Table Supplementary material 2. In the phylogenetic tree constructed for ThrRS intein sequences, the intein CreRBP2 (from the RNA polymerase II of the *Chlamydomonas reinhardtii*) was used as the outgroup. Furthermore, the resultant phylogeny, despite corroborating the relationship among isolates from *psilosis* group and from *C. tropicalis* species, does not confirm the phylogeny already proposed, using the rDNA region (ITS) (Tavanti et al., 2005), for the *psilosis* complex, because *C. parapsilosis* and *C. metapsilosis*, rather than *C. metapsilosis* and *C. orthopsilosis*, share a more recent common ancestor (Figure 5) given that CorThrRS has an unusual sequence divergence compared to the other species from the *psilosis* complex.

Discussion

Inteins are parasitic genetic elements whose protein splicing is considered vital for the host cell because intein excision is essential for host protein functionality (Liu, 2000). Thus, most of the inteins occur in conserved regions of housekeeping proteins (Swithers et al., 2009). These conserved sites do not disturb the splicing function and also maintain the homing site for intein invasion by homing endonuclease (Liu, 2000). This is the case of both host proteins, whose inteins were herein studied. The VMA protein is highly conserved due to its important functions of transporting ions and acidifying organelles (Yamashiro et al., 1990), while the threonyl-tRNA synthetase encodes an aminoacyl-tRNA synthetase, responsible for engaging the threonine tRNA with the corresponding anticodon (Alberts et al., 2002). Both of them encode for proteins that are essential to cell physiology. In this work we evaluated both inteins, from VMA and ThrRS host protein, in some species of the *Candida* genus, in relation to their sequence variability, splicing, HE domains and phylogenetic relationships.

VMA intein

Usually, the canonical inteins have a Cys residue as the first nucleophile of block A and Asn as the last of block G, a motif Thr-x-x-His in block B as well as a Cys, Thr or Ser in the first C-terminal residue of the host protein. All these amino acid residues are involved in protein splicing and in the present study were found in all VMA sequences, with the exception of the motif Thr-x-x-His in block B, known to assist in the initial acyl rearrangement at the N-terminal splicing junction by hydrogen bonding to main chain atoms and hold the residue preceding the intein (Klabunde, 1998 and Poland, 2000). However, any residue that can perform similar hydrogen bonds can replace these conserved residues (Perler et al., 2000). Besides the presence of the residues required for protein splicing, the highly conserved blocks A, B, F and G indicate that the splicing function is operating both in CglVMA and CtrVMA.

On the other hand the HE domains of CglVMA and CtrVMA do not have the two aspartates known to be essential for the activity of endonucleases from the LAGLIDADG family. According to structural studies on SceVMA these aspartates, Asp218 and Asp326 are directly involved in the double-strand break (Gimble & Stephens, 1995; Posey et al., 2004; Koufopanou and Burt, 2005). The absence of these residues, taken together with the high degeneration observed in the HE domain, leads us to speculate that CglVMA and CtrVMA homing endonucleases, like most of the yeast VMA inteins, are not functional (Posey et al., 2004). This is in agreement with the homing cycle proposed by Burt & Koufopanou (2004): once most of the alleles in a population are occupied by intein invasion, due to the homing process, there is no more constrained selection of a functional endonuclease, which might degenerate and become non-functional, so that empty sites may emerge again.

Curiously, the phylogeny obtained for the VMA intein revealed *C. glabrata* to be a sister species of *C. tropicalis*, contrary to the phylogenetic analysis using the entire genome, in

which *C. glabrata* is clearly closer to *S. cerevisiae* than to any other *Candida* species (Wang, 2009). The topology obtained herein corroborates the phylogenetic study carried out by Poulter et al. (2007). The authors analyzed VMA inteins from many distinct species and concluded that horizontal transfer might have occurred sporadically among the host species. Despite this, the VMA intein should not be ruled out as a suitable molecular marker because it clearly differentiates the species *C. glabrata* from *C. tropicalis*.

ThrRS intein

The CpaThrRS and CmeThrRS are mini-inteins while CtrThrRS and CorThrRS are full-length inteins with a highly degenerated HE domain.

The sequence homogeneity of CpaThrRS among the *C. parapsilosis* (there was no amino acid polymorphism among the sequences) supports the clonal and extremely low variability already documented for this species (Fundyga et al., 2004 and Tavanti et al., 2005). The same cannot be inferred for the threonyl intein among the other species of the *psilosis* complex due to the low numbers of representative isolates used.

The amplification of two bands of the threonyl intein in *C. metapsilosis* was not expected. Initially, gradient PCRs were performed by increasing the annealing temperature; however, both bands always appeared together in the agarose gel (data not shown). After sequencing, it was clear that the intein was present in the two bands, whose size differed by only four amino acid substitutions. The alignment of the partial sequences showed that the N and C portions of the threonyl host protein (extein) present some indels that may be responsible for the size difference between the bands. Besides that, the primers used for this PCR reaction were not specific for *C. metapsilosis*, but were based on sequences from *C. parapsilosis*, so that we cannot rule out a possible annealing difference between the two alleles. The most intriguing fact about these two alleles is that one of them has a Trp as +1 C-

extein residue while the other has a Cys. Since the C-extein +1 residue known to be involved in splicing is either Cys or Thr or Ser, its replacement with any other residue may affect the splicing function. As a diploid microorganism *C. metapsilosis* presents two different alleles coding for the ThrRS intein and, given our sequencing results, we suggest that these inteins might differ in splicing efficiency.

Another unexpected observation was the many sequencing polymorphisms found in the intein CorThrRS when compared to the other threonyl inteins. Firstly, it is the largest threonyl intein (~1.6Kb) and presents many polymorphic sites even in the splicing domain, not in the blocks A, B, F and G, but in the sequences located among them. As can be inferred by its long branch in the phylogenetic tree (Figure 5), this intein seems to have the highest evolutionary rate among all the threonyl inteins examined in the present work.

According to Poulter et al. (2007) the threonyl intein is closely related to the inteins from the RNA Polymerase II of different species. If the threonyl gene was invaded by an intein from RNA Polymerase, it would not be the first such case in the literature; for instance the TopA intein is present both in topoisomerase I and reverse gyrase (Chute et al., 1998). Due to this phylogenetic relationship, the intein from the RNA Polymerase II of *Chlamydomonas reinhardtii* was herein used as the outgroup. The phylogeny obtained (Figure 5) revealed that this intein is a potential tool for distinguishing the species from the *psilosis* complex because they are quite different, probably due to an ancient origin and a long evolutionary history. Additional isolates of *C. metapsilosis* and *C. orthopsilosis* properly identified should be included in future studies to clarify this subject. In addition to the evolutionary aspect of the threonyl intein in *psilosis* complex, their sequencing differences among *C. parapsilosis*, *C. metapsilosis* and *C. orthopsilosis* will contribute to the development of molecular diagnosis strategies and therefore to a correct treatment of this important mycosis.

Acknowledgments

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP Grants No. 2010/02674-0).

References

- Alberts, B. et al., 2002. *Molecular Biology of the Cell*. New York: Garland Science Publishing. 4th ed. 1400p.
- Aperis, G., Myriounis N., Spanakis, E.K., Mylonakis E. , 2006. Developments in the treatment of candidiasis: more choices and new challenges. *Expert Opin Investig Drugs*. 15, 1319-1336.
- Bruder-nascimento, A., Camargo, C. H., Sugizaki, M. F., 2010. Species distribution and susceptibility profile of *Candida* species in a Brazilian tertiary hospital. *BMC Research Notes*. 3, 1 doi:10.1186/1756-0500-3-1.
- Burt, A., Koufopanou, V., 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Gen. Develop.* 14, 609-615.
- Butler, M.I., Gray, J., Goodwin, T.J., Poulter, R.T.M., 2006. The distribution and evolutionary history of the PRP8 intein. *BMC Evol. Biol.* 6, 1-26.
- Butler, M.I., Poulter, R.T., 2005. The PRP8 inteins in *Cryptococcus* are a source of phylogenetic and epidemiological information. *Fungal Genet. Biol.* 42, 452-463.
- Chen, Y.C., Eisner, J.D., Kattar, M.M., Rassouljian-Barrett, S.L., LaFe, K., Bui, U., Yarfitz, S.L., Limaye, A.P., Cookson, B.T., 2001. Polymorphic Internal Transcribed Spacer Region 1 DNA Sequences Identify Medically Important Yeasts. *J. Clin. Microbiol.* 39, 4004-4051.
- Chen, Y.C., Eisner, J.D., Kattar, M.M., Rassouljian-Barrett, S.L., LaFe, K., Yarfitz, S.L., Limaye, A.P., Cookson, B.T., 2000. Identification of medically important yeasts using PCR-based detection of DNA sequence polymorphisms in the internal transcribed spacer 2 region of the rRNA genes. *J. Clin. Microbiol.* 38, 2302-2310.

- Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F.B., Xu, M.Q., 1996. Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J. Biol. Chem.* 271, 22159-22168.
- Chute, I.C., Hu, Z., Liu, X.Q., 1998. A topA intein in *Pyrococcus furiosus* and its relatedness to the r-gyr intein of *Methanococcus jannaschii*. *Gene* 210, 85-92.
- Ciardo, D.E., Schär, G., Böttger, E.C., Altwegg, M., Bosshard, P.P., 2006. Internal Transcribed Spacer Sequencing versus Biochemical Profiling for Identification of Medically Important Yeasts. *J. Clin. Microbiol.* 44, 77-84.
- Colombo, A. L.; Guimarães, T.; Silva, L. R. B. F.; Monfardini, L. P. A.; Cunha, A. K. B.; Rady, P.; Alves, T.; Rosas, R. C., 2007. Prospective Observational Study of Candidemia in São Paulo, Brazil: Incidence Rate, Epidemiology, and Predictors of Mortality. *Infect Control Hosp Epidemiol.* 28, 570-576.
- Colombo, A.L., Nucci, M., Salomão, R., Branchini, M.L., Richtmann, R., Derossi, A., Wey, S.B., 1999. High rate of non-albicans candidemia in Brazilian tertiary care hospitals. *Diagn. Microbiol. Infect. Dis.*, 34, 281-286.
- Fundyga, R. E., Kuykendall, R. J., Lee-Yang, W., Lott, T. J., 2004. Evidence for aneuploidy and recombination in the human commensal yeast *Candida parapsilosis*. *Infect. Genet. Evol.* 4, 37-43.
- Gimble, F.S., Stephens, B.W., 1995. Substitutions in conserved dodecapeptide motifs that uncouple the DNA binding and DNA cleavage activities of PI-SceI endonuclease. *J. Biol. Chem.* 270, 5849-5856.

- Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K., Anraku, Y., 1990. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H (+)- translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. J. Biol. Chem. 265, 6726-6733.
- Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebel, M., Stevens, T.H., 1990. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science 250, 651-657.
- Klabunde, T., Sharma, S., Telenti, A., Jacobs, W.R.Jr., Sacchettini, J.C., 1998. Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. Nat Struct Biol 5, 31-36.
- Koufopanou, V., Goddard, M.R., Burt, A., 2002. Adaptation for horizontal transfer in a homing endonuclease. Mol. Biol. Evol. 19, 239-246.
- Koufopanou, V., Burt, A., 2005. Degeneration and domestication of a selfish gene in yeast: molecular evolution versus site-directed mutagenesis. Mol. Biol. Evol. 22, 1535–1538.
- Kurtzman, C.P., Fell, J.W., 1998. *The Yeasts, a taxonomic study*. Amsterdam: Elseviers Scientific., 4th ed.
- Leaw, S.N., Chang, H.C., Sun, H.F., Barton, R., Bouchara, J.P., Chang, T.C., 2006. Identification of Medically Important Yeast Species by Sequence Analysis of the Internal Transcribed Spacer Regions. J. Clin. Microbiol. 94, 693-699.
- Linton, C.J., Borman, A.M., Cheung, G., Holmes, A.D., Szekely, A., Palmer, M.D., Bridge, P.D., Campbell, C.K., Johnson, E.M., 2007. Molecular identification of unusual pathogenic yeast isolates by large ribosomal subunit gene sequencing: 2 years of experience at the United kingdom mycology reference laboratory. J. Clin. Microbiol. 45, 1152-1158.

- Liu, X.Q., 2000. Protein-Splicing Intein: Genetic Mobility, Origin and Evolution. *Ann. Rev. Genet.* 34, 61-76.
- Lott, T.J., Kuykendall R.J., Reiss E., 1993. Nucleotide sequence analysis of the 5.8S rDNA and adjacent ITS2 region of *Candida albicans* and related species. *Yeast.* 11, 1199-1206.
- Mannarelli, B.M., Kurtzman, C.P., 1998. Rapid identification of *Candida albicans* and other human pathogenic yeasts by using short oligonucleotides in a PCR. *J Clin Microbiol.* 36, 1634-1641.
- McCullough, M.J., DiSalvo, A.F., Clemons, K.V., Park, P., and Stevens, D. A., 2000. Molecular epidemiology oh *Blastomyces dermatitidis*. *Clin. Infect. Dis.* 30, 328-335.
- Nucci, M., Colombo, A.L., 2007. Candidemia due to *Candida tropicalis*: clinical, epidemiologic, and microbiologic characteristics of 188 episodes occurring in tertiary care hospitals. *Diagn. Microbiol. Infect. Dis.* 58, 77–82.
- Passos, X.S., Costa, C.R., Araújo, C.R., Nascimento, E.S., Souza, L.K., Fernandes Ode, F., Sales, W.S., Silva Mdo, R., 2007. Species distribution and antifungal susceptibility patterns of *Candida* spp. bloodstream isolates from a Brazilian tertiary care hospital. *Mycopathologia.* 163, 145-151.
- Perler, F.B., 2000. InBase, the Intein Database. *Nucleic acids res.* 28, 344-345.
- Pfaller, M.A., Diekema, D.J., Gibbs, D.L., Newell, V.A., Ng, K.P., Colombo, A.L., Finkelievich, J., Barnes ,R., Wadula, J., 2008. The Global Antifungal Surveillance Group: Geographic and temporal yrends in isolation and antifungal Susceptibility of *Candida parapsilosis*: a global assessment from the ARTEMIS DISK antifungal surveillance program, 2001 to 2005. *J. Clin. Microbiol.*, 46, 842-849.

- Pfaller, M. A., et al. 1998. National surveillance of nosocomial bloodstream infections due to species of *Candida* other than *Candida albicans*: frequency of occurrence and antifungal susceptibility in the SCOPE program. *Diagn. Microbiol. Infect. Dis.* 39, 121-129.
- Poland, B.W., Xu, M.Q., Quirocho, F.A., 2000. Structural insights into the protein splicing mechanism of PI-SceI. *J. Biol. Chem.* 275, 16408-16413.
- Posey, K.L., Koufopanou, V., Burt, A., Gimble, F.S., 2004. Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. *Nucleic acids res.* 32, 3947-3956.
- Poulter, R.T., Goodwin, T.J., Butler, M.I., 2007. The nuclear-encoded inteins of fungi. *Fungal Genet. Biol.* 44, 153-179.
- Richardson, M. D., 2005. Changing patterns and trends in systemic fungal infections. *J. Antimicrob. Chemother.* 56, 5-11.
- Suh, S.O., Kurtzman, C.O., Lachance, M.A., 2006. Phylogenetics of Saccharomycetales, the ascomycete yeasts. *The Mycol. Soc. Of America. Mycologia.* 98, 1006-1017.
- Swithers, K.S., Senejani, A.G., Fournier, G.P., Gogarten, P., 2009. Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol Biol.* 9, 303 doi:10.1186/1471-2148-9-303.
- Tamura, K., 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution.* 9, 678-687.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* (In Press).

- Tavanti, A., Davidson, A.D., Gow, N.A., Maiden, M.C., Odds, F.C., 2005. *Candida orthopsilosis* and *Candida metapsilosis* spp. nov. to replace *Candida parapsilosis* groups II and III. J. Clin. Microbiol. 43, 284-292.
- Theodoro, R.C., Bagagli, E., Oliveira, C., 2008. Phylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complex. Fungal Genet. Biol. 45, 1284-1291.
- Trofa, D., Gácsér, A., Nosanchuck, J. D., 2008. *Candida parapsilosis*, an emerging fungal pathogen. Clin. Microbol. Rew. 214, 606-625.
- Yamashiro, C.T., Kane, P. M., Wolczyk, D. F., 1990. Role of vacuolar acidification in protein sorting and zymogen activation: a genetic analysis of the yeast vacuolar proton-translocating ATPase. Mol Cell Biol. 10, 3737-3749.
- Yokoyama, K., Biswas, S.K., Miyaji, M., Nishimura, K., 2000. Identification and phylogenetic relationship of the most common pathogenic *Candida* species inferred from mitochondrial cytochrome b gene sequences. J. Clin. Microbiol. 38, 4503-4510.
- Wang, H., Xu, Z., Gao, L., Hao, B., 2009., A fungal phylogeny based on 82 complete genomes using the composition vector method. BMC Evol. Biol. 10,9-195.
- Wu, H., Hu, Z., Liu, X.Q., 1998. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. Proc. Nat. Acad. Sci. USA 95, 9226-9231.
- Xu, M., Southworth, M.W., Mersha, F.B., Hornstra, L.J., Perler, F.B., 1993. In vitro protein splicing of purified precursor and the identification of a branched intermediate. Cell. 75, 1371-1377.

SceVMA_Ref.	CFAGKGTNVL	ADGSIIECIEN	IEVGDKVMGK	DGRPREVIKL	PRGRETMYSV	VQKSQHRAHK	SDSSREVPEL	LKFTCNATHE
CglVMA_Ref.	CFAGKTQVIM	ADGSNQSIEN	IKIGDKVMGQ	DGKPRNVTAL	PRGYDDIMYNV	ELDGETD---	-----	LSYTCNSNHT
CglVMA_S-130	CFAGKTQVIM	ADGSNQSIEN	IKIGDKVMGQ	DGKARNVTAL	PRGYDDIMYNV	ELDGETD---	-----	LSYTCNSNHT
CglVMA_S-1	CFAGKTQVIM	ADGSNQSIEN	IKIGDKVMGQ	DGKPRNVTAL	PRGYDDIMYNV	ELDGETD---	-----	LSYTCNSNHT
CglVMA_AC-33	CFAGKTQVIM	ADGSNQSIEN	IKIGDKVMGQ	DGKARNVTAL	PRGYDDIMYNV	ELDGETD---	-----	LSYTCNSNHT
CtrlVMA_Ref.	CFTKGTQVMM	ADGADKSIES	IEVGDKVMGK	DGMPREVVL	PRGYDDIMYKV	RQLSSTRRNA	KSEG-----L	MDFTVSADHK
CtrlVMA_S-36	CFTKGTQVMM	ADGADKSIES	IEVGDKVMGK	DGMPREVVL	PRGYDDIMYKV	RQLSSTRRNA	KSEG-----L	MDFTVSADHK
CtrlVMA_S-39	CFTKGTQVMM	ADGADKSIES	IEVGDKVMGK	DGMPREVVL	PRGYDDIMYKV	RQLSSTRRNA	KSEG-----L	MDFTVSADHK
CtrlVMA_S-35	CFTKGTQVMM	ADGADKSIES	IEVGDKVMGK	DGMPREVVL	PRGYDDIMYKV	RQLSSTRRNA	KSEG-----L	MDFTVSADHK
CtrlVMA_S-158	CFTKGTQVMM	ADGADKSIES	IEVGDKVMGK	DGMPREVVL	PRGYDDIMYKV	RQLSSTRRNA	KSEG-----L	MDFTVSADHK
	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
A					B			
SceVMA_Ref.	LVLKTEQNVL	LAGNTVS---	YFALGALIDE	TNGRA----	EVVQEVQETF	ESNIS-----	---ASDFAAN	IN-REPISWT
CglVMA_Ref.	LVLKTEQNVL	LAGNTVS---	YFALGALIDE	TNGRA----	EVVQEVQETF	ESNIS-----	---ASDFAAN	IN-REPISWT
CglVMA_S-130	LVLKTEQNVL	LAGNTVS---	YFALGALIDE	TNGRA----	EVVQEVQETF	ESNIS-----	---ASDFAAN	IN-REPISWT
CglVMA_S-1	LVLKTEQNVL	LAGNTVS---	YFALGALIDE	TNGRA----	EVVQEVQETF	ESNIS-----	---ASDFAAN	IN-REPISWT
CglVMA_AC-33	LVLKTEQNVL	LAGNTVS---	YFALGALIDE	TNGRA----	EVVQEVQETF	ESNIS-----	---ASDFAAN	IN-REPISWT
CtrlVMA_Ref.	LILKTKQDVK	IATRKIGGNT	YTGVTFFVLE	KTKTG----	ELVKAKTKVF	GHHIHGQNGA	EEKAATFAAG	IDSKEYIDWI
CtrlVMA_S-36	LILKTKQDVK	IATRKIGGNT	YTGVTFFVLE	KTKTG----	ELVKAKTKVF	GHHIHGQNGA	EEKAATFAAG	IDSKEYIDWI
CtrlVMA_S-39	LILKTKQDVK	IATRKIGGNT	YTGVTFFVLE	KTKTG----	ELVKAKTKVF	GHHIHGQNGA	EEKAATFAAG	IDSKEYIDWI
CtrlVMA_S-35	LILKTKQDVK	IATRKIGGNT	YTGVTFFVLE	KTKTG----	ELVKAKTKVF	GHHIHGQNGA	EEKAATFAAG	IDSKEYIDWI
CtrlVMA_S-158	LILKTKQDVK	IATRKIGGNT	YTGVTFFVLE	KTKTG----	ELVKAKTKVF	GHHIHGQNGA	EEKAATFAAG	IDSKEYIDWI
	* * *		*		* *			*
B					C			
SceVMA_Ref.	IEARDLSLLG	SHVRKATYQT	YAPILYENDH	FFDYMQKSKF	HLTIEGPKVL	AYLLGLWIGD	GLSDRAITFSV	DSRD-TSLME
CglVMA_Ref.	LEIRDIDYLS	ERVVMFTKQS	VNPVLLTPT	LAKQLESNE-	-STATN---L	AYLLGTWIAS	KATTAGTISV	PTTKADLLSK
CglVMA_S-130	LEIRDIDYLS	ERVVMFTKQS	VNPVLLTPT	LAKQLESNE-	-STATN---L	AYLLGTWIAS	KATTAGTISV	PTTKADLLSK
CglVMA_S-1	LEIRDIDYLS	ERVVMFTKQS	VNPVLLTPT	LAKQLESNE-	-STATN---L	AYLLGTWIAS	KATTAGTISV	PTTKADLLSK
CglVMA_AC-33	LEIRDIDYLS	ERVVMFTKQS	VNPVLLTPT	LAKQLESNE-	-STATN---L	AYLLGTWIAS	KATTAGTISV	PTTKADLLSK
CtrlVMA_Ref.	IEARDYVQVD	EIVKTSTTQM	INPVHFESGK	LGNWLHEHKQ	NKSLAP--QL	GYYLGTWAGI	GNVKSSAFTM	NSKDDVKLAT
CtrlVMA_S-36	IEARDYVQVD	EIVKTSTTQM	INPVHFESGK	LGNWLHEHKQ	NKSLAP--QL	GYYLGTWAGI	GNVKSSAFTM	NSKDDVKLAT
CtrlVMA_S-39	IEARDYVQVD	EIVKTSTTQM	INPVHFESGK	LGNWLHEHKQ	NKSLAP--QL	GYYLGTWAGI	GNVKSSAFTM	NSKDDVKLAT
CtrlVMA_S-35	IEARDYVQVD	EIVKTSTTQM	INPVHFESGK	LGNWLHEHKQ	NKSLAP--QL	GYYLGTWAGI	GNVKSSAFTM	NSKDDVKLAT
CtrlVMA_S-158	IEARDYVQVD	EIVKTSTTQM	INPVHFESGK	LGNWLHEHKQ	NKSLAP--QL	GYYLGTWAGI	GNVKSSAFTM	NSKDDVKLAT
	* **	* **	* **			* ****		*
					C			
SceVMA_Ref.	RVTEYAEKLN	LCAEYKDRKE	PQVAKTVN-L	YSKVVRGNGI	RNNLN-----	-----	-----TEN	PLWDAIVGLG
CglVMA_Ref.	VKSALSSLSI	DYSSESINSV	STYRRQTQIP	LMENGKHVGN	ANITAEQIE	ETMEVLSLN-	-----VTN	HSSKLFHDLA
CglVMA_S-130	VKSALSSLSI	DYSSESINSI	STYRRQTQIP	LMENGKHVGN	ANITAEQIE	ENMEMLSLN-	-----VTN	HSSKLFHDLA
CglVMA_S-1	VKSALSSLSI	DYSSESINSI	STYRRQTQIP	LMENGKHVGN	ANITAEQIE	ENMEMLSLN-	-----VTN	HSSKLFHDLA
CglVMA_AC-33	VKSALSSLSI	DYSSESINSV	STYRRQTQIP	LMENGKHVGN	ANITAEQIE	ETMEVLSLN-	-----VTN	HSSKLFHDLA
CtrlVMA_Ref.	RIMNYSSKLG	MTCSSSTESGE	LNVAENEEEF	FNNLGAEKDE	AGDFTFDEFT	DAMDELTINV	HGAAASKKNN	LLWNALKSLG
CtrlVMA_S-36	RIMNYSSKLG	MTCSSSTESGE	LNVAENEEEF	FNNLGAEKDE	AGDFTFDEFT	DAMDELTINV	HGAAASKKNN	LLWNALKSLG
CtrlVMA_S-39	RIMNYSSKLG	MTCSSSTESGE	LNVAENEEEF	FNNLGAEKDE	AGDFTFDEFT	DAMDELTINV	HGAAASKKNN	LLWNALKSLG
CtrlVMA_S-35	RIMNYSSKLG	MTCSSSTESGE	LNVAENEEEF	FNNLGAEKDE	AGDFTFDEFT	DAMDELTINV	HGAAASKKNN	LLWNALKSLG
CtrlVMA_S-158	RIMNYSSKLG	MTCSSSTESGE	LNVAENEEEF	FNNLGAEKDE	AGDFTFDEFT	DAMDELTINV	HGAAASKKNN	LLWNALKSLG
								* *

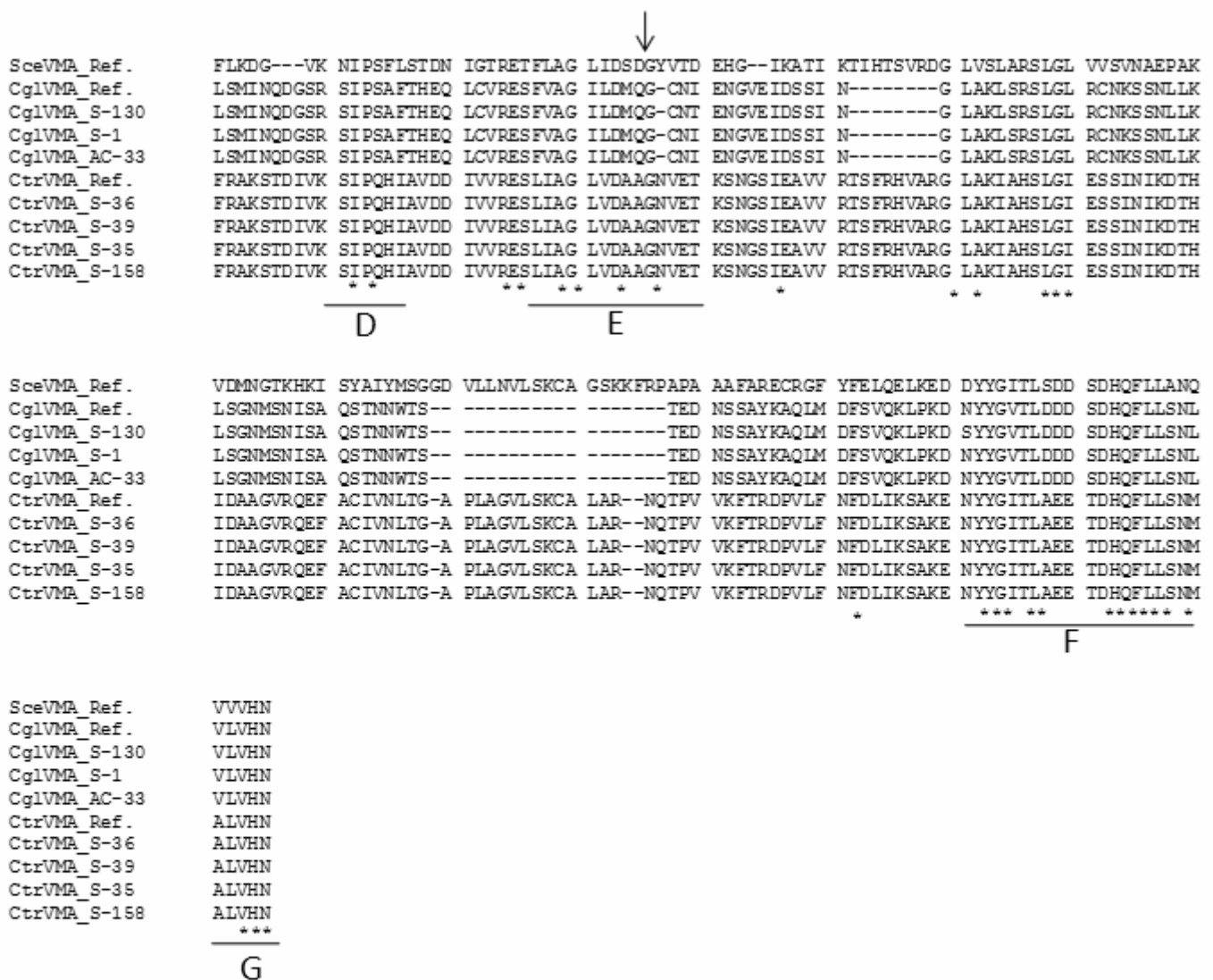


Figure 1: Alignment of the deduced amino acid sequences of SceVMA, CtrVMA and CglVMA inteins. Deduced amino acid sequences of SceVMA, CtrVMA and CglVMA inteins. The arrows indicate the position of the two essential aspartate residues according to VMA intein of *S. cerevisiae*. The * indicates amino acid identity.

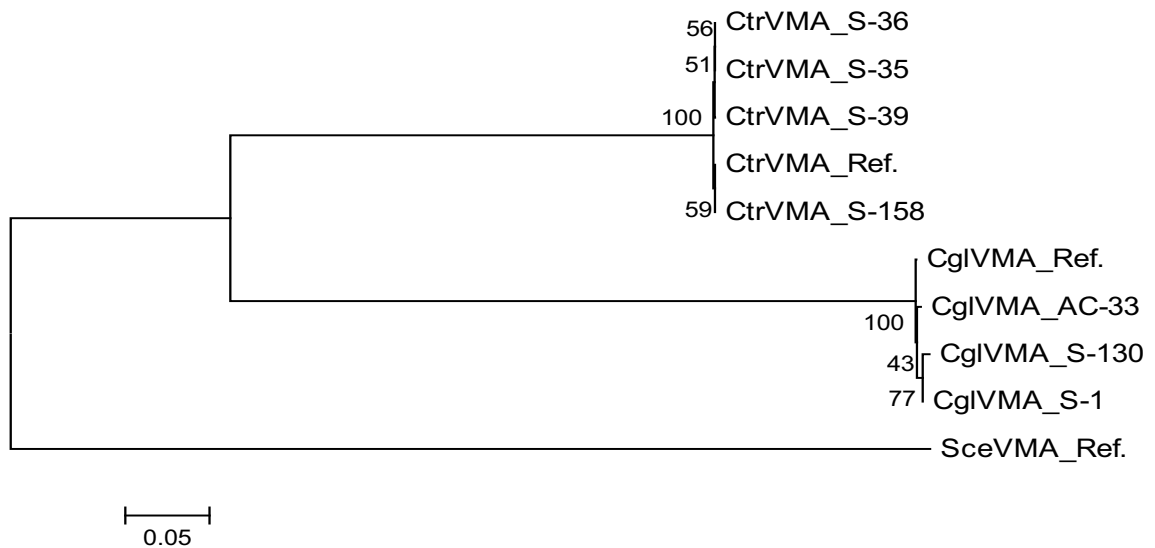


Figure 2: Neighbor Joining Tree based on the amino-acids sequences of the VMA intein from *C. tropicalis*, *C. glabrata* and *Saccharomyces cerevisiae*. Bootstrap proportions are shown above branches.

CtrThrRs_Ref.	AANEANSAWL	SKAISELEGK	VDGTTITDSK	LSALFNGSEQ	SFAEWVWDLA	KDELRSVVHG
CtrThrRs_S-36	AANEANSAWL	SKAISELEGK	VDGTTITDSK	LSALFNGSEQ	SFAEWVWDLA	KDELRSVVHG
CtrThrRs_S-48	AANEANSAWL	SKAISELEGK	VDGTTITDSK	LSALFNGSEQ	SFAEWVWDLA	KDELRSVVHG
CtrThrRs_S-5	AANEANSAWL	SKAISELEGK	VDGTTITDSK	LSALFNGSEQ	SFAEWVWDLA	KDELRSVVHG
CorThrRS_ATCC	-----	-----	-----	----YTPIKA	KYDNWVINYA	EDGSPS----
CmeThrRs_B1	-----	-----	-----	-----	-----A	VDESSSF---
CmeThrRs_B2	-----	-----	-----	-----	-----A	VDETSSF---
CpaThrRs_S-18	-----	-----	-----	-----	-----A	AEESSASSAHA
CpaThrRs_S-33	-----	-----	-----	-----	-----A	AEESSASSAHA
CpaThrRs_S-41	-----	-----	-----	-----	-----A	AEESSASSAHA
CpaThrRs_S-45	-----	-----	-----	-----	-----A	AEESSASSAHA
CpaThrRs_S-46	-----	-----	-----	-----	-----A	AEESSASSAHA
CpaThrRs_S-50	-----	-----	-----	-----	-----A	AEESSASSAHA
CpaThrRs_Ref.	-----	-----	-----	-----	-----A	AEESSASSAHA
					*	*
↓						
CtrThrRs_Ref.	FARASGDENK	KIYTSSVILR	DELVRVLLHA	GYTSRFELNA	TKGWEITYVE	DVAQCVNPVL
CtrThrRs_S-36	FARASGDENK	KIYTSSVILR	DELVRVLLHA	GYTSRFELNA	TKGWEITYVE	DVAQCVNPVL
CtrThrRs_S-48	FARASGDENK	KIYTSSVILR	DELVRVLLHA	GYTSRFELNA	TKGWEITYVE	DVAQCVNPVL
CtrThrRs_S-5	FARASGDENK	KIYTSSVILR	DELVRVLLHA	GYTSRFELNA	TKGWEITYVE	DVAQCVNPVL
CorThrRS_ATCC	----GK---	-----	-----	-----	-----	---QCVRPNL
CmeThrRs_B1	-----	-----	-----	-----	-----	----TKTSL
CmeThrRs_B2	-----	-----	-----	-----	-----	----TKTSL
CpaThrRs_S-18	F-----	-----	-----	-----	-----	----VNITL
CpaThrRs_S-33	F-----	-----	-----	-----	-----	----VNITL
CpaThrRs_S-41	F-----	-----	-----	-----	-----	----VNITL
CpaThrRs_S-45	F-----	-----	-----	-----	-----	----VNITL
CpaThrRs_S-46	F-----	-----	-----	-----	-----	----VNITL
CpaThrRs_S-50	F-----	-----	-----	-----	-----	----VNITL
CpaThrRs_Ref.	F-----	-----	-----	-----	-----	----VNITL
						*
CtrThrRs_Ref.	YSDKNVKVVD	DYFGRVWCVT	VPTGLIIVQR	VVKNAEDVVV	KASRPITIVGN	
CtrThrRs_S-36	YSDKNVKVVD	DYFGRVWCVT	VPTGLIIVQR	VVKNAEDVVV	KASRPITIVGN	
CtrThrRs_S-48	YSDKNVKVVD	DYFGRVWCVT	VPTGLIIVQR	VVKNAEDVVV	KASRPITIVGN	
CtrThrRs_S-5	YSDKNVKVVD	DYFGRVWCVT	VPTGLIIVQR	VVKNAEDVVV	KASRPITIVGN	
CorThrRS_ATCC	YSDTEIKRVD	NYVGRVWCVT	VPTGLIIAQR	VHRDADNNVI	RASRPITIVGN	
CmeThrRs_B1	NSKTEIKAVG	DYNGRVWCVT	VPTGLIIAHR	VHRGADGNVI	RSKPIVIVGN	
CmeThrRs_B2	NSKTEIKAVG	DYNGRVWCVT	VPTGLIIAHR	VHRGADGNVI	RASEPVVIGN	
CpaThrRs_S-18	NSTTEIRAVG	DYQGRVWCVT	VPTGLIVAGR	VHRDTSGNVV	CASKPVVIGN	
CpaThrRs_S-33	NSTTEIRAVG	DYQGRVWCVT	VPTGLIVAGR	VHRDTSGNVV	CASKPVVIGN	
CpaThrRs_S-41	NSTTEIRAVG	DYQGRVWCVT	VPTGLIVAGR	VHRDTSGNVV	CASKPVVIGN	
CpaThrRs_S-45	NSTTEIRAVG	DYQGRVWCVT	VPTGLIVAGR	VHRDTSGNVV	CASKPVVIGN	
CpaThrRs_S-46	NSTTEIRAVG	DYQGRVWCVT	VPTGLIVAGR	VHRDTSGNVV	CASKPVVIGN	
CpaThrRs_S-50	NSTTEIRAVG	DYQGRVWCVT	VPTGLIVAGR	VHRDTSGNVV	CASKPVVIGN	
CpaThrRs_Ref.	NSTTEIRAVG	DYQGRVWCVT	VPTGLIVAGR	VHRDTSGNVV	CASKPVVIGN	
	*	*	*	*****	*****	* * *
				F	G	

Figure 3: Alignment of the deduced amino acid sequences of CtrThrRS, CmeThrRS, CorThrRS and CpaThrRS inteins. Deduced amino acid sequences of CtrThrRS, CmeThrRS, CorThrRS and CpaThrRS inteins. The arrows indicate the position of the two essential aspartic acids according to VMA intein of *S. cerevisiae*. The * indicates amino acid identity.

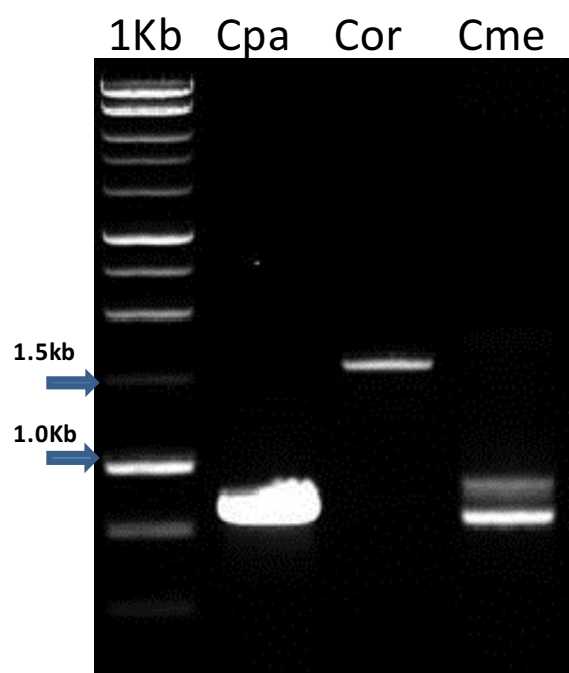


Figure 4: Agarose gel electrophoresis of the PCR carried out with P78 and P79 primers for amplification of ThrRS inteins from *C. parapsilosis* (Cpa), *C. orthopsilosis* (Cor) and *C. metapsilosis* (Cme). Lane 1: 1Kb DNA Ladder (Promega).

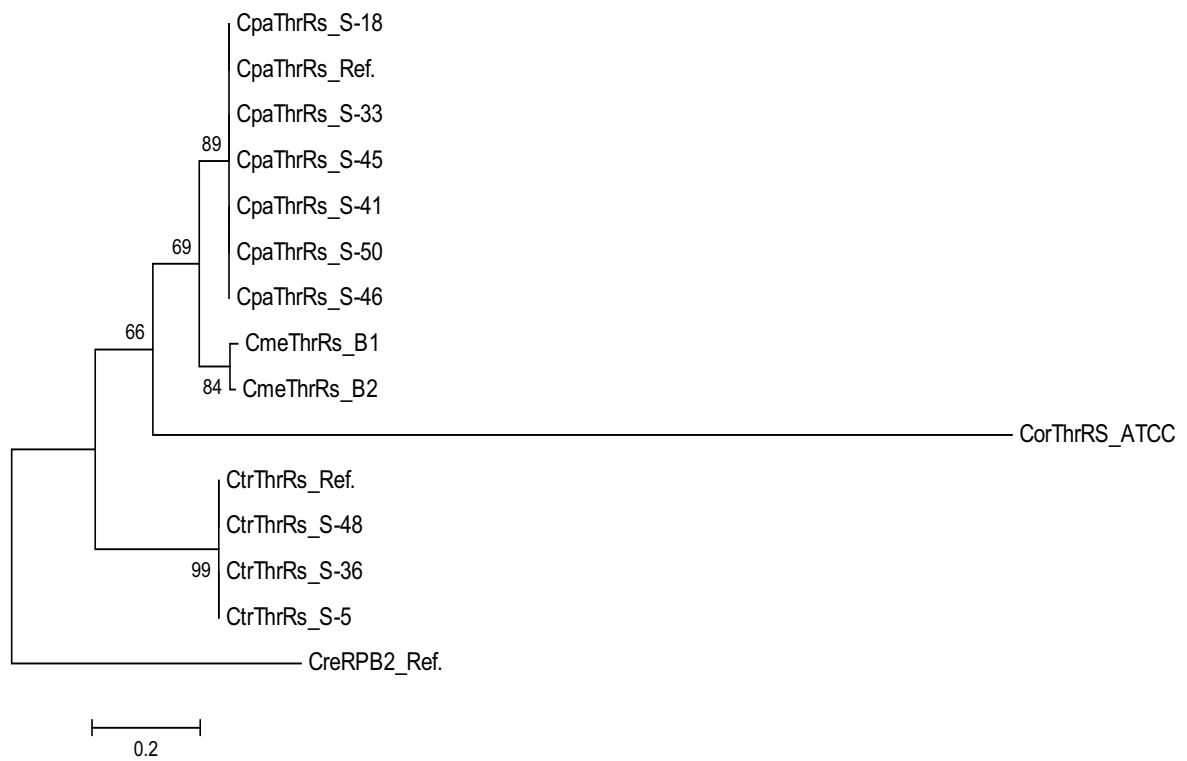


Figure 5: Neighbor Joining Tree based on the amino-acids sequences of the ThrRS intein from *C. tropicalis*, *C. parapsilosis*, *C. metapsilosis*, *C. orthopsilosis* and *Chlamydomonas reinhardtii*. Bootstrap proportions are shown above branches.

	1	2	3	4	5	6	7	8	9	10
1. SceVMA_Ref.		0.0789	0.0800	0.0775	0.0800	0.0562	0.0561	0.0562	0.0561	0.0562
2. CglVMA_Ref.	1.0827		0.0025	0.0019	0.0014	0.0414	0.0415	0.0415	0.0415	0.0414
3. CglVMA_S-130	1.0878	0.0081		0.0017	0.0023	0.0415	0.0416	0.0416	0.0416	0.0415
4. CglVMA_S-1	1.0710	0.0048	0.0032		0.0024	0.0412	0.0413	0.0414	0.0413	0.0412
5. CglVMA_AC-33	1.0903	0.0024	0.0073	0.0073		0.0417	0.0419	0.0419	0.0419	0.0417
6. CtrVMA_Ref.	0.9602	0.6918	0.7009	0.6967	0.6935		0.0012	0.0010	0.0012	0.0000
7. CtrVMA_S-36	0.9604	0.6919	0.7010	0.6968	0.6936	0.0021		0.0007	0.0000	0.0012
8. CtrVMA_S-39	0.9606	0.6920	0.7011	0.6969	0.6937	0.0014	0.0007		0.0007	0.0010
9. CtrVMA_S-35	0.9604	0.6919	0.7010	0.6968	0.6936	0.0021	0.0000	0.0007		0.0012
10. CtrVMA_S-156	0.9602	0.6918	0.7009	0.6967	0.6935	0.0000	0.0021	0.0014	0.0021	

SM1: Estimates of Evolutionary Divergence between VMA intein sequences. The number of base substitutions per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal. Analyses were conducted using the Tamura 3-parameter model (Tamura, 1992). The analysis involved 10 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1431 positions in the final dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CreRPB2_Ref.		0.1386	0.1386	0.1386	0.1386	3.3389	0.1314	0.1309	0.1324	0.1324	0.1324	0.1324	0.1324	0.1324	0.1324
2. CtrThrRs_Ref.	0.9154		0.0000	0.0000	0.0000	0.3353	0.0670	0.0687	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721
3. CtrThrRs_S-36	0.9154	0.0000		0.0000	0.0000	0.3353	0.0670	0.0687	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721
4. CtrThrRs_S-48	0.9154	0.0000	0.0000		0.0000	0.3353	0.0670	0.0687	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721
5. CtrThrRs_S-5	0.9154	0.0000	0.0000	0.0000		0.3353	0.0670	0.0687	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721	0.0721
6. CorThrRS_ATCC	3.1485	1.6883	1.6883	1.6883	1.6883		0.3028	0.3149	0.3649	0.3649	0.3649	0.3649	0.3649	0.3649	0.3649
7. CmeThrRs_B1	0.9555	0.4563	0.4563	0.4563	0.4563	1.5318		0.0135	0.0322	0.0322	0.0322	0.0322	0.0322	0.0322	0.0322
8. CmeThrRs_B2	0.9397	0.4567	0.4567	0.4567	0.4567	1.5658	0.0232		0.0302	0.0302	0.0302	0.0302	0.0302	0.0302	0.0302
9. CpaThrRs_S-18	0.9146	0.4635	0.4635	0.4635	0.4635	1.7251	0.1248	0.1166		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10. CpaThrRs_S-33	0.9146	0.4635	0.4635	0.4635	0.4635	1.7251	0.1248	0.1166	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000
11. CpaThrRs_S-41	0.9146	0.4635	0.4635	0.4635	0.4635	1.7251	0.1248	0.1166	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000
12. CpaThrRs_S-45	0.9146	0.4635	0.4635	0.4635	0.4635	1.7251	0.1248	0.1166	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000
13. CpaThrRs_S-46	0.9146	0.4635	0.4635	0.4635	0.4635	1.7251	0.1248	0.1166	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000
14. CpaThrRs_S-50	0.9146	0.4635	0.4635	0.4635	0.4635	1.7251	0.1248	0.1166	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000
15. CpaThrRs_Ref.	0.9146	0.4635	0.4635	0.4635	0.4635	1.7251	0.1248	0.1166	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

SM2: Estimates of Evolutionary Divergence between ThrRS intein sequences.

The number of base substitutions per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal. Analyses were conducted using the Tamura 3-parameter model (Tamura, 1992). The analysis involved 10 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1431 positions in the final dataset.