

UNIVERSIDADE ESTADUAL PAULISTA – UNESP

INSTITUTO DE BIOCÊNCIAS

Raquel Cordeiro Theodoro

**Gênero *Paracoccidioides*: especiação e
análise do intein PRP8**

Tese apresentada ao Instituto de Biociências da Universidade Estadual Paulista – UNESP, Campus de Botucatu, como parte dos requisitos para a Defesa de Doutorado em Ciências Biológicas, Área de concentração: Genética.

Orientador: Dr. Eduardo Bagagli

Botucatu, 2010

Theodoro, Raquel Cordeiro

Gênero *Paracoccidioides*: especiação e análise do *intein* PRP8
/ Raquel Cordeiro Theodoro. - Botucatu, 2010

Tese (doutorado) - Instituto de Biociências de Botucatu, Universidade
Estadual Paulista, 2010

Orientador: Eduardo Bagagli

Capes: 20202008

1. *Paracoccidioides* 2. Paracoccidioidomicose

Palavras-chave: Ajellomycetaceae; endonuclease; espécies crípticas;
intein PRP8; marcadores moleculares; *Paracoccidioides*; *splicing*
protéico

Agradecimentos

Àqueles diretamente envolvidos nesta tese de doutorado:

Ao orientador Prof. Dr. Eduardo Bagagli, por sempre me estimular a encarar novos e interessantes desafios, mostrando que qualquer obstáculo pode ser vencido com curiosidade e vontade de aprender.

Ao supervisor Prof. Dr. Xiang-Qin Liu (Dalhousie University), por abrir as portas de seu laboratório, por sua atenção, paciência e apoio oferecidos.

Ao supervisor Dr. Gerrit Volkmann (Dalhousie University), não somente pela ajuda prestada dia-a-dia na bancada, mas também por discutir os resultados comigo e ampliar minha visão sobre a importância do trabalho realizado, além de ter se tornado um grande amigo.

Ao Prof. Dr. Cláudio de Oliveira, por pacientemente me ajudar nas análises filogenéticas do intein PRP8 e esclarecer várias dúvidas sobre evolução.

À Profa. Dra. Maria Sueli Felipe Soares e Marcus de Melo Teixeira (UNB), pela ajuda nas análises biogeográficas e pelas interessantes discussões sobre evolução e filogenética.

Ao Prof. Dr. Paulo Martins Ribolla, Dra. Karina dos Santos Paduan e Carine Spenassatto (IBB-UNESP), por todo suporte oferecido para as análises de SNPs.

À Profa. Dra. Luzia A. Trinca e ao Prof. Reinaldo José da Silva (IBB-UNESP), pela realização das análises estatísticas dos experimentos de morfometria de levedura.

À Profa. Dra. Gioconda San-Blas, pelas valiosas sugestões dadas para o capítulo 1 desta tese.

Àqueles que ficaram torcendo nos bastidores:

À minha mãe, Salete Cordeiro, ao meu pai, Pedro Luiz Theodoro, à minha avó Emília M. Cordeiro, ao meu avô, Francisco Cordeiro Filho (em memória) e ao meu namorado Mateus Ferrareze Feitosa por todo amor, carinho e preocupação e por sempre me incentivarem a fazer aquilo que gosto, mesmo quando tivemos que ficar distantes por um tempo.

Aos tios, tias, primos e primas, que são parte importante da minha vida, pois juntos com meus pais e avós sempre me deram muito amor e me ensinaram o quanto é bom ter uma família unida.

Aos amigos de bancada, Sandra, Assis, Gabriel, Juliana (Bio), Juliana (Biomed), Michele, Ariane, Tarsila, Tâmara, Thales, Mariana, Virgínia, Keila e Lidiany, pela amizade e pelo ambiente divertido e descontraído do nosso lab.

Aos funcionários Sônia, Nice, Ademival e Lula, pelo carinho e atenção sempre dados aos alunos.

Às amigas de bancada da Dalhousie, LingLing, Yi, Yang Fei e Xi Mei, pela troca de experiências, pelo carinho com que me receberam, pelas

comidas gostosas que me davam para experimentar, pelos passeios e risadas e pela amizade que construímos.

Às agências financiadoras

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

À CAPES, pela concessão da bolsa sandwich (programa PDEE) e oportunidade dada para o desenvolvimento de parte desta tese na universidade de Dalhousie, Halifax, Canadá.

Índice

	Pág.
Apresentação da Tese	1
<u>Capítulo 1: Especiação do Gênero Paracoccidioides</u>	3
Revisão de Literatura.	
<i>Paracoccidioides brasiliensis</i> : um complexo de espécies	5
<i>Paracoccidioides brasiliensis</i> : taxonomia, etiologia e ecologia	6
<i>P. brasiliensis</i> : variabilidade intra-específica ou especiação críptica?	9
Identificando novos isolados	21
Referências Bibliográficas	28
Artigo 1	35
<u>Capítulo 2: Intein PRP8</u>	78
Revisão Bibliográfica	
<i>Intein</i> : um elemento genético parasita	80
<i>Intein</i> : uma nova descoberta da biologia molecular	80
Tipos, estrutura e distribuição dos <i>Inteins</i>	82
Domínio <i>Splicing</i>	85
Domínio <i>Homing</i> Endonuclease e o ciclo <i>homing</i>	88
Origem dos <i>inteins</i>	97
<i>Inteins</i> em fungos	98
<i>Intein</i> PRP8	101
Uso de <i>inteins</i> em análises filogenéticas	107
<i>Intein</i> PRP8 como alvo terapêutico	108
Aplicações biotecnológicas	113
Referências Bibliográficas	119
Artigo 1	124
Artigo 2	133
Artigo 3	142

Apresentação da Tese

Esta tese de doutorado está dividida em dois capítulos, de forma a abranger com maiores detalhes os temas Especiação do Gênero *Paracoccidioides* e *Intein* PRP8. Ambos os capítulos são compostos por uma revisão de literatura em português, seguida dos artigos científicos, publicados ou a serem submetidos, resultantes dos experimentos realizados durante o doutorado.

O capítulo 1 discorre sobre aspectos evolutivos e biogeográficos das espécies de fungos patogênicos do gênero *Paracoccidioides*, recentemente descritas, e sua implicação para o diagnóstico da Paracoccidioidomicose. Sendo assim, este capítulo possui duas vertentes, uma evolutiva e outra aplicada. A primeira explora os possíveis eventos geológicos e biogeográficos que teriam resultado na atual distribuição geográfica das quatro espécies do gênero *Paracoccidioides*, e a segunda explora marcadores morfológicos e moleculares para um rápido reconhecimento de espécies dentro do gênero.

O capítulo 2 explora um tema novo e promissor, o *intein* PRP8. Trata-se de um elemento genético parasita que ocorre no gene *PRP8* e na proteína Prp8 de fungos do gênero *Paracoccidioides* e em outras espécies patogênicas. Este elemento sofre um *splicing* protéico pós-traducional que garante a funcionalidade normal da proteína hospedeira. Além disso, alguns destes *inteins* codificam uma *homing* endonuclease, o que os torna móveis, garantindo uma rápida fixação na população. O interesse inicial em se estudar este gene foi o de explorar seu potencial filogenético para distinguir isolados

fúngicos das diferentes espécies do gênero *Paracoccidioides*, um dos objetivos centrais desta tese, bastante discutido no capítulo 1.

Após o sucesso do uso do *intein* PRP8 nas análises filogenéticas aqui realizadas, o potencial terapêutico deste elemento genético nos chamou a atenção, pois ele ocorre em uma proteína vital para a célula fúngica (a Prp8 faz parte do centro catalítico do spliceossomo) e, uma vez o *splicing* inibido a proteína hospedeira não assumiria sua conformação funcional, inviabilizando a sobrevivência do fungo. Além disso, o *intein* PRP8, como vários outros *inteins*, possui aspectos evolutivos peculiares devido à função *homing*, responsável pela sua rápida dispersão nas populações. Em um estudo mais detalhado sobre este gene, investigamos as funções *splicing* e *homing* endonuclease em modelos não nativos de gene/proteína e célula hospedeiras.

Capítulo 1:

Especiação do Gênero Paracoccidioides

“Existe uma grandeza nessa visão de mundo, com seus vários poderes, tendo sido originalmente insuflados em poucas formas, ou em apenas uma; enquanto este planeta continuou girando segundo as leis da gravidade, desde o mais simples começo, infindáveis formas, as mais belas e mais maravilhosas, evoluíram ou estão evoluindo.”

Charles Darwin, A origem das espécies

Revisão de Literatura

***Paracoccidioides brasiliensis*: um complexo de espécies**

A partir de do ano de 2006 a Paracoccidioidomicose (PCM) passou a ser vista como uma micose sistêmica causada por pelo menos três espécies crípticas (S1, PS2 e PS3 do complexo *P. brasiliensis*) e, no ano de 2009, uma quarta espécie, *Paracoccidioides lutzii*, foi definida, sendo a mais divergente do gênero. Apesar das inúmeras controvérsias sobre o conceito e o reconhecimento do que chamamos de espécie, estes quatro “grupos” de isolados, aparentam não trocar alelos entre si na natureza, ou seja, segundo análises *in silico*, estão isolados reprodutivamente.

Definir variedades, agrupamentos genéticos e, principalmente, espécie é essencial não somente para um estudo biológico e evolutivo básico, mas, neste caso, para um objetivo aplicado óbvio: o correto diagnóstico e tratamento da PCM. Porém, definir a espécie à qual um novo isolado de *Paracoccidioides* pertence por sequenciamento de *multi loci* (MLST), como feito inicialmente para a descoberta das espécies crípticas, seria um tanto laborioso. Sendo assim, um dos objetivos centrais dessa tese de doutorado foi o de identificar potenciais marcadores morfológicos e/ou moleculares para um rápido reconhecimento de espécie dentro do gênero *Paracoccidioides*.

Esta revisão de literatura discorre a respeito dos aspectos biológicos básicos do gênero *Paracoccidioides*, dos conceitos de espécie e sua problemática em microrganismos, bem como sobre as estratégias hoje disponíveis e adotadas nesta tese para o reconhecimento de espécie. Aspectos biogeográficos do gênero *Paracoccidioides* serão também apresentados na tentativa de compreender os padrões atuais da distribuição geográfica das diferentes espécies, suas origens e processos de divergência e dispersão.

***Paracoccidioides brasiliensis*: taxonomia, etiologia e ecologia.**

P. brasiliensis é a fase anamórfica (mitospórica ou assexuada) de um ascomiceto pertencente à Ordem Onygenales, Família Ajellomycetaceae, a qual representa a um clado monofilético que inclui também os gêneros *Blastomyces*, *Emmonsia*, *Histoplasma* (Untereiner et al., 2004) e *Lacazia* (Vilela et al., 2009).

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). Este fungo apresenta distribuição geográfica restrita aos países latino-americanos, ocorrendo com maior incidência no Brasil, Colômbia, Venezuela e Argentina (Wanke & Londero, 1994).

Este patógeno possui um dimorfismo termo-dependente, a 35-37°C (*in vitro*, em meios enriquecidos ou no tecido do hospedeiro em condição de parasitismo) possui forma de levedura com multibrotamentos, aspecto conhecido como “roda de leme” (Lacaz et al., 1984) e a 25°C (*in vitro* ou em condições de saprobiose) cresce na forma de micélios aéreos curtos bastante aderidos ao meio de cultura, as hifas são septadas, com artrósporos, clamidósporos intercalares e aleuriósporos globosos (Franco, 1987).

P. brasiliensis tem sua ecologia praticamente desconhecida devido aos dados pouco conclusivos a respeito de seu isolamento ambiental. O caráter casual e não repetitivo das observações, aliado às dificuldades de isolamento ambiental, dificultam a exata localização deste patógeno no ambiente. Além disso, a falta de surtos epidêmicos, o prolongado período de latência da doença e frequentes migrações das populações de áreas endêmicas impossibilitam a identificação dos locais onde as infecções foram adquiridas (Restrepo, 1985).

Apesar destas dificuldades, existe um “certo consenso”, baseado, também, em outras micoses sistêmicas provocadas por fungos dimórficos relacionados, onde, o pulmão é o órgão mais afetado, de que o *P. brasiliensis* vive saprobioticamente no solo, produzindo

estruturas assexuadas na forma de artroconídias (Bustamante-Simon et al., 1985), as quais seriam as responsáveis pela infecção principalmente pelo trato respiratório (McEwen et al., 1987).

Existem alguns casos de isolamento do *P. brasiliensis* em solo (Shome & Batista, 1963; Negroni, 1968; Albornoz, 1971 e Silva-Vergara et al., 1998), sendo que os dois primeiros isolamentos não foram totalmente confirmatórios para *P. brasiliensis*, já que o isolado nem mesmo foi mantido em coleção (Franco et al., 2000). Além destes isolamentos, o patógeno também já foi isolado em ração canina (Ferreira et al., 1990) e fezes de pingüim (Gesuele, 1989) e detectado em trato intestinal de morcegos (Grose & Tramsitt, 1995).

Um novo indício ambiental da presença do *P. brasiliensis* em solo foi a constatação da ocorrência do mesmo, em alta frequência, em tatus da espécie *Dasypus novemcinctus* (Naiff et al., 1986, Macedo et al., 1999; Silva-Vergara et al., 2000; Corredor et al., 1999; Bagagli et al., 1998; 2003), animais com intenso hábito fossorial e escavatório, cuja distribuição geográfica costuma coincidir com a observada na paracoccidioidomicose (Restrepo, 2000). O contato com estes animais está associado a um aumento dos fatores de riscos para a infecção em pessoas residentes nas áreas endêmicas da doença (Cadavid & Restrepo, 1993).

Além da metodologia direta (cultura) e indireta (inoculação animal), métodos moleculares vem sendo empregados para a detecção do *P. brasiliensis* em solo através da PCR (Reação da Polimerase em Cadeia) de regiões do rDNA, que compreendem o ITS1 (Internal transcribed spacer 1), 5.8S e ITS2 (Internal transcribed spacer 2), específicas do *P. brasiliensis* (Theodoro et al., 2005a). A técnica molecular empregada se mostrou eficaz na detecção do DNA do patógeno em tocas de tatu, infectados naturalmente pelo fungo, sendo também adaptada para detecção do patógeno em animais atropelados e amostras clínicas (Richini-Pereira et al., 2007; 2009; Terçarioli et al., 2007).

Por apresentar temperatura corporal e imunidade relativamente baixas (Purtilo et al., 1975; Ulrich et al., 1976), o tatu parece favorecer o desenvolvimento de certas doenças infecciosas, podendo ter desempenhado um importante papel na evolução do *P.*

brasiliensis à condição zoofílica (adaptada ao tecido animal) (Bagagli et al., 1998), o que explicaria uma maior tendência para o estabelecimento de lesões crônicas nos hospedeiros e a baixa produção de conídios quando em saprobiose (McEwen et al., 1987), explicando assim o seu difícil isolamento ambiental (Franco et al., 2000).

Os isolados de tatus apresentam significativa diversidade em relação à virulência (Sano et al., 1999; Hebeler-Barbosa et al., 2003a, 2003b). Hebeler-Barbosa et al (2003a; 2003b) caracterizaram a biodiversidade de isolados de *P. brasiliensis* de tatus, comparando-os com isolados clínicos da região endêmica de Botucatu, com relação aos diferentes perfis de virulência e análises moleculares. Por meio de inoculação animal e contagem de UFC (Unidades Formadoras de Colônias), foi possível caracterizar os isolados quanto à virulência (muito alta, alta, intermediária e baixa). Apesar de não fornecer uma correlação clara com os perfis de virulência, o padrão de RAPD mostrou que os isolados de tatu provenientes de uma mesma área geográfica são mais similares entre si do que isolados provenientes de áreas distantes uma das outras. Estes dados são particularmente úteis para o mapeamento do patógeno na natureza, visto que estes animais possuem uma “home range” pequena (Loughry & McDonough, 1998), locomovendo-se em trilhas relativamente constantes no interior de pequenas matas de galerias (Taber, 1945, Tamage & Buchanan, 1954), não tendo hábito migratório como no caso dos seres humanos. Com relação ao sequenciamento da região ITS1, 5.8S e ITS2, enquanto o gene 5.8S apresentou-se altamente conservado, tanto entre os diferentes isolados como entre *P. brasiliensis* e *Blastomyces dermatitidis*, a região ITS apresentou-se significativamente variável entre as duas espécies. Apesar de relativamente conservada entre os diferentes isolados de *P. brasiliensis*, tal região mostrou algumas variações, como por exemplo, o isolado T10B1, que apresentou maior variabilidade em relação aos demais, tendo uma substituição no ITS1 e duas no ITS2 (Hebeler-Barbosa et al., 2003b).

Análises comparativas das sequências da região ITS e também do gene codificador da glicoproteína antigênica *GP43* confirmaram a similaridade entre os isolados clínicos e de tatus (Hebeler-Barbosa et al., 2003b), indicando que tanto o homem quanto os tatus estão

se infectando pelos mesmos “ecopatogenotipos” (Bagagli et al., 1998; Franco et al., 2000; Restrepo et al., 2000). A caracterização molecular dos isolados de tatu e sua comparação com os isolados clínicos pode ser de grande utilidade no mapeamento de genótipos do *P. brasiliensis* que estão causando infecções humanas.

Além da espécie *D. novemcinctus*, o fungo também foi isolado de um tatu *Cabassous centralis*, em Caldas, Colômbia, indicando que outras espécies além do tatu-galinha também podem estar infectadas com o patógeno (Corredor et al., 2005).

Recentemente foram relatados dois casos de PCM em cães (Ricci et al., 2004, Farias et al., 2005), sendo obtida cultura fúngica apenas para o segundo caso. O isolado foi caracterizado molecular e morfológicamente (Bosco et al., 2005). Este achado é de grande importância para o estudo epidemiológico da PCM, pois representa o primeiro isolamento de *P. brasiliensis* de um animal doméstico, cujo contato com humanos é infinitamente superior se comparado com animais silvestres.

***P. brasiliensis*: variabilidade intra-específica ou especiação críptica?**

Espécie, em biologia, é um termo difícil de definir. Existem diferentes conceitos, aplicados aos diferentes organismos. Segundo Ernest Mayr (1963), um dos mais influentes biólogos evolucionistas, espécie é uma linhagem única proveniente de um ancestral comum, que possui uma identidade genética e constitui uma unidade reprodutiva e ecológica e que tem suas próprias tendências evolutivas. Esta parece ser, e de fato é, uma ótima definição, mas como aplicá-la para reconhecer espécie, ou então para diferenciar espécies bem parecidas?

A forma de reconhecimento mais difundida é o morfológico. Ora, é só olhar para qualquer paisagem natural que somos capazes de identificar com clareza diferentes espécies de seres vivos. Mas, obviamente, quanto menor o organismo, mais difícil é achar diferenças fenotípicas para distinguí-las corretamente as espécies, sendo comum, no caso de fungos, a junção de indivíduos de diferentes espécies, quando o reconhecimento é apenas

baseado em características como pigmentação, forma e aspecto da colônia assexuada ou da estrutura reprodutiva, uso de diferentes substratos, entre outras.

Considerar espécie como grupo de indivíduos intercruzantes que originam descendentes férteis é o conceito mais difundido e usado para os mais diversos grupos de organismos. Porém, obviamente não é aplicável para populações microbianas assexuadas e mesmo em populações sexuadas, pode levar a um conflito entre fluxo gênico real e fluxo gênico potencial, uma vez que, após o isolamento geográfico (alopatria), muitas outras características podem divergir antes mesmo da perda do potencial de inter cruzamento (as espécies já estão isoladas e não há seleção para a formação de barreiras intrínsecas ao fluxo gênico) (Taylor et al., 2006).

Há ainda um conceito ecológico de espécie, o qual a define como um grupo de indivíduos que ocupam o mesmo nicho ecológico, sendo este o conjunto de interações com o meio biótico e abiótico bem como sua sobrevivência em diferentes condições e recursos proporcionados pelo ambiente (nicho ecológico, portanto não é o mesmo que habitat). Este conceito está intimamente conectado ao conceito biológico, uma vez que indivíduos que se reproduzem com outros de mesma adaptação ecológica possuem maior valor adaptativo. Sabe-se que a adaptabilidade dos híbridos pode, em muitos casos, ser menor do que a das espécies parentais (Ridley, 2006). A aplicabilidade deste conceito é inviabilizada pelo fato de que o nicho ecológico é, na verdade, n dimensional, sendo de difícil definição, ainda mais para microrganismos, para os quais envolve interações bioquímicas com o substrato, seja ele ambiental/saprobótico ou o próprio hospedeiro.

No caso de microrganismos, novas espécies vêm sendo descritas de acordo com o conceito filogenético, segundo o qual espécie é o menor grupo de organismos individuais derivado de um único ancestral comum (monofilia). Os membros de uma espécie compartilham apomorfias (características derivadas ou novidades evolutivas) não existentes em outras espécies. O reconhecimento filogenético de espécie detecta divergência genética através do sequenciamento de DNA de vários *loci* (*Multi Locus Sequencing Type*, MLST), determinando assim quais populações estão isoladas reprodutivamente (não há

recombinação entre alelos) pela concordância de genealogia de genes. Esta técnica também nos permite inferir o tipo de reprodução da espécie, se clonal ou sexuada segundo o índice de associação de alelos (este índice é baixo em populações sexuadas, uma vez que há maior recombinação). O uso de apenas um locus poderia considerar grupos de indivíduos com polimorfismo intra-específico como espécies diferentes, o que não seria correto, uma vez que toda espécie possui uma variabilidade genética intrínseca e compartilhada por seus membros (Taylor et al., 2000).

O uso de sequências de DNA para estudos filogenéticos é altamente informativo, principalmente no caso de microrganismos como fungos, para os quais existem poucos caracteres morfológicos devido à simplicidade e baixa diferenciação celular dos mesmos. Além disso, existem diversos modelos matemáticos disponíveis para o aprimoramento dos cálculos de distância genética, que vão desde modelos simples, como o de Jukes-Cantor, para o qual a taxa de substituição é similar para todas as regiões gênicas e é a mesma para todos os nucleotídeos, até modelos como o de Kimura-2-Parâmetros e Tajima-Nei que levam em consideração a diferença entre a frequência de transições e transversões e a diferença de frequência dos quatro nucleotídeos, respectivamente (Nei & Kumar, 2000). Assim, cada nucleotídeo é considerado uma característica diferente para a construção da árvore filogenética. Existem diferentes métodos de construção de árvores filogenéticas, sendo que nesta tese foram utilizados os métodos de Neighbor Joining (NJ), Máxima Parcimônia (MP), Máxima Verossimilhança (Maximum Likelihood, ML) e análise Bayesiana.

O método de NJ é um método de inferência filogenética por distância baseado no mesmo princípio do *Minimal Evolution* (ME), para o qual a soma de todos os comprimentos de ramos é computada para todas as topologias possíveis, sendo que aquela que possuir o menor comprimento é considerada a melhor árvore. Porém, o NJ é mais rápido por não examinar todas as topologias possíveis. Este método baseia-se no conceito de “vizinhos”, (dois taxa conectados por um único nó em uma árvore não enraizada). Inicialmente o método considera a árvore como uma grande politomia (todos os ramos saindo de um nó só), ou seja, parte de uma árvore “estrela”, cuja soma dos comprimentos dos ramos deverá

ser maior do que a soma dos ramos da árvore final. Como o método não considera mutações reversas ou mutações paralelas, realiza-se um teste estatístico para avaliar a confiabilidade da árvore obtida. Em geral, aplica-se o teste de *bootstrap*, que nada mais é do que a exclusão de um carácter e a duplicação de outro de forma aleatória e reconstrução da árvore. Comumente são usadas mil réplicas de *bootstrap*, sendo que quanto maior o valor (até 100), maior a confiabilidade dos clados da árvore (Nei & Kumar, 2000).

Inferências filogenéticas pelo método da MP buscam a topologia com o menor número de substituições que explique o processo evolutivo. Geralmente este método é indicado quando a divergência entre as sequências é baixa (distância menor ou igual a 0.1), a taxa de mutação é constante e muitos nucleotídeos serão avaliados. Este método usa apenas sítios informativos, os quais apresentam pelo menos dois nucleotídeos diferentes, cada um sendo representado por pelo menos dois taxa. A presença de caracteres homoplásicos, ou seja, derivados de mutações paralelas que surgem independentemente em ramos diferentes da árvore, prejudica a confiabilidade da árvore de MP, por esta razão deve-se calcular o índice de consistência da árvore (CI), que nada mais é que a razão entre o menor número possível de substituições e o menor número necessário de substituições. Portanto, quanto maior o CI menor o número de homoplasias da árvore e maior a confiabilidade da topologia (Nei & Kumar, 2000).

A ML por sua vez é um método estatístico bem estabelecido, conhecido por apresentar uma variância menor quando o n analisado é grande. Este método é muito utilizado para se estimar valores de recombinação e frequência alélica em genética clássica. O princípio básico do método de verossimilhança consiste em estimar a probabilidade, com base em um determinado modelo, de um conjunto de dados representar um processo que de fato ocorreu. Em termos de evolução de sequências de DNA, o método irá calcular a probabilidade de que as sequências em estudo tenham sido geradas seguindo as premissas do modelo evolutivo escolhido. O problema com a ML é a reconstrução da topologia da árvore. Métodos de ML não incluem parâmetros para

topologias. Dentre todas as árvores possíveis, a topologia ideal é simplesmente escolhida segundo os maiores valores de máxima probabilidade, supondo, portanto que a topologia com melhor estimativa de comprimento de ramos seja a árvore verdadeira. Isto nem sempre é verdadeiro, principalmente quando o número de substituições nucleotídicas varia muito entre os ramos, mesmo quando o n (número de nucleotídeos) é grande. (Pereira et al., 2001)

A análise Bayesiana, bastante usada em diversas análises estatísticas, começou a ser usada também em inferências filogenéticas nos últimos anos. Esta análise é baseada no valor conhecido como probabilidade posterior (PP) de uma árvore, segundo o Teorema de Bayes:

$$Pr[Tree | Data] = (Pr[Data | Tree] \times Pr[Tree]) / Pr[Data], \text{ onde:}$$

$$Pr[Tree | Data] = \text{Probabilidade posterior, pois depende dos dados;}$$

$$Pr[Data | Tree] = \text{Verossimilhança (probabilidade condicional dos dados dado a árvore)}$$

$$Pr[Tree] = \text{Probabilidade anterior da árvore sem levar em consideração os dados (todas possuem igual probabilidade anterior)}$$

$$Pr[Data] = \text{Probabilidade anterior dos dados (constante normalizadora)}$$

A probabilidade anterior é a probabilidade antes da análise, ou seja, todas as árvores são consideradas igualmente prováveis. A verossimilhança, num sentido estatístico, é uma versão inversa de probabilidade, ou seja, enquanto probabilidade nos permite prever um resultado não conhecido baseado em parâmetros conhecidos, a verossimilhança nos permite estimar parâmetros não conhecidos baseado em resultados conhecidos. No caso das análises filogenéticas, o “resultado conhecido” seriam as diferentes combinações de árvores. A PP de uma árvore pode ser interpretada como a probabilidade da mesma estar correta, assim, quanto maior este valor, melhor é a estimativa da filogenia. A PP envolve a análise de todas as árvores possíveis e da integração de todas as combinações possíveis de comprimentos dos ramos e valores de parâmetros de substituição nucleotídica. O valor de PP é calculado por um algoritmo

chamado *Markov chain Monte Carlo* (MCMC) que envolve basicamente dois passos: uma perturbação aleatória da árvore e sua posterior aceitação ou rejeição. Caso aceita, a árvore é sujeita a nova perturbação e assim sucessivamente. No fim, a proporção do número de vezes que uma árvore foi perturbada é a aproximação válida da sua probabilidade posterior (Huelsenbeck et al., 2001)

Métodos de análise filogenética usando sequências de DNA como caracteres vêm sendo extensivamente usados para análises populacionais em microrganismos, incluindo descrição de novas espécies. Usando estas ferramentas, trabalhos recentes demonstraram que a variabilidade genética do *P. brasiliensis* pode ir além de um simples polimorfismo intra-específico, caracterizando a presença de espécies crípticas, ou seja, espécies em divergência cujos caracteres fenotípicos até então conhecidos não são suficientes para distingui-las (Futuyama, 2002). Matute et al (2006a) analisaram sequências de DNA de oito regiões a partir de cinco genes nucleares codificadores: quitina sintase (*CHS2*), β -glucana sintase (*FKS*), fator adenil de ribosilação (*ARF*), α -tubulina (*TUB*) e *PbGP43* por análise de MP e análise Bayesiana. Foram estudados 65 isolados de *P. brasiliensis*, abrangendo seis áreas endêmicas de PCM. Através da análise por concordância de genealogia de genes foi possível detectar três espécies crípticas: S1, encontrada no Brasil, Argentina, Paraguai, Peru e Venezuela; PS2, encontrada no Brasil e Venezuela e PS3, encontrada apenas na Colômbia. S1 e PS2 são simpátricas, portanto a divergência genética entre as duas sugere a existência de uma barreira não geográfica ao fluxo gênico.

A espécie S1 parece estar mais amplamente distribuída, enquanto a espécie PS2 parece se distribuir preferencialmente na região sul/sudeste do Brasil. Devido à alta similaridade genética entre S1 e PS3, Matute et al (2006a) atribuíram a especiação do grupo PS3 a um possível evento de migração seguido de isolamento geográfico de isolados S1 para a Colômbia. Neste trabalho os autores não atribuíram à S1 o status de espécie filogenética, pois isolados PS2 e PS3 fazem dela um grupo parafilético. Isso porque em apenas duas, das oito árvores geradas (análise individual dos oito *loci* sequenciados no

trabalho), o agrupamento S1, PS2 e PS3 foi bem definido, nas demais foi possível observar ou alguns indivíduos da Colômbia (PS3) ou do PS2 agrupados com indivíduos S1.

Espécies crípticas também já haviam sido detectadas, por concordância de genealogias de genes, em outros fungos patogênicos e oportunistas, como, por exemplo, *Pneumocystis* (Aliouat-Denis et al., 2008), *Coccidioides immitis* (separado em duas espécies sexuadas, *C. immitis* e *C. posadasii*) (Koufopanou et al., 2001) e *Histoplasma capsulatum* (separado em pelo menos 7 espécies filogenéticas) (Kasuga et al., 1999, 2003).

O gênero *Pneumocystis* é composto por espécies de fungos não-cultiváveis, altamente diversificados e capazes de se fixarem em células alveolares do tipo I e se proliferarem pelos alvéolos pulmonares causando uma severa pneumonia. Primeiramente os microrganismos deste gênero eram considerados protozoários, porém com o advento da taxonomia molecular, provou-se que na verdade se tratavam de fungos verdadeiros (Ascomycota, Taphrinomycotina). Nos últimos anos a variabilidade destes patógenos foi melhor estudada, documentando-se uma significativa divergência genética entre cepas de *Pneumocystis* provenientes de diferentes mamíferos. Adicionalmente, experimentos de infecção cruzada comprovaram a associação entre padrão genético das cepas e hospedeiro. O reconhecimento filogenético de espécie pela análise simultânea de sete *loci* considerou estes grupos genéticos como cinco espécies distintas: *P. carinii* e *P. wakefieldiae* em ratos, *P. jirovecii* em humanos, *P. murina* em camundongos e *P. oryctolagi* em coelhos. Estas espécies também apresentam aspectos fenotípicos distintos (Aliouat-Denis et al., 2008; Keely & Stringer, 2005). O mais interessante a respeito da especiação deste grupo de fungos é o fato de existir uma concordância topológica entre a árvore filogenética das espécies de *Pneumocystis* e a árvore dos hospedeiros (Keely & Stringer, 2005). Essa especificidade por hospedeiro sugere que as espécies do gênero *Pneumocystis* tenham co-evoluído ou co-especiado juntamente com o hospedeiro. Esta hipótese foi explorada recentemente pela comparação da filogenia de hospedeiros mamíferos com a filogenia de *Pneumocystis* parasitas. O estudo mostrou correspondência entre as filogenias, concluindo que o principal evento de especiação em *Pneumocystis* foi o

de co-evolução (Guillot et al., 2001), sendo um evento simpátrico (mesma área geográfica) e nicho-específico.

Fungos do gênero *Coccidioides* são pertencentes à família Onygenaceae *sensu lato*, são dimórficos, ocorrendo na forma de micélio no solo e na forma de levedura nos hospedeiros (seres humanos, roedores e tatus). A infecção se dá via inalação de artroconídios produzidos assexuadamente no solo. Uma vez nos pulmões, estes esporos formam esférulas repletas de leveduras em divisão que posteriormente são liberadas atingindo diferentes órgãos e tecidos. Estas espécies se distribuem desde o norte da Califórnia até a Argentina e têm preferência por solos de ambientes áridos e semi-áridos, como o nordeste Brasileiro, onde é comum o isolamento do fungo de tatus e surtos de infecção humana após caçadas de tatus. Até recentemente todos os indivíduos isolados eram considerados de uma única espécie, *C. immitis*, até que este foi separado em duas espécies, *C. immitis* e *C. posadasii* segundo a concordância de três genealogias de genes (Koufopanou et al., 2001). Mais tarde, Fisher et al (2001), estudando microssatélites nestas duas espécies, observaram que elas estão divergindo a cerca de 12,8 milhões de anos, que o *C. immitis* ocorre somente na Califórnia e o *C. posadasii* nas demais regiões, sendo as populações encontradas na América do Sul mais homogêneas e semelhantes às populações do Texas, sendo provavelmente resultantes de migração de poucos exemplares deste local, o que causou um “efeito fundador” na população, resultando na baixa a variabilidade genética da população sul americana. Como a distância geográfica entre estas populações (isolados do Texas e da América do Sul) é grande, se comparada à distância genética, os autores propuseram que o transporte de indivíduos *C. posadasii* para a América do Sul tenha ocorrido via ameríndios ou pequenos roedores há cerca de 134 mil anos. Os autores ainda especulam que, as populações californianas (*C. immitis*) e não-californianas (*C. posadasii*) tenham se especiada alopaticamente devido a Montanha Tehachapi, barreira geográfica entre o Central Valley e o resto Sul da Califórnia.

Isolados fúngicos conhecidos como *H. capsulatum* são, na verdade, um complexo de espécies da família Ajellomycetaceae de distribuição geográfica ampla, ocorrendo nos

continentes americano, europeu, asiático e africano. Até pouco tempo atrás eram subdivididos em três variedades, a *duboisii* (Africana), a *capsulatum* (no Novo Mundo) e a *farciminosum* (no Velho Mundo) as duas primeiras variedades são patogênicas humanas e a terceira causa doença em equinos (Kwon-Chung & Bennett, 1992). Assim como o exemplo anterior, trata-se de um fungo termo dimórfico, cujos esporos estão em contato com o solo e fezes de aves e morcegos, hospedeiros voadores que contribuem para a ampla distribuição do patógeno. A infecção também se dá via inalação e após converter para a forma leveduriforme, no pulmão, o patógeno é disseminado via corrente sanguínea e linfática. Usando a genealogia de quatro genes nucleares codificadores de proteínas, Kasuga et al (1999; 2003) detectaram sete espécies filogenéticas mais uma espécie da Eurásia que emergiu do clado da América do Sul, grupo A. A divergência do *H. capsulatum* teve início entre 3 a 13 milhões de anos atrás, na América Latina, e os clados da África, Austrália e América do Norte são resultantes da dispersão inicial. A baixa variabilidade genética encontrada em alguns clados, como o da Austrália, por exemplo, foi justificada como uma possível herança de um evento *bottle neck* causado durante o período de glaciação do Pleistoceno, evento que diminuiu o habitat de um número considerável de espécies, principalmente em áreas temperadas. Uma interessante observação foi a presença de isolados típicos da variedade *farciminosum* em diferentes espécies crípticas do complexo *H. capsulatum*, indicando que *farciminosum* é um grupo polifilético e, portanto artificial, ou seja, a habilidade em se parasitar equinos evoluiu independentemente, em diferentes linhagens, constituindo um exemplo de homoplasia (convergência evolutiva ou paralelismo).

A especiação do gênero *Paracoccidoides* é ainda controversa entre os diferentes pesquisadores da área. Para alguns parece estar claro que existem quatro espécies, para outros, o que chamamos de S1, PS2 e PS3 seriam na verdade variantes geográficas de *P. brasiliensis*, assim, existiriam apenas duas espécies, *P. brasiliensis* e *P. lutzii*. Muitos autores reforçam esta idéia pelo fato de que o nível de divergência entre *P. lutzii* e os demais grupos genéticos é consideravelmente maior do que a divergência existente dentro

do complexo S1, PS2 e PS3. Teixeira et al (2009) evidenciaram a existência da espécie *P. lutzii*, descrita neste ano (Teixeira et al., não publicado), mas na época conhecida como *Pb01-like*, pela genealogia dos mesmos oito *loci* usados por Matute et al (2006a) na identificação dos grupos S1, PS2, PS3. Os autores realizaram análises filogenéticas das sequências isoladas e também concatenadas. Como o clado *Pb01-like* apareceu em todas as filogenias independentes e também na filogenia de *multi loci*, os autores o consideraram como uma nova espécie. Além dos oito *loci*, os autores também avaliaram a região rRNA (ITS1, 5.8S e ITS2). Apesar de diferenciar claramente o clado *Pb01-like* dos demais isolados, esta região não permitiu distinção entre S1, PS2 e PS3. Aliado a estes dados, *Pb01-like* também se mostrou isolado reprodutivamente das demais espécies pela análise de *Split decomposition* e apresentou-se como a espécie mais divergente do gênero *Paracoccidioides*. Tal divergência pode ser evidenciada em outros trabalhos, usando-se outros genes nucleares codificadores de proteínas, como *CH4*, *Actin*, *ODC*, *URA3*, *FKS1*, *HSP70*, *Hydrophobin*, *Kex*, *Catalase A*, *Catalase P*, *Formamidase* and *Glyoxalase* (Carrero et al., 2008). Além destes genes, o *intein* PRP8 também foi usado em análises filogenéticas em *Paracoccidioides* e separou com altos valores de *bootstrap* as quatro espécies, sendo que, mais uma vez, a *Pb01-like* apresentou maior grau de divergência comparado com os demais (Theodoro et al., 2008b, artigo 1 do capítulo 2 desta tese).

Neste ano, dados adicionais também foram publicados envolvendo isolados das diferentes espécies. Marini et al (2010) realizaram um trabalho inédito de busca por elementos genéticos transponíveis (transposons) no genoma das espécies S1, PS2 e *P. lutzii*. Os autores encontraram pelo menos oito famílias de Tc1/mariner (A-H), referidas como Trem (**T**ransposable **e**lement **m**ariner), sendo que a distribuição destes elementos varia segundo a espécie. Assim, Trem C e H estão presentes em S1, PS2 e PS3, Trem A, B e F estão presentes em S1 e PS2, mas ausentes em *P. lutzii*, enquanto Trem D e E são específicos para S1 e *P. lutzii*, respectivamente, estando ausentes em PS2.

Em outro trabalho, recentemente publicado, cinco genes mitocondriais (*COB*, *COX3*, *RNS*, *RNL* e *ATP6*) foram sequenciados e analisados por concordância genealógica para

isolados do gênero *Paracoccidioides* pertencentes às espécies S1, PS2, PS3 e *P. lutzii* e para os grupos externos *Epidermophyton floccosum*, *Penicillium marneffeii* e *Aspergillus niger*. As inferências filogenéticas foram realizadas por MP e análise Bayesiana e corroboraram a origem monofilética de PS2 e PS3 e a parafilia de S1. Adicionalmente os autores propuseram uma subespécie críptica (CS1), clado composto por isolados venezuelanos. Os autores também consideraram o grau de diferenciação genética segundo o comprimento dos ramos, calculado a partir das árvores não enraizadas. Como a divergência genética é significativamente maior entre *P. lutzii* e os demais grupos genéticos e os isolados PS2, PS3 e CS1 tendem a se agrupar geograficamente, os autores consideraram que o gênero *Paracoccidioides* é composto apenas por duas espécies: *P. lutzii* e *P. brasiliensis*, cujos grupos genéticos reprodutivamente isolados representam, na verdade, variantes geográficas e não espécies distintas. Além disso, os autores afirmam que se PS2, PS3 e CS1 fossem considerados espécies, S1 seria ou parafilético e isso não seria válido para propostas taxonômicas (Salgado-Salazar et al., 2010).

E agora? O que considerar espécie e o que considerar variante geográfico? Ora, o fato destas espécies possuírem uma tendência de agrupamento geográfico reforça a idéia de isolamento genético já evidenciado por sequenciamento *multi loci*. Chamar estes agrupamentos de espécie ou variantes geográficas não altera o fato de que os mesmos não estão trocando alelos na natureza. Na verdade, a ausência de fluxo gênico caracteriza espécie segundo o conceito filogenético, extensivamente discutido por Taylor (2000) em sua revisão sobre conceito e reconhecimento de espécie. Considerar divergência genética para inferir espécie parece ser um tanto arbitrário, uma vez que distância é relativo, ou seja, caso uma ou mais espécies novas no gênero *Paracoccidioides* forem descobertas e apresentarem uma divergência maior em relação às espécies atualmente conhecidas, *P. brasiliensis* e *P. lutzii* não seriam mais “tão” diferentes. Este é um assunto ainda muito pouco discutido no ambiente acadêmico, porém de grande importância biológica.

Entender melhor a diversidade de isolados do gênero *Paracoccidioides* e definir espécies ou mesmo simples variantes geográficas significa compreender de forma mais

abrangente a dinâmica populacional, as diferentes formas clínicas e a importância de se considerar o *Paracoccidioides* não como uma única cepa, mas sim como um conjunto complexo de genótipos que devem ser levados em consideração nas medidas terapêuticas. Uma vez que estas espécies estão em divergência genética, é de se esperar que isso seja traduzido em diferentes estratégias de interação com o ambiente, o que poderia resultar em maior ou menor produção de esporos assexuados no meio saprobiótico, diferentes temperaturas de transição micélio-levedura, diferentes aspectos clínicos, de diagnóstico e tratamento, etc.

Um exemplo claro da importância prática do reconhecimento de espécies em *Paracoccidioides* é o não reconhecimento de antígenos produzidos em um local por soros de pacientes provenientes de locais diferentes. Em um estudo recente desenvolvido por Batista Jr et al (2009), duas cepas de *P. brasiliensis* provenientes de duas áreas geográficas distintas foram comparadas em termos da sensibilidade e especificidade em relação ao diagnóstico da PCM. Exoantígenos do isolado 550B do Mato-Grosso e do isolado B-339 de São Paulo (usado como antígeno de referência) foram comparados em testes de imunodifusão com soros de pacientes do Mato-Grosso e de São Paulo. Antígenos e soros provenientes do mesmo estado apresentaram de 90 a 100% de positividade nos testes de imunodifusão, porém, testes cruzados (antígeno de um estado cruzado com soro de pacientes do outro estado), apresentaram de 26 a 42% de positividade. Estes resultados sugerem que as diferenças na composição antigênica estão provavelmente relacionadas às peculiaridades filogenéticas dos isolados de *P. brasiliensis* da região centro-oeste, hoje denominados *P. lutzii*. Entretanto, os autores apenas supõem que o isolado 550B seja pertencente a mais nova espécie descrita, *P. lutzii*, sendo ainda necessários estudos sorológicos mais abrangentes, com isolados devidamente identificados, pertencentes às quatro espécies (ou variantes geográficas), uma vez que quanto mais cedo for o correto diagnóstico, maior a chance de se obter sucesso no tratamento da PCM.

Identificando novos isolados

Nos trabalhos comentados anteriormente, os autores usaram sequenciamento e análise filogenética de várias regiões codificadoras. Entretanto, adquirir sequências de vários *loci* com variação genética suficiente para detectar espécies filogenéticas pode ser extremamente trabalhoso, tornando o uso de marcadores hipervariáveis atrativo para um rápido reconhecimento de isolados pertencentes às diferentes espécies crípticas (Taylor et al., 2000).

Técnicas moleculares como RFLP, RAPD, pesquisa de microssatélites e estudo de regiões variáveis como a de rDNA são promissoras para a busca de marcadores moleculares a fim de diferenciar as espécies do gênero *Paracoccidioides*.

O uso de RAPD em nosso laboratório (Hebeler-Barbosa et al., 2003a) permitiu a separação dos isolados de *P. brasiliensis* em dois grupos com cerca de 70% de similaridade genética. Em um grupo estão os isolados T10B1 e Bt84 e no outro estão os demais isolados. Esta separação foi confirmada também pela análise filogenética que constatou a existência das espécies crípticas, na qual os isolados T10 e Bt84 pertencem à espécie PS2 e os demais isolados à espécie S1 (Matute et al., 2006a).

Regiões variáveis e de múltiplas cópias no genoma, como as de rDNA (codificador do RNA ribossômico), são de grande valia para estudos filogenéticos pois estão presentes em todos os seres vivos e acumulam mutações tanto em baixa (regiões 18S, 5.8S e 28S) como alta (regiões ITS1 e ITS2) frequência (White et al., 1990, Guarro et al., 1999), sendo também utilizadas para a detecção e identificação de fungos patogênicos (Henry et al., 2000).

A região de rDNA também permite a diferenciação de isolados, isso pôde, por exemplo, ser evidenciado em fungos dermatófitos (Jackson et al., 1999) e também em isolados de *P. brasiliensis* (Hebeler-Barbosa et al., 2003b). Porém, como já mencionado, tal

região não distingui claramente as espécies S1, PS2 e PS3 do complexo *P. brasiliensis* (Teixeira et al., 2009)

Outra opção para a detecção de marcadores moleculares são os microsatélites ou repetições de sequências simples (SSRs) que são repetições em *tanden* de fragmentos de DNA compostos por 1 a 6 pb de extensão. Trata-se de um marcador genético codominante, com polimorfismo neutro e de alta reprodutibilidade e especificidade (Dettman & Taylor, 2004). Erros de pareamento durante a replicação do DNA são considerados os principais mecanismos mutacionais para os SSRs (Schlotterer & Tautz, 1992). Durante a replicação, a cadeia de DNA que está sendo polimerizada pode se dissociar da fita molde e, devido à natureza repetitiva dos SSRs, as fitas podem se re-anelar erroneamente. Assim, ao final da replicação, a fita de DNA recém polimerizada pode conter um número de repetições diferente da fita molde. Este tipo de mutação pode gerar uma homoplasia de tamanho (Dettman & Taylor, 2004). Este é um fator complicador, pois altas taxas de mutação nos *loci* microsatélites (10^{-2} a 10^{-5}) resultam em polimorfismo de tamanho destas regiões, o que pode contrabalancear os efeitos da deriva genética. Assim, alelos que foram perdidos por deriva podem reaparecer na população, resultando em uma homoplasia. Portanto a existência de alelos que são idênticos em tamanho pode não representar um carácter homólogo (de mesma origem, em um ancestral comum), mas sim um carácter homoplásico (Fisher et al., 2000a). Existem modelos mutacionais de SSRs que levam em consideração a ocorrência de homoplasia, sendo necessária a determinação de um modelo mais apropriado para inferir distâncias genéticas.

Aplicações recentes da análise de polimorfismo de microsatélites se mostraram bem sucedidas, como no caso dos fungos *Penicillium marneffe* (Fisher et al., 2004, Lascker & Ran, 2004), *H. capsulatum* (Carter et al., 2001) e *C.immitis* (Fisher et al., 2000a) por corroborarem estudos prévios de genealogia de *multi loci*.

Em *P. brasiliensis* foram selecionadas algumas regiões de microsatélites para uso como marcadores moleculares para estudos clínicos e ambientais (Nascimento et al., 2004). Apesar de não ter sido observada nenhuma correlação entre as diferentes formas

clínicas dos isolados e os quatro SSRs selecionados, foi possível detectar altas taxas de polimorfismo nestas regiões, sugerindo seu potencial uso para estudos epidemiológicos e filogenéticos.

Matute et al (2006b) avaliaram cinco regiões de microssatélites para o correto diagnóstico de isolados pertencentes às espécies crípticas S1, PS2 e PS3. Apesar de não ter sido possível o uso individual de nenhum dos microssatélites para a distinção entre as espécies, o uso dos mesmos em pares foi bem sucedido, sendo escolhido, dentre as dez combinações possíveis, três pares de microssatélites para a separação dos isolados pertencentes às espécies crípticas distintas. Ao todo foram identificados 43 genótipos provenientes da análise dos cinco microssatélites, um deles foi encontrado em alta frequência na espécie PS3 (57,2%), indicando variabilidade genética, devido à reprodução clonal detectada neste grupo.

O estudo de SNPs (Single Nucleotide Polimorphism) também parece ser promissor para a separação dos isolados de *Paracoccidioides* de espécies diferentes, bem como para o desenvolvimento de uma técnica mais simples e pontual para a identificação de isolados recentes provenientes de tatus e pacientes humanos. SNPs são polimorfismos de uma única base em uma localização física exata. Uma vez que já se conhece a sequência de nucleotídeos de oito regiões nucleares codificadoras para um número de isolados representativos das três espécies (Matute et al., 2006a), é possível identificar SNPs que, analisados simultaneamente, poderão ser úteis para a diferenciação de isolados.

Em *C. immitis*, SNPs vêm sendo utilizados em estudos populacionais que visam entender a estrutura reprodutiva deste patógeno, bem como diferenciar isolados ambientais e humanos e caracterizar isolados recentes (Fisher et al., 1999 e 2000b, Sheff et al., 2010). Em *Mycobacterium tuberculosis* foi realizado um estudo filogenético com 212 SNPs, envolvendo cepas de todo o mundo, sendo possível inferir a história evolutiva deste patógeno no que diz respeito ao local de origem e disseminação (Filliol et al., 2006).

Nesta tese, SNPs de três genes nucleares (*ARF*, *GP43* e intein *PRP8*) foram detectados por SNaPshot® e/ou por Real Time PCR (genotipagem pela técnica de

TaqMan[®]) para a identificação de espécies do gênero *Paracoccidoides*. O kit Snapshot[®] Multiplex da Applied Biosystems foi utilizado para as reações de minissequenciamento (método dideoxi com extensão de uma única base). Uma vez que todos os nucleotídeos na reação são dideoxi, utiliza-se um *primer* sense localizado logo na posição *upstream*, flanqueando o SNP a ser detectado, de modo que o único nucleotídeo a ser incorporado corresponda ao SNP e seja identificado de acordo com a fluorescência emitida e lida durante a corrida dos “mini-amplicons” em gel de poliacrilamida no aparelho ABI 377 (Applied Biosystems). A genotipagem por Real Time, usando sondas TaqMan[®] (Applied Biosystems) específicas para cada SNP, trata-se de uma reação comum de PCR, com *primers* sense e antisense que amplificam a região contendo o SNP, mais sondas específicas para cada SNP, contendo a molécula *reporter* na ponta 5' e o *quencher* na 3'. Enquanto a sonda estiver íntegra o *quencher* inibe a emissão de fluorescência do fluoróforo. Uma vez anelada ao sítio contendo o SNP, a sonda é destruída pela atividade de exonuclease da enzima DNA Polimerase usada para a amplificação do fragmento, liberando o *reporter* do *quencher*, permitindo assim a emissão da fluorescência cuja cor será específica para cada SNP. Este método é menos trabalhoso e mais rápido, pois diferentemente da técnica de SNaPshot, dispensa a preparação do gel de Poliacrilamida e pode ser feito diretamente com o DNA extraído, sem necessidade de PCR prévio do gene que contém o SNP.

Além do DNA ribossômico e dos SNPs, o intein PRP8, um elemento genético parasita localizado no gene *PRP8* é também uma região promissora para estudos filogenéticos e epidemiológicos (Butler et al., 2001 e Butler & Poulter, 2005). Para uma descrição mais detalhada sobre biologia, evolução e aplicabilidade biotecnológica dos inteins, veja revisão bibliográfica do capítulo 2 desta tese.

Além de marcadores moleculares, é também de grande importância o estudo de marcadores morfológicos para uma melhor interpretação das diferentes interações que este patógeno pode ter com seu ambiente. Se as espécies estão separadas geneticamente é de se esperar que com o passar do tempo esta divergência seja detectada por diferenças

morfológicas que se traduzem em exploração de diferentes nichos ecológicos, tanto quando em saprobiose, como quando em parasitismo. Sabe-se que nas espécies de *H. capsulatum* pode-se detectar diferenças morfológicas e fisiológicas entre isolados pertencentes às diferentes espécies (Kasuga et al., 2003). No caso do *Coccidioides* spp, observou-se que a espécie *C. posadasii* cresce mais lentamente em meio contendo alta concentração de sais quando comparada com a espécie *C. immitis*, porém este fenótipo não é usado como diagnóstico (Fisher et al., 2002).

No complexo de espécies *P. brasiliensis* e *P. lutzii*, existem muitos fenótipos variáveis, como crescimento, aspecto de colônia miceliana, produção de conídios, transição micélio/levedura, microscopia leveduriforme (brotamentos, forma e tamanho das células, etc), virulência, termotolerância, entre outros (Theodoro et al., 2005b).

Theodoro et al. (2008a), estudando caracteres morfológicos envolvidos no dimorfismo micélio-levedura em associação com a expressão do gene *hsp70*, sugeriram o uso de alguns marcadores fenotípicos para a diferenciação entre as espécies de *P. brasiliensis*. Neste estudo, o sequenciamento parcial do gene *hsp70* permitiu a separação dos isolados em dois grupos, correspondentes às duas espécies crípticas simpátricas (S1 e PS2) que ocorrem na região hiper endêmica de Botucatu, SP. Além disso, foram observadas leveduras alongadas nos isolados Bt84 e T10 (pertencentes à espécie PS2), enquanto os demais isolados (pertencentes à espécie S1) apresentaram células arredondadas. A produção de conídios também parece ser um diferencial entre membros de espécies diferentes. Os isolados Bt84 e T10 não são capazes de produzir conídios em meio Agar extrato de solo, enquanto os demais isolados parecem ter uma produção exuberante neste meio (Terçarioli et al., 2007).

Os dados até então obtidos indicam que estes genótipos do grupo PS2 devem ocorrer de forma bem menos freqüente que os genótipos do grupo S1, tanto em hospedeiros humanos como nos tatus (Matute et al., 2006a). Particularmente em nossos isolados de tatus, todos provenientes de uma mesma região hiper endêmica, esta proporção é de 1:9. Esta menor freqüência de isolamento de *P. brasiliensis* da espécie PS2

poderia ser explicada pela menor capacidade ou pela não produção de conídias infectantes neste grupo? Estes genótipos poderiam estar associados aos eventuais casos de infecção pela via traumática e não inalatória como na maioria das infecções por PCM?

Com o objetivo de elucidar questões básicas como as descritas acima, caracteres morfológicos das fases de micélio e levedura foram aqui avaliados em isolados pertencentes às diferentes espécies do gênero *Paracoccidioides*. Além disso, marcadores moleculares como sequenciamento do *intein* PRP8 e análise de SNPs dos genes *GP43* e *ARF*, previamente sequenciados por Matute et al (2006) e Teixeira et al (2009) e também do *intein* PRP8, sequenciado nesta tese (Theodoro et al., 2008b, artigo 1, capítulo 2), foram realizadas com a finalidade de se identificar novos isolados, provenientes de tatus ou de pacientes humanos.

Com intuito de se compreender melhor a história biogeográfica do gênero, dados filogenéticos e de localização geográfica dos isolados foram analisados nos programas Nested Clade Analysis (NCA) e DIVA (Divergence-Vicariance Analysis). Além disso, foi estimado o tempo de divergência entre as espécies segundo a taxa de substituição do gene *ARF*, previamente calculada por Kasuga et al (2003).

Os principais processos considerados pela biogeografia histórica são extinções, expansões geográficas e dois tipos de especiação alopátrica: vicariância e dispersão (Futuyama, 2002). A especiação por vicariância ocorre após o aparecimento da barreira geográfica, enquanto a especiação por dispersão ocorre segundo uma ativa colonização através de uma barreira pré-existente. Portanto, na vicariância a colonização supostamente ocorreu antes da formação da barreira e na dispersão, depois. Em geral, o termo expansão geográfica é usado para colonizações que favorecem a ampla distribuição da espécie e dispersão se refere às colonizações que resultam em especiação (Kodandaramaiah, 2009).

O NCA é um método que combina informações filogenéticas provenientes das sequências de DNA com informações geográficas, ou seja, com os locais de onde estas sequências foram obtidas. Resumidamente, redes haplotípicas são geradas (um haplótipo é o conjunto de sequências com similaridade entre si maior ou igual a 99%) para dados de

um único locus e a relação entre a distribuição geográfica e os haplótipos contem informações a respeito da história biogeográfica destas populações no que diz respeito a possíveis migrações e isolamentos reprodutivos, por exemplo (Panchal, 2007; Clemente et al., 2000 e Posada et al., 2000).

O DIVA é um programa simples para reconstrução da distribuição ancestral de uma filogenia, usando análise de dispersão-vicariância, um método no qual a distribuição ancestral é inferida sem a suposição prévia da existência de padrões biogeográficos gerais. Portanto, o DIVA é útil na reconstrução da história de distribuição de um grupo de organismos na ausência de uma hipótese biogeográfica dos taxa. Para usar o programa é essencial fornecer uma árvore filogenética verdadeira e as áreas de ocorrência de cada taxon. Como resultado o programa apresenta uma árvore de área, na qual os nós representam áreas ancestrais (Ronquist, 1996. DIVA version 1.1). Este método estabelece um “custo” de “1” para colonizações e extinções e nenhum “custo” para eventos de vicariância, favorecendo hipóteses que maximizem os eventos de vicariância, uma vez que qualquer hipótese de origem de uma espécie pode ser explicada por diversas dispersões (Kodandaramaiah, 2009). Este programa vem sendo muito usado para espécies de animais e plantas e este é o primeiro trabalho no qual foi aplicado para fungos patogênicos humanos.

O uso do NCA e do relógio molecular em associação com a análise de divergência-vicariância do DIVA foi importante, pois este programa não distingue perfeitamente entre os eventos de vicariância e dispersão, pois não considera *a priori* os tempos de divergência relativos a formação de barreiras geográficas.

A metodologia, os principais resultados e a interpretação dos mesmos estão descritos no artigo 1 deste capítulo.

Referências Bibliográficas

- Albornoz, M.B., 1971. Isolation of *Paracoccidioides brasiliensis* from rural soil in Venezuela. *Sabouraudia* 9, 248-253.
- Aliouat-Denis, C.M., Chabé, M., Demanche, C., Aliouat E.M., Viscogliosi, E., Guillot, J., Delhaes, L., Dei-Cas, E., 2008. Pneumocystis species, co-evolution and pathogenic power. *Infec. Genet. Evol.* 8, 708-726.
- Bagagli, E., Sano, A., Coelho, K.I.R., Alquati, S., Miyaji, M., Camargo, Z.P., Gomes, G., Franco, M., Montenegro, M.R., 1998. Isolation of *Paracoccidioides brasiliensis* from armadillos (*Dasypus novemcinctus*) captured in an endemic area of paracoccidioidomycosis. *Am J Trop Med & Hyg.* 58, 505-512.
- Bagagli, E., Franco, M., Bosco, S. M. G., Hebelers-Barbosa, F., Trinca, L. A., Montenegro, M. R., 2003. High frequency of *Paracoccidioides brasiliensis* infection in armadillo (*Dasypus novemcinctus*): an ecological study. *Med. Mycol.* 41, 217-223.
- Batista Jr, J., Camargo, Z.P., Fernandes, G.F., Vicentini, A.P., Fontes C.J.F., Hahn, R.C., 2008. Is the geographical origin of a *Paracoccidioides brasiliensis* isolate important for antigen production for regional diagnosis of paracoccidioidomycosis? *Mycosis* 53, 176-180.
- Bosco, S.M.G., Theodoro, R.C., Macoris, S.A.G., Farias, M.R., Muro, M., Ribeiro, M. G., Bagagli, E., 2005. Morphological and molecular characterization of the first isolate of *Paracoccidioides brasiliensis* from dog (*Canis familiaris*). *Rev. Inst. Med. Trop. São Paulo* 47 (14), 62-63.
- Butler, M.L., Goodwin, T.J.D., Poulter, R.T.M., 2001. A nuclear-encoded intein in the fungal pathogen *Cryptococcus neoformans*. *Yeast* 18:1365-1370.
- Butler, M.L., Poulter, R.T.M., 2005. The PRP8 inteins in *Cryptococcus* are a source of phylogenetic and epidemiological information. *Fungal. Genet. Biol.* 42, 452-463.
- Carrero LL, Niño-Vega G, Teixeira MM, Carvalho MJA, Soares CMA, Pereira M, Jesuino RSA, McEween JG, Mendoza L, Taylor J W, Felipe MS, San-Blas G 2008. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. *Fungal Genet. Biol.* 45(5), 605-612.
- Carter, D.A., Taylor, J.W. Dechairo, B., Burt, A., Koenig, G.L., White, T.J., 2001. Amplified Single-Nucleotide Polymorphisms and a (GA)_n Microsatellite Marker Reveal Genetic Differentiation between Populations of *Histoplasma capsulatum* from the Americas. *Fungal Gen. Biol.* 34(1), 37-48.
- Clement, M. Posada, D., Crandall, K.A., 2000. TCS: A computer program to estimate genes genealogies. *Mol. Ecol.* 9(10), 1657-1659.
- Corredor, G.G, Peralta, L.A., Castaño, J.H., Zuluaga, J.S., Henao, B., Arango, M., Tabares, A.M., Matute, D.R., McEwen, J.G., Restrepo, A., 2005. The naked-tailed armadillo *Cabassous centralis* (Miller 1899): a new host to *Paracoccidioides brasiliensis*. *Med. Mycol.* 43(3), 275-280.
- Corredor, G. G., Castaño, J. H., Peralta, A., Díez, S., Arango, M., McEwen, J., Restrepo, A., 1999. Isolation of *Paracoccidioides brasiliensis* from the nine-banded armadillo

- Dasyopus novemcinctus*, in an endemic area for paracoccidioidomycosis in Colombia. *Rev. Iberoam. Micol.* 16(4), 216-220.
- Corredor, G.G., Peralta, L.A., Castaño, J.H., Zuluaga, J.S., Henao, B., Arango, M., Tabares, A.M., Matute, D.R., McEwen, J.G., Restrepo, A., 2005. The naked-tailed armadillo *Cabassous centralis* (Miller 1899): a new host to *Paracoccidioides brasiliensis*. Molecular identification of the isolate. *Med. Mycol.* 43(3), 275-280.
- Dettman, J.R., Taylor, J.W., 2004. Mutation and evolution of microsatellite *loci* in *Neurospora*. *Genetics* 168, 1231-1248.
- Farias, M.R., Werner, J., Muro, M. D., Marques, S.A., Marques, M.E.A., Franco, M.F., Ribeiro, M.G., Custodio, C.C., Condas, L.A.Z., Bosco, S.M.G., Bagagli, E., 2005. Canine paracoccidioidomycosis: case report of generalized lymphadenitis. *Rev. Inst. Med. Trop.* 47(14), 64.
- Ferreira, M. S.; Freitas, L. S.; Lacaz, C. S.; Del Negro, G. M.; Aiello, N. T.; Garcia, M. N.; Assis, C. M.; Salebian, A.; Heris-Vaccari, E. M., 1990. Isolation and characterization of a *Paracoccidioides brasiliensis* strain from dog food probably contaminated with soil in Uberlândia, Brazil. *J. Med. Vet. Mycol.* 38, 253-256.
- Filliol, I., Motiwala, A.S., Cavatore, M. QI, W. et al., 2006. Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set. *J. Bacteriol.* 188, 759-772.
- Fisher, M.C., Aenensen, D., Hoog, S., Vanittanakom, N., 2004. Multilocus microsatellite typing system for *Penicillium marneffei* reveals spatially structured populations. *J. Clin. Microbiol.* 42, 5065-5069.
- Fisher, M.C., Koenig, G.L., White, T.J., Taylor, J.W., 2000a. A test for concordance between the multilocus genealogies of genes and microsatellites in the pathogenic fungus *Coccidioides immitis*. *Mol. Biol. Evol.* 17, 1164-1174.
- Fisher, M.C., Koenig, G.L., White, T.J., Taylor, J.W., 2000b. Pathogenic Clones versus Environmentally Driven Population Increase: Analysis of an Epidemic of the Human Fungal Pathogen *Coccidioides immitis*. *J. Clin. Microbiol.* 38, 807-813.
- Fisher, M.C., Koenig, G.L., White, T.J., Taylor, J.W., 2002. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. *Mycologia*, 94 (1), 73-84.
- Fisher, M.C., Koenig, G.L., White, T.J., San Blas, G., Negroni, R., Alvarez, I.G., Wanke, B., Taylor, J.W., 2001. Biogeographic range expansion into South America by *Coccidioides immitis* mirrors New World patterns of human migration. *Proc. Natl. Acad. Sci. U. S. A* 98, 4558-4562.
- Fisher, M.C., White, T.J., Taylor, J.W., 1999. Primers for genotyping single nucleotide polymorphisms and microsatellites in the pathogenic fungus *Coccidioides immitis*. *Mol. Ecol.* 8, 1075-1092.
- Franco, M., Bagagli, E., Scapolio, S., Lacaz, C.S., 2000. A critical analysis of isolation of *Paracoccidioides brasiliensis* from soil. *Med. Mycol.* 38, 185-191.

- Franco, M., 1987. Host-parasite relationships in paracoccidioidomycosis. *J. Med. Vet. Mycol.* 25, 5-18.
- Futuyama, D.J., 2002. *Biologia Evolutiva*. 2 ed. Ribeirão Preto. Funpec, 117p.
- Gesuele, E., 1989. Aislamento de *Paracoccidioides* sp. De heces de pinguino de la Antártida. In: International meeting on paracoccidioidomycosis, Proceedings Caracas, Abstract B2.
- Grose, E.; Tamsitt, J. R., 1965. *Paracoccidioides brasiliensis* recovered from intestinal tract of three bats (*Artibeus lituratus*) in Colombia. *Sabouraudia* 4, 124-125.
- Guarro, J., Gené, J., Stchigel., A.M., 1999. Developments in fungal taxonomy. *Clin. Microbiol. Rev.* 12, 454-500.
- Guillot, J., Demanche, C., Hugot, J.P., Berthelemy, M., Wakefield, A.E., Dei-Cas, E., Chermette, R., 2001. Parallel phylogenies of *Pneumocystis* species and their mammalian hosts. *J. Eukaryot. Microbiol.* 47, S113-S115.
- Hebeler-barbosa, F., Montenegro, M.R., Bagagli, E., 2003a. Virulence profiles of ten *Paracoccidioides brasiliensis* isolates obtained from armadillos (*Dasypus novemcinctus*). *Med. Mycol.* 41, p.89-96.
- Hebeler-Barbosa, F., Morais, F. V., Montenegro, M. R., Kuramae, E. E., Taylor, J. W., Montes, B., McEwen, J. G., Puccia, R., Bagagli, E., 2003b. Sequence comparison of the internal transcribed spacer regions and gp 43 in *Paracoccidioides brasiliensis* for patients and armadillos *Dasypus novemcinctus*. *J. Clin. Microbiol.* 41, 5735-5737.
- Henry, T., Iwen, P.C., Hinrichs, S.H., 2000 Identification of *Aspergillus* Species Using Internal Transcribed Spacer Regions 1 and 2. *J. Clin. Microbiol.* 38, 1510-1515.
- Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian Inference of Phylogeny and its impact on Evolutionary Biology. *Science* 294, 2310-2314.
- Jackson, C.J., Barton, R.C., Evans, E.G., 1999. Species Identification And Strain Differentiation Of Dermatophyte Fungi By Analysis Of Ribosomal-Dna Intergenic Spacer Regions. *J. Clin. Microbiol.* 37, 931-936.
- Kasuga, T., Taylor, J.W., White, T.J., 1999. Phylogenetics relationships of varieties and geographical groups of the human pathogenic fungus *Histoplasma capsulatum* darling. *J. Clin. Microbiol.* 37, 653-663.
- Kasuga, T., White, T.J., Koenig, G., McEwen, J., Restrepo, A., Castañeda E., Lacaz, C.S., Heins-Vaccari, E.M., Freitas R.S., Zancopé-Oliveira, R.M., Qin, Z., Negroni, R., Carter, D.A., Mikami, Y., Tamura, M., Taylor, M.L., Miller, G.F., Poonwan, N., Taylor, J.W., 2003. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol. Ecol.* 12, 3383-3401.
- Keely, S.P., Stringer, J.R., 2005. Nomenclature and genetic variation of *Pneumocystis*. In: Walzer, P.D., Cushion, M.T. (Eds), *Pneumocystis Pneumonia*, third ed. Marcel Dekker, New York, pp. 39-59.
- Kodandaramaiah, U. 2010. Use of dispersal-vicariance analysis in biogeography – a critique. *J. Biogeogr.* 37, 3-11.

- Koufopanou, V., Burt, A., Szaro, T., Taylor, J.W., 2001. Gene genealogies, cryptic species, and molecular evolution in the human pathogen *Coccidioides immitis* and relatives (Ascomycota, Onygenales). *Mol. Biol. Evol.* 18, 1246-1258.
- Kwon-Chung, K.J., Bennett, J.E. 1992. *Medical Mycology*. Lea & Febiger, Malvern, PA.
- Lacaz, C.S., Porto, E., Martins, J.E.C., 1984. Paracoccidioidomicose In: LACAZ, C.S., PORTO, E., Martins, J.E.C. *Micologia médica*. São Paulo: Sarvie, 189-216p.
- Lascker, B.A, Ran, Y., 2004. Analysis Of Polymorphic Microsatellite Markers For Typing *Penicillium Marneffei* Isolates. *J. Clin. Microbiol.* 42, 1483-1490.
- Loughry, W.J., McDonough, C.M., 1998. Spatial patterns in a population of nine-banded armadillos (*Dasyus novemcinctus*). *Am. Midl. Nat.* 140, 161-168.
- Macedo, R.C.L., Lazera, M.S., Trilles, L., Bulcão, A.S., Silva Jr, N.J., Oliveira, N.A., Wanke, B., 1999. Paracoccidioides brasiliensis- In: VII International Meeting On Paracoccidioidomycosis, Campos do Jordão. Proceedings São Paulo: Brasil, p.128.
- Matute, D.R., McEwen, J.G., Montes, B.A, San-Blas, G., Bagagli, E., Rauscher, J.T., Restrepo, A., Morais, F., Nino-Veja, G., Taylor, J.W., 2006a. Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Mol. Biol. Evol.* 23, 65-73.
- Matute, D.R., Sepulveda, E.V., Quesada, L.M., Goldman, G.H., Taylor, J.W., Restrepo, A., McEwen, J.G., 2006b. Microsatellite Analysis of three phylogenetic species of *Paracoccidioides brasiliensis*. *J. Clin. Microbiol.*, 44, 2153-2157.
- Marini, M.M., Zanforlin, T., Santos, P.C., Barros, R.R.M., Guerra, A.C.P., Puccia, R., Felipe M.S.S, Brigido, M., Soares, C.M.A., Ruiz, J.C., Silveira, J.F., Cisalpino P.S., 2010. Identification and characterization of Tc1/mariner-like DNA transposons in genomes of the pathogenic fungi of the *Paracoccidioides* species complex. *BMC genomics*, 11 doi: 10.1186/1471-2164-11-130.
- Mayr, E., 1963. *Animal species and evolution*. Harvard University Press, Cambridge, MA. 797p.
- McEwen, J.G., Bedoya, V., Patino, M.M., Salazar, M.E., Restrepo, A., 1987. Experimental murine paracoccidioidomycosis induced by the inhalation of conidia. *J. Med. Vet. Mycol.* 25, 165-175.
- Naiff, R. D., Ferreira, L. C. P., Barthe, T. V., Naiff, M. F. and Arias, J. R., 1986 Paracoccidioidomicose enzoótica em tatus (*Dasyus novemcinctus*) no Estado do Pará. *Rev. Inst. Med. Trop. São Paulo* 28, 19-27.
- Nascimento, T., Martinez, R., Lopes, A.R., Bernardes, L.A.D., Barco, C.P., Goldman, M.H.S., Taylor, J.W., McEwen, J.G., Nobrega, M.P., Nobrega, F.G., Goldman, G.H., 2004. Detection and selection of microsatellites in the genome of *Paracoccidioides brasiliensis* as molecular markers for clinical and epidemiological studies. *J. Clin. Microbiol.* 42, 5007-5014.
- Negróni, P., 1966. El *Paracoccidioides brasiliensis* vive saprotíficamente en el suelo Argentino. *Prensa Med. Argent.* 53, 2381-2382.

- Nei, M., Kumar, S., 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Panchal, M., 2007. The automation of Nested Clade Phylogeographic Analysis. *Bioinformatics*, 23, 509-510.
- Pereira, S.L., Miyaki, C.Y., Russo, C.A.M., 2001. Reconstrução filogenética: Métodos probabilísticos. In *Biologia Molecular e Evolução*, Holos Editora, p.117-129.
- Posada, D., Crandall, K.A., Templeton, A.R., 2000. GeoDis: A program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9(4), 487-488.
- Purtilo, D.T., Walsh, G.P., Storrs, E.E, Gannon, C., 1975. The immune system of the nine-banded armadillo (*Dasypus novemcinctus* Linn). *Anat. Rec. Philad.* 181, 725-734.
- Restrepo, A., Baumgardner, D.J., Bagagli, E., Cooper, C.R.Jr., McGinnis, M.R., Lázeras, M.S., Barbosa, F.H., Bosco, S.M.G., Camargo Z.P.F.R., Coelho, K.I.R., Fortes, S.T., Franco, M., Montenegro, M.R., Sano, A., Wankes, B., 2000. Clues to the presence of pathogenic fungi in certain environments. *Med. Mycol.* 38, 67-77.
- Restrepo, A., 1985. The ecology of *Paracoccidioides brasiliensis*: A puzzle still unsolved. *J. Med. Vet. Mycol.* 23, 324-334.
- Ricci, G., Mota, F.T., Wakamatsu, A., Serafim, R.C., Borra, R.C., Franco, M., 2004. Canine paracoccidioidomycosis. *Med. Mycol.* 42, 379-83.
- Richini-Pereira, V. B. ; Bosco, S.M.G. ; Griese, J ; Theodoro, R. C. ; Macoris, S.A.G.; da Silva, RJ ; Barrozo, L. ; Tavares, PM ; Zancoppe-Oliveira, RM ; Bagagli, E., 2007. Molecular detection of *Paracoccidioides brasiliensis* in road-killed wild animals. *Med. Mycol.* 18, 1-6.
- Richini-Pereira, V.B.; Bosco, S.M.G. ; Theodoro, R. C. ; Macoris, S.A.G.; Bagagli, E., 2009. Molecular approaches for eco-epidemiological studies of *Paracoccidioides brasiliensis*. *Mem. Inst. Oswaldo Cruz* 104, 636-643.
- Ridley, M., 2006. *Evolução*. 3 ed. Porto Alegre. Artmed, 481-484p.
- Ronquist, F., 1996. DIVA version 1.1 Computer Program and Manual Available by Anonymous FTP from Uppsala University. <http://www.ebc.uu.se/systzoo/research/diva/diva.html>.
- Salgado-Salazar, C., Jones, L.R., Restrepo, A., McEwen, J.G., 2010. The human fungal pathogen *Paracoccidioides brasiliensis* (Onygenales: Ajellomycetaceae) is a complex of two species: phylogenetic evidence from five mitochondrial markers. *Cladistics* 26, 1-12.
- Sano, A., Tanaka, R., Yokoyama, K., Franco, M., Bagagli, E., Montenegro, M. R., Mikami, Y., Miyaji, M., Nishimura, K. 1999. Comparison between human and armadillo *Paracoccidioides brasiliensis* by random amplified polymorphic DNA analysis. *Mycopathologia*, 143, 165-169.
- Schlotterer, C., Tautz, D., 1992 Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20, 211-215.

- Shome, S.K.; Batista, A.C., 1963. Occurrence of *Paracoccidioides brasiliensis* in the soil of Recife, Brazil. *Rev. Fac. Med. Univ. Fed. Ceará* 3, 90-94.
- Silva-Vergara, M.L., Martinez, R., Camargo, Z.P., 2000. Isolation of *Paracoccidioides brasiliensis* form armadillos (*Dasypus novemcinctus*) in an area where the fungus was recently isolated from soil. *Med. Mycol.* 38, 193-199.
- Silva-Vergara, M. L.; Martínez, R.; Chadu. A.; Madeira, M.; Freitas-Silva, G., 1998. Isolation of *Paracoccidioides brasiliensis* strain from the soil of a coffee plantation in Ibiá, State of Minas Gerais, Brazil. *Med. Mycol.* 36, 37-42.
- Sheff, K.W., York, E.R., Driebe, E.M., Barker, B.M., Rounsley, S.D., Waddell, V.G., Beckstrom-Sternberg, S.M., Beckstrom-Sternberg, J.S., Keim, P.S., Engelthaler, D.M., 2010. Development of a rapid, cost-effective TaqMan Real-Time PCR Assay for identification and differentiation of *Coccidioides immitis* and *Coccidioides posadasii*, *Medical Mycology*, 48, 466–469.
- Taber, F.W., 1945. Contribution on the life story and ecology of the nine-banded armadillo. *J. Mammal.* 26, 211-216.
- Talmage, R.V., Buchanan, G.D., 1954. The armadillo *Dasypus novemcinctus* – a review of its natural history, ecology, anatomy and reproductive physiology. Rice Inst. Pamphlet Houston 41.
- Taylor, J. W., Jacobson, D.J., Kroken, S., Kasuga, T., Gêiser, D.M., Hibbett, D.S., Fisher, M.C., 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Gen. Biol.* 31(1), 21-32.
- Taylor, J.W., Turner, E., Townsend, J.P., Dettman, J.R., Jacobson, D., 2006. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1947-1963.
- 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., 2009. Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides* genus. *Mol Phylogenet Evol.* 52, 273-283.
- Terçarioli, G.R., Theodoro, R.C., Bosco, S.M.G., REIS, G.M., Macoris, A.S.G., Simões, L.B., Bagagli, E., 2007. Ecological study of *Paracoccidioides brasiliensis* in soil: growth ability, conidia production and molecular detection.. *BMC Microbiology (Online)*, v. 22, p. 7(1):92.
- Theodoro, R.C., Bagagli, E., Oliveira, C., 2008b. Phylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complex. *Fungal Genet. Biol.* 45: 1284-1291.
- Theodoro, R. C., Bosco, S. M. G., Araújo, J. P. JR., Candeias, J. M., Macoris S. A. G., Trinca, L. A., Bagagli, E., 2008a. Dimorphism, Thermal Tolerance, Virulence and Heat Shock Protein 70 Transcription in different isolates of *Paracoccidioides brasiliensis*. *Mycopathologia.* 165, 355-365.
- Theodoro, R.C., Bosco, S.M.G., Sugizaki, M.F.; Bagagli, E., 2005b. Variation on mycelial-to-yeast phase transition and production of conidia in different isolates of *Paracoccidioides brasiliensis* obtained from human, Armadillo and dog. *Rev. Inst. Med. Trop. São Paulo* 47(14), 45.

- Theodoro, R.C. Candeias, J.M.G., Araújo Jr., J.P., Bosco, S.M.G., Macoris, S.A.G., Padula Jr., L.O., Franco, M., Bagagli, E., 2005a. Molecular detection of *Paracoccidioides brasiliensis* in soil. *Med. Mycol.* 43(8), 725-729.
- Ulrich, M., Convit, J., Centeno, M., Rapetti, M., 1976. Immunological characteristics of the armadillo, *Dasypus sabanicola*. *Clin. Exp. Immunol.* 25, 170-176.
- Untereiner, W.A, Scott, J.A., Naveau, F.A., Sigler, L., Bachewish, J., Angus, A., 2004. The Ajellomycetaceae, a new family of vertebrate-associate Onygenales. *Mycol.* 96(4), 812-821.
- Vilela, R., Rosa P.S., Belone, A.F., Taylor, J.W., Diorio, S.M., Mendoza, L., 2009. Molecular phylogeny of animal pathogen *Lacazia loboi* inferred from rDNA and DNA coding sequences. *Mycol. Res.* 113, 851-857.
- Wanke, B., Londero, A.T., 1994. Epidemiology and Paracoccidioidomycosis Infection. In: *Paracoccidioidomycosis*. FRANCO, M.F., LACAZ, C.S., RESTREPO, A., DEL NEGRO, G. (Eds). Boca Raton, London, CRC Press, cap.7.
- White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds). *PCR Protocols: A guide to methods and applications*. San Diego: Academic Press, 315-322.

Artigo 1

A ser submetido

Paracoccidioides genus: species recognition and biogeographic aspects.

Raquel Cordeiro Theodoro¹, Marcus de Mello Teixeira², Maria Sueli Soares Felipe², Karina dos Santos Paduan¹, Paulo Martins Ribolla¹, Gioconda San-Blas³, Eduardo Bagagli^{1*}

¹ – Universidade Estadual Paulista, Campus de Botucatu - UNESP, SP, Brazil

² – Universidade de Brasília – UNB, DF, Brazil

³ - Venezuelan Institute for Scientific Research, Center of Microbiology and Cell Biology, Caracas 1020 A, Venezuela

*Corresponding author:

e-mail: bagagli@ibb.unesp.br

Dept. de Microbiologia e Imunologia

Instituto de Biociências

UNESP-Botucatu

Distrito de Rubiao Junior, s/n

Botucatu-Sao Paulo- Brazil

Zip Code: 18618-970

Phone +55 14 3811-6058

Fax +55 14 3815-3744

Abstract

Paracoccidioidomycosis is a systemic mycosis caused by *Paracoccidioides brasiliensis* (cryptic species S1, PS2, PS3), and *Paracoccidioides lutzii* (Pb01-like species). For conclusive species recognition within the genus *Paracoccidioides*, each of them was subjected to an analysis of morphological characters in their yeast and mycelial phases, followed by single nucleotide polymorphism (SNP) studies on *GP43*, *ARF* and *PRP8* intein.

To evaluate the biogeographic events that led to the current geographic distribution of *P. brasiliensis* S1, PS2, PS3, *P. lutzii* and their sister species *Lacazia loboi*, Nested Clade (NCA) and Dispersal-Vicariance (DIVA) analyses were applied, in association with the divergence time calculated on the basis of *ARF* substitution rate. Our results suggest that the radiation of *Paracoccidioides* started around 20 million years ago (mya), in the Brazilian Shield. The main events involved in the speciation of this genus are related to vicariance, responsible for the divergence among S1, PS2 and *P. lutzii*, plus a recent dispersal that generated the PS3 species, restricted to the Colombian territory. Taking into account the ancestral areas revealed by DIVA analysis, and the major geographic distribution of *L. loboi* in the Amazon basin, a region extremely affected by the Andes uplift and marine incursions in the Cenozoic era, we also speculate about the effect these geological events had on the observed vicariance between *Paracoccidioides* and *L. loboi*.

Key-words: *Paracoccidioides*, species recognition, molecular markers, biogeography, marine incursions, *Lacazia loboi*, Ajellomycetaceae

Introduction

The genus *Paracoccidioides* encompasses thermo-dimorphic fungal pathogens from the family Ajellomycetaceae, Order Onygenales (Untereiner et al., 2004); it grows as yeast cells at 37°C or in mammal tissues, and as mycelia, producing the infective asexual spores or conidia, at 25°C or in soil (Franco et al., 1987). All members of this family, which also includes the pathogenic species *Blastomyces dermatitidis* (teleomorph *Ajellomyces dermatitidis*), *Histoplasma capsulatum* (teleomorph *Ajellomyces capsulatus*), *Emmonsia parva* and *E. crescens* (teleomorph *Ajellomyces crescens*), might have evolved in close association with vertebrate hosts, presenting a saprobe phase in soil and/or feces and a parasitic phase in host tissue (Untereiner et al., 2004; Herr et al., 2001; Bagagli et al., 2006). In a recent study, the pathogenic fungus *Lacazia loboi* was also included in this group, as a *Paracoccidioides* sister species (Vilela et al., 2009). This fungus, incapable of growing in culture media, is known to cause a subcutaneous mycosis in dolphins and humans, especially those from the Amazon basin (Taborda et al., 1999).

Until 2006, the genus *Paracoccidioides* was believed to include only one species, *Paracoccidioides brasiliensis*, as the etiologic agent of paracoccidioidomycosis (PCM), one of the most frequent systemic mycoses in Latin America (Franco 1987). This pathogen has been repeatedly recovered from human clinical samples and tissues of some species of armadillos, such as the nine-banded armadillo, *Dasypus novemcinctus* (Bagagli et al. 2003) and occasionally from *Cabassus centralis* (Corredor et al., 2005). It was also detected in and isolated from dogs (Ricci et al. 2004, Farias et al. 2005). Molecular polymorphism analysis suggested that the same “ecopathogenotypes” can infect humans and other mammals (Hebeler-Barbosa et al. 2003). It was only through Multi Loci Sequencing Type (MLST) analysis that the genetic variability, formerly known as merely intraspecific and geographic polymorphism, revealed the existence of four cryptic species: S1, PS2 and PS3, from the *P. brasiliensis* complex (Matute et al., 2006a, 2006b) and *Paracoccidioides lutzii* (originally called Pb01-like) (Teixeira et al., 2009; Carrero et al., 2008). S1 is the most widely distributed species, occurring in Brazil,

Argentina, Paraguay, Peru and Venezuela, followed by PS2 in Venezuela and Brazil, in sympatry with S1, while PS3, the only clonal species from the genus, is endemic to Colombia (Matute et al., 2006a). *P. lutzii*, the last discovered species, is the most divergent and occurs prevalently, though not exclusively, in the Mid-Western region of Brazil (Teixeira et al., 2009).

Although there is no clear agreement on S1, PS2 and PS3 being geographical variants of one single (Salgado-Salazar et al., 2010) or distinctly separated species (Matute et al., 2006 a, b), these clades and *P. lutzii* are reproductively isolated in nature, as revealed by Split Decomposition Analysis (Matute et al., 2006a; Teixeira et al., 2009). Since reproductive or genetic isolation is the first step for species divergence (Taylor et al., 2000), this might eventually lead to morphological and physiological differences, with important consequences for the diagnosis and treatment of PCM. For instance, Batista Jr et al (2009) demonstrated that the *Paracoccidioides* standard gp43 reference antigen for PCM, when tested against sera from patients living in geographically distant areas, gave a high frequency of false negative results in immunodiffusion tests, possibly due to the assays being performed in different *Paracoccidioides* species.

This work aimed to evaluate morphological and molecular markers for fast species recognition in the genus *Paracoccidioides*, without the need of multiloci sequencing. Morphological analysis included yeast cell morphometry and conidia production in the mycelial phase of a significant number of isolates from each species. For molecular analysis, SNPs from the genes *GP43* (43KDa immunodominant glycoprotein), *ARF* (ADP-ribosylation factor) (Matute et al., 2006a) and *PRP8* intein (intervening parasitic genetic element from the *PRP8* gene) (Theodoro et al., 2008a) were analyzed. Additionally, phylogenetic data and geographic locations of every isolate were associated by Nested Clade (NCA), and Dispersal-Vicariance analysis, in an attempt to resolve the ancestral areas and the main biogeographic events that might have taken place during the radiation of this genus in South America and its divergence from its sister species, *L. loboii*.

Material and Methods

Fungal isolates and growth conditions: The isolates are listed in Table 1. The yeast phase was kept in GPYA (2% glucose, 1% peptone, 0.5% yeast extract and 1% agar) at 36°C, and the mycelial phase in PDA (Potato dextrose agar – Oxoid Ltd., Basingstoke, UK), at 25°C.

Morphometry of yeast cell: Microscope slides of yeast cells on the 7th day of growth were stained with lactophenol cotton blue. Five random microscopic fields were analyzed with the Leica Qwin V3 software. About 30 yeast cells were measured by microscopic field. Two diameter measures of each cell were registered onto an Excel spreadsheet to determine the area, according to the following formula: (diameter 1 x 0.5) x (diameter 2 x 0.5) x 3.1416. The relationship between area and species was evaluated by using the R statistical software.

Conidia production: Isolates were kept at 25°C for 30-40 days, as giant colonies in Petri dishes, both on PDA and Soil Extract Agar (SEA), prepared according to Kwon-Chung (1973). Production of conidia was evaluated by using adhesive tape (Durex, 3M), which was softly touched onto the surface of the colonies, previously killed by formaldehyde vapor for 48 h, and also by the slide culture technique, using both culture media. Microscopic observation was carried out in a Photomicroscope (Olympus, BX60); the Leica Qwin V3 software was used for estimation of the conidial sizes.

Single nucleotide polymorphism (SNP) analysis: the DNA was extracted according to McCullough et al (2000), quantified in *Nanovue* (GE Healthcare) and diluted in ultra pure water to 15ng/μL. The SNP analysis was carried out by using the SNaPshot[®] Multiplex System and/or by Real Time PCR with TaqMan[®] Probes (Applied Biosystems). For the SNaPshot technique, the *ARF* gene and the *GP43* exon 2 were amplified by PCR.

Reactions carried out in 25 µl of reaction mixture containing 10 ng genomic DNA, 1× PCR buffer (50 mM KCl, 10 mM Tris HCl and 1.5 mM MgCl₂), 0.2 mM dNTP, 10 pmoles of each primer and 1 unit of *Taq* polymerase (GE Healthcare). Amplifications were performed in a thermocycler (Eppendorf). Thermal cycling conditions were as follows: 94°C for 4 min, 40 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min and a final cycle at 72°C for 5 min. Primers sequences and amplicon lengths are listed in Supplementary Material 1.

The *ARF* and *GP43* sequences, previously obtained by Matute et al. (2006a) for isolates belonging to S1, PS2 and PS3 species, and by Teixeira et al. (2009) for *P. lutzii* isolates, were aligned by Clustal W, in Mega v4.0 software (Kumar et al., 2004). Species-specific SNPs were selected for species recognition of 34 not yet identified *Paracoccidioides* isolates. For SNaPshot, forward primers (SNP-primers) were designed on flanking SNPs regions. Cytosines were added to their 5' end so that the amplified fragments could differ in length and therefore, position, in the polyacrylamide gel. Primer sequences, number of added cytosines, SNP positions and expected SNP for each species are listed in Table 2.

SNaPshot reactions were carried out as follows: 6 µL ultra pure water, 1 µL ½ TERM buffer (200 mM Tris HCl, 5 mM MgCl₂, pH 9.0), 1 µL SNaPmix (Applied Biosystems), 1 µL (1 pmol/µL) SNP-primer and 1 µL of the PCR product, previously purified with ExoSAP-IT (GE Healthcare) according to the supplier's instructions. Reactions were performed in a Whatman Biometra® (T Gradient) thermo cycler. The thermal cycling consisted of 40 cycles of 95°C for 10 sec, 50°C for 5 sec and 60°C for 30 sec.

Since the unincorporated [F]ddNTPs can co-migrate with the fragment of interest, the 5' phosphoryl groups were removed by enzymatic treatment with SAP (shrimp alkaline phosphatase, GE Healthcare), according to the supplier's instructions. Reaction samples were diluted 1:1 with Loading Buffer [45% HI DI formamide (v/v) (Applied Biosystems), 2.26 µM EDTA pH 8.0 and 2.26% Blue Dextran (w/v)]. Samples were heated at 95°C for 3 min and placed on ice. One micro liter of each sample was loaded into 12%

polyacrylamide gel and submitted to electrophoresis, in an ABI PRISM 377 DNA sequencer (Applied Biosystems). The SNPs were analyzed according to their colors.

For Real Time PCR, TaqMan probes and the forward and reverse primers for *GP43*, *ARF* and *PRP8* intein were designed (table 2). Polymorphic sites of the *PRP8* intein has been previously evaluated (Theodoro et al., 2008a). For each reaction, 1.5 µL ultra pure water, 5 µL Master Mix (Fermentas), 0.5 µL 20X primers and probe mixture (Applied Biosystems) and 3 µL DNA diluted to 15 ng/µL, were used. Thermal cycling conditions were 95°C for 10 min, followed by 40 cycles 92°C for 15 sec and 60°C 1 min, in the Step One Plus™ equipment (Applied Biosystems). To determine SNPs, measurement of end point fluorescence was done by means of the Step One Plus software (Applied Biosystems).

Biogeographic analysis

Nested Clade Analysis (NCA): To study the association between phylogenetic data and geographic location of our isolates (Supplementary Material 2), the software ANeCA (Templeton et al., 2005; Panchal, 2007; Clemente et al., 2000; Posada et al., 2000) was used. Nucleotide sequences from *ARF* (407 nt), *GP43* (427 nt from exon 2) and *TUB* (α -tubulin) (243 nt) were used in separate analyses. A haplotype genealogy, a population distribution profile for each haplotype and a nested design (hierarchical clustered haplotypes known as “clades”) were constructed by using the TCS and nesting algorithms. The geographic information (latitude, longitude and the approx. 200 Km² range of isolate occurrence) was used as an input file for running the geographic association test in GeoDis program. Finally, the automated inference key was run.

Dispersal-Vicariance analysis: *ARF* nucleotide sequences from each of the four *Paracoccidioides* species (7 haplotypic sequences) and from outgroup species *L. loboi*, *H. capsulatum* and *B. dermatitidis* were used for this analysis (Figure 1). The phylogenetic

tree obtained by Neighbor Joining, using Tamura Nei (Tamura & Nei, 1993) as evolutionary model, in Mega v5.0 software (Kumar et al., 2004) was written in parenthesis and used as input information together with the area distribution of each taxon. Thirteen areas were settled for the biogeographic analysis (Figure 1). Dispersal-Vicariance analysis was performed on DIVA software (Ronquist, 1996) to infer ancestral distributions and events involved in the biogeographic history of the genus *Paracoccidioides* and its sister species, *L. loboii*.

In order to place the biogeographic events, suggested by the NCA and DIVA analysis, in the geological time scale, we estimated the time of *Paracoccidioides* radiation and the divergence time between this genus and *L. loboii*, using the DNA substitution rate of 0.86×10^{-9} , previously calculated by Kasuga et al (2003) for the gene *ARF*. The genetic distance between *P. brasiliensis* S1, PS2, PS3 isolates, and *P. lutzii* and other Ajellomycetaceae species, was determined by using the Tamura Nei (Tamura & Nei, 1993) model in Mega v5.0 software (Kumar et al., 2004). The distance value was used to work out the divergence time between groups according to the formula $T=d/2\lambda$, where λ is the substitution rate, d the genetic distance and T the time of divergence (Nei, 1987).

Results

Morphological analysis: The yeast cell area seems unrelated to the genetic group as no significant variations on yeast size among the species were found, with the only exception of Pb01, the first *P. lutzii* isolate ever reported (Carrero et al., 2008; Teixeira et al., 2009); this isolate shows very large yeast cells when compared to other isolates. The PS2 isolates, except Pbdog and Pb927, presented elongated cell shapes when compared to yeast cells from all other species, as reported previously (Theodoro et al., 2008b) (Supplementary Material 3).

Most S1 isolates (18 out 26) produced conidia, being more abundant in SEA than in PDA medium. The Pbdog was the only PS2 isolate to produce as many conidia as the

S1 isolates. Two Colombian PS3 isolates were unable to generate conidia in either media. *P. lutzii* isolates produced an intermediate to high level of conidia (Table 3 and Supplementary Material 4). Most of the Pb01 conidia, and few EE conidia, both *P. lutzii*, are longer (the length, from septa to septa, ranged from 2.17 to 22.89 μm , being 11.07 μm the median value) than S1 and S2 conidia whose length, from septa to septa, ranged from 2.25 to 7.99 μm (median value, 5.13 μm).

Single Nucleotide Polymorphism analysis: Analysis of four SNPs, two in the exon 2 of *GP43*, one in *ARF* gene and one in the PRP8 intein (Table 4 and Figure 2), allowed us to tell apart thirty four new *Paracoccidioides* isolates into species. Of them, 30 belonged to S1 and 4 to PS2.

Inference of the biogeographic history of the genus *Paracoccidioides*: The statistically significant historical events generated by the Nested Clade Analysis are summarized in Table 5. In general, the results indicate a continuous expansion of isolates S1 and a dispersal event that gave rise to the PS3 species in Colombia. Inside the S1 species, a possible fragmentation event gave rise to a S1 subclade within the Venezuelan isolates, corroborating the idea that they could be considered a S1 cryptic subspecies (Salgado-Salazar et al., 2010). In addition, the DIVA analysis generated a tree with ancestral areas in its nodes (Figure 1), suggesting a long dispersal event for the PS3 species, and vicariant events that played an important role not only in the divergence among S1, PS2 and *P. lutzii*, but also between the genus *Paracoccidioides* and its sister species *L. loboii*.

Discussion

Concerning speciation of the genus *Paracoccidioides*, both a rather simple identification method and an evolutionary view are presented herein. The first one is

focused on a fast and firm diagnosis of *Paracoccidioides* species, from clinical and/or environmental samples, suitable for epidemiological studies and eventually, for clinical use, should these different species turn out to induce different host responses. The second view infers the possible historical events that have resulted in the current biogeography of this genus and its sister species, *L. loboii*.

Morphological and molecular markers for *Paracoccidioides* species

Because of subtle or inexistent morphological differences among cryptic species, molecular techniques become the tool of choice to tell them apart. For example, *Coccidioides posadasii* grows slowly in high concentrations of salt when compared to *C. immitis* (Fisher et al., 2002), but this phenotype is not used for species recognition. The same is valid for the morphological study on *Paracoccidioides*. Here we observed that yeast cell size can vary in the same isolate and in isolates from the same species, but it does not distinguish S1, PS2, PS3 and *P. lutzii*. However, trends in yeast cell shape suggest that most PS2 isolates, except the Pbdog and Pb927, tend to present elongated yeast cells, just like pseudohyphae. These isolates do not easily transform from the mycelial to the yeast phase (data not shown). In fact, one of them, the Pb22 isolate, although obtained as yeast from human patient, does not grow at all in the yeast morphology. It is important to highlight here that all isolates from table 1, including Pb22, were submitted to ITS1-5.8S-ITS2 sequencing, being confirmed as *Paracoccidioides* isolates (data not shown).

Production of conidia, the infective asexual spore of the mycelia phase, was higher in SEA than in PDA medium, probably because the former resembles the natural conditions in which this process might occur (Terçarioli et al., 2007). Most S1 isolates (69%) produced conidia under these conditions; with the exception of Pbdog (PS2), all PS2 (9 samples) and PS3 isolates (2 samples) produced very little or no conidia. It is important to mention that the Pbdog isolate has other peculiarities, such as a diffusible

brown pigment in agar media and a smooth, instead of wrinkled, aspect of the yeast culture. Sequencing of the eight loci for species recognition (Matute et al., 2006a), has confirmed that Pbdog does belong to the PS2 species (Bosco et al., unpublished data). Teixeira et al. (2009) observed very long conidia in Pb01, a characteristic shared with another *P. lutzii* isolate, EE.

Therefore, the limited morphological studies done so far to characterize each *Paracoccidioides* species prevent us from providing distinctive clues to tell them apart. On the contrary, the chosen SNP markers (SNP1-ARF or SNP3-ARF for PS2, SNP2-GP43-exon 2 or SNP5-PRP8 intein for *P. lutzii*, and SNP4-GP43-exon2 for PS3) are better suited for conclusive species recognition within the genus *Paracoccidioides*; with them it was possible to classify 34 additional isolates, that were included in a map of geographic distribution of *Paracoccidioides* species in South America (Figure 3).

Biogeographic history of *Paracoccidioides* genus

Inferring the biogeographic history of *Paracoccidioides*, is not simple because of: i) the constant migration of human hosts; ii) the long latency period of PCM and iii) the scarcity of environmental isolates. With the exception of one *P. lutzii* strain isolated in the Southern region of Brazil (Paraná State; Takayama et al., 2009), there is a strong clustering in the species distribution of *Paracoccidioides*, according to their origin (Figure 3), aspect already documented by Calcagno et al. (1998), long before the discovery of cryptic species in this pathogen.

All environmental and armadillo (*D. novemcinctus*; *C. centralis*) isolates so far available in culture collections (Naiff et al., 1986; Macedo et al., 1999; Silva-Vergara et al., 2000; Corredor et al., 1999; Bagagli et al., 1998; 2003; Corredor et al., 2005), belong to *P. brasiliensis* S1, PS2 and PS3. They were isolated in Southern Brazil (*D. novemcinctus*) or Caldas, Colombia (*C. centralis*). An evaluation of wild animals, mainly armadillos, in Mid-Western Brazil, is pending in order to confirm the geographic distribution of *P. lutzii*.

NCA analysis suggests a long dispersal event, rather than a vicariance process for the emergence of PS3 species. This may have been due to migration of S1 individuals to Colombia (Matute et al., 2006a), crossing preexisted geographic barriers. DIVA analysis also corroborates a dispersal hypothesis, rather than a vicariance process, because the area “F”, which refers to Colombia, is a derived and not an ancestral area.

Since most PS3 isolates available to date come from Antioquia, Colombia, a region located between the Andean Western and Central Cordilleras, we believe that this geographically secluded location may have contributed to the isolation of PS3. Although the *ARF* phylogeny makes S1 paraphyletic, that is S1 and PS3 isolates together in some clades, the PS3 speciation by dispersal to Colombia might have occurred less than 4.9mya, which is the estimated time for the divergence between S1/PS3 and PS2 according to the *ARF* substitution rate and Tamura Nei distance (Figure 4). This recent dispersal contrasts with the 20 mya radiation of the genus *Paracoccidioides* in the South-Eastern and Mid-Western regions of the Brazilian Shield, the most probable ancestral areas (Figure 1). A possible explanation for this recent speciation could be the impact of the two major geological events that took place in South America from 60-8 mya, which were the Andes uplift and the episodic marine incursions, caused by global sea level rises during the extensional tectonic phases (Hoorn, 1993; Hoorn et al., 1995; Lunenberg et al., 1998).

In contrast to the geological instability of the Andean region, the neighbouring Brazilian and Guiana shields are extremely stable, not having undergone any deformation in the past 10^8 - 10^9 years. The Guiana shield encompasses most of the Guianas surface, Southern Venezuela, South-Eastern Colombia and Northern Amazonia, while the Brazilian shield includes Mid-Western and South-Eastern Brazil. Between these shields is the Amazonian depression, originated in the Paleozoic, and reactivated in the Triassic while the separation of tectonic plates was taking place, with the end product of Africa and South America as different continents. After that, the convergence of tectonic plates gave rise to the Andes Cordillera, changing the drainage

of the Amazonian rivers from west to east. This new drainage system, and the coincident marine incursions, created a large wetland portion that allowed the interflow between fresh and marine organisms (Lundberg et al., 1998). At least five marine incursions took place in South America, two independent ones in Southern Paraná and North-Western Amazon basin, in the late Cretaceous (61 mya), another one in the early Eocene (around 50 mya) in Colombia, a fourth one in the Miocene (around 20 mya) in Colombia and the West-Northern regions of the Amazon basin, and finally, also in this area, a late Miocene (11.8-7.0 mya) incursion. In this way, a 1 million Km² Pebas/Solimões lake was formed (Wesselingh, 2002; 2006, Wesselingh et al., 2006; Lundberg et al., 1998), covering most of today's Colombian territory. Until then, while *Paracoccidioides* was going through a radiation process in South America, the Colombian territory was submerged, avoiding any possible dispersal event.

Vertebrate hosts might have played an important role in this long dispersal event. Indeed, the long latency period, the tendency to chronicity and the mitigation of virulence traits have been pointed as evidence for a long coexistence with vertebrates, which could have shaped the current geographic distribution of *Paracoccidioides* species. Richini-Pereira et al (2009) suggested the importance of animals from Xenarthra order, in special armadillos, for eco-epidemiological studies on PCM. The authors demonstrated a high frequency of molecular detection of *Paracoccidioides* spp in tissue samples from different Xenarthra and associated their data with the weak immune response and lower body temperature of these animals (Purtilo et al., 1975 and Ulrich et al., 1976). In addition, the ancient origin of Xenarthra order (around 65 mya, in South America) and its coexistence with fungi from Onygenaceae, sensu lato, family, which includes *Paracoccidioides* genus, suggest that these mammals could have had an important role for the evolution of parasitism in *Paracoccidioides* (Bagagli et al., 2006).

The remaining species from *Paracoccidioides* genus occur almost sympatrically across their geographic range, except *P. lutzii*, which is prevalent in the Brazilian Middle Western. This regionalization suggests an alopatric speciation. However, according to

DIVA, the most probable ancestral area of the genus encompasses the Middle-West and Southeast regions of the Brazilian Shield which is an extremely stable territory, that is, no clear geographic barrier is known to have separated the *P. lutzii* from S1 and PS2 species. Still, defining geographic barriers for microorganisms is not as simple as for more complex organisms, such as plant and animals. Small variations in weather, temperature, humidity and environmental resources might create ecological barriers, being enough for species divergence and emergence of different adaptations. Since the geographic distance between *P. lutzii* and S1/PS2 is relatively not so far, their divergence could have been a result of a reproductive isolation leading to a parapatric speciation due to ecological variation, rather than distance as observed for PS3. A restricted gene flow between these two populations could have been intensified by Natural Selection and genetic drift. In order to better explore this hypothesis, the interaction of these pathogens with saprobic substrates should be evaluated. Until now, the single study carried out in this manner analyzed the fungal growth in different soil textures, with different water availability (Terçarioli et al., 2007), but without including isolates from *P. lutzii* species, which, by the way, had not been described yet.

The sympatric distribution of S1 and PS2 may be linked to other factors, as documented in other fungal species, i.e. *Pneumocystis* (Keely and Stringer, 2005) or *Botrytis* (Fournier & Giraud, 2008) species, which may have diverged after a specialization process towards different animal or plant hosts. This possibility is ruled out to explain S1 and PS2 sympatry, inasmuch as they infect the same hosts, human and armadillos. The probability of substrate and resource preferences in their saprobic life styles (Disruptive selection) may be considered as a way to establish different adaptation processes in nature (Coyne & Orr, 2004). According to the Neutral theory, sympatric speciation may also occur, even without geographic barriers, ecological gradients or Natural selections (Aguiar et al., 2009).

Divergence between the genus *Paracoccidioides* and its sister species, *L. lobo*

Although *L. lobo* belongs to the Ajellomycetaceae family (Vilela et al., 2009), as sister species of the genus *Paracoccidioides*, it displays a peculiar life style when compared to other Ajellomycetacean pathogens: i) it is impossible to culture *in vitro*; ii) it infects dolphins, being associated with aquatic environment, and iii) it causes a subcutaneous, not systemic, mycosis both in humans and dolphins (Taborda et al., 1999).

Lacaziosis is prevalent in tropical and subtropical areas, particularly in the Amazon basin (Brazil and Colombia) (Figure 3), but there are reports in Costa Rica, Venezuela, Peru, French Guiana, Suriname, Panama, Guiana, Ecuador, Bolivia, Mexico, Canada and United States (Ramos & Silva et al., 2009). A single European case was described in a dolphin caretaker (Symmer, 1983) and the first reported case in the US came from a patient who had traveled to Venezuela (Burns et al., 2000).

Two dolphin species, *Tursiops truncatus* (marine) and *Sotalia fluviatilis* (fluvial), are responsible for the current distribution of *L. lobo*. The first species has a global distribution in tropical and equatorial regions and the second one occurs in the Amazon basin, the Atlantic coast from Florida to Paraná State (Brazil), and the Suriname rivers (Migaki et al., 1971; Cowan, 1993; De Vries & Laarman, 1973).

Since many cases of lacaziosis are the result of transmission from dolphins to human, water is believed to be the environmental source of the pathogen as well as the reason for its large geographic distribution (Ramos & Silva, 2009). Bagagli et al (2006) put forward the hypothesis according to which *L. lobo* may be in an evolutionary phase of strong reduction of the saprobic form and specialization towards parasitism, so that it depends on the host for its growth and dissemination. If this hypothesis is confirmed, the large geographic distribution of this pathogen will be associated with host migration rather than water itself. The high prevalence of *L. lobo* in the Amazon basin (Figure 3) and its phylogenetic proximity to all members of the genera *Paracoccidioides* and *Histoplasma*,

suggest that this peculiar Ajellomycetacean has emerged in South America, as have the other two.

According to our DIVA and divergence time analysis, the ancestral of *Paracoccidioides* and *L. lobo* lived in South America, in the Mid-Western, South-Eastern and Northern regions of Brazil around 36,7 mya. While being phylogenetically close to *Paracoccidioides*, *L. lobo* inhabits a different ecological niche (aquatic vs. soil), and causes a subcutaneous (not systemic) mycosis, transmitted by trauma and not by inhalation of spores. The habitat of this pathogen, mainly in the Amazon basin corroborates to the vicariant event suggested by our DIVA analysis, which explains the divergence of the ancestral population giving rise to *Lacazia* and *Paracoccidioides* genera. This event may have been triggered by the formation of new and empty ecological niches (wetland and/or totally submerged regions), and the simultaneous emergence of riverine Cetacean mammals in the Oligocene and early Miocene (Hamilton et al., 2001; Cassens et al., 2000), in which *L. lobo* ancestors may have found an ideal refuge away from the continuous instability represented by those saline/lacustre environments.

The phylogeography of the Ajellomycetaceae has not yet been fully resolved. Most of the information comes from the pathogenic members of the family, while little if any information is available for *Emmonsia* spp. and *Ajellomyces grisea*, the only non-pathogenic fungus in the group. While *H. capsulatum* and probably *Paracoccidioides* spp and *L. lobo* have been originated in South America (Kasuga et al., 2003), *B. dermatitidis* (teleomorph *Ajellomyces dermatitidis*) is the only species of this family that is endemic to North America. Because of a high genotypic similarity with the North American genotype, a few isolates in Africa and India are probably derived from human migration (McCullough et al., 2000). The *ARF* substitution rate observed in our data leads to a divergence time of ~199 million years between *B. dermatitidis* and *H. capsulatum* vs. *Paracoccidioides* and *L. lobo*, in the Jurassic. Then, the Nearctic and Neotropical land masses were still connected (Gondwana). Could the ancestors of the present Ajellomycetaceae family have occupied

all this area and diversified after the isolation of continents? In that era, the last glacial Pleistocene period affected the Northern hemisphere, which remained 80% covered by thick ice layers. On the contrary, few South American places went undisturbed (Brown & Lomolino, 2006). If so, could the low species diversity of the Ajellomycetaceae in US and Canada be related to such climatic changes?

These questions are just the starting point for more research on the evolution of the Ajellomycetaceans, studies that should take account more species and isolates, and also new calibration points for molecular clock estimatives. This will also help in the understanding of survival strategies in saprobic environment and/or vertebrate hosts.

Acknowledgments

We thank Reinaldo José da Silva and Luzia A. Trinca, for their assistance in yeast morphometry, Sílvio Alencar Marques, Zoilo Pires de Camargo and Adriana Vicentini-Moreira for supplying fungal isolates, and Carine Spenassatto for helping in Real Time PCR assays. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp Grant number: 07/01306-5).

References

- Aguiar, M.A.M., Baranger, M., Baptestini, E.M., Kaufman, L., Bar-Yam, Y., 2009. Global patterns of speciation and diversity. *Nature* doi:10.1038/nature08168.
- Antoneli, A., Nylanden, J.A.A., Persson, C., Sanmarti, I., 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *PNAS* 106, 9749-9754.
- Bagagli, E., Bosco, S.M.G., Theodoro, R.C., Franco, M., 2006. Phylogenetic and evolutionary aspects of *Paracoccidioides brasiliensis* reveal a long coexistence with animal hosts that explain several biological features of the pathogen. *Infect. Genet. Evol.* 6, 344-351.
- Bagagli, E., Sano, A., Coelho, K.I.R., Alquati, S., Miyaji, M., Camargo, Z.P., Gomes, G., Franco, M., Montenegro, M.R., 1998. Isolation of *Paracoccidioides brasiliensis* from armadillos (*Dasypus novemcinctus*) captured in an endemic area of paracoccidioidomycosis. *Am J Trop Med & Hyg.* 58, 505-512.

- Bagagli, E., Franco, M., Bosco, S. M. G., Hebler-Barbosa, F., Trinca, L. A., Montenegro, M. R., 2003. High frequency of *Paracoccidioides brasiliensis* infection in armadillo (*Dasyus novemcinctus*): an ecological study. *Med. Mycol.* 41, 217-223.
- Bagagli, E., Bosco, S.M.G., Theodoro, R.C., Franco, M., 2006. Phylogenetic and evolutionary aspects of *Paracoccidioides brasiliensis* reveal a long coexistence with animal hosts that explain several biological features of the pathogen. *Infect. Genet. Evol.* 6, 344-351.
- Bagagli, E., Theodoro, R.C., Bosco, S.M.G., McEwen, J.G., 2008. *Paracoccidioides brasiliensis*: phylogenetic and ecological aspects. *Mycopathologia.* 165, 197-207.
- Batista Jr, J., Camargo, Z.P., Fernandes, G.F., Vicentini, A.P., Fontes C.J.F., Hahn, R.C., 2008. Is the geographical origin of a *Paracoccidioides brasiliensis* isolate important for antigen production for regional diagnosis of paracoccidioidomycosis? *Mycosis* 53, 176-180.
- Brown, J.H., Lomolino, M.V., 2006. Glaciação e dinâmica Biogeográfica do Pleistoceno. In *Biogeografia* (2ed), Ribeirão Preto, SP, Funpec (Ed), 177-219.
- Burns, R.A., Roy, J.S., Woods, C., Padhey, A.A., Warnock, D.W., 2000. Report of the first human case of lobomycosis in the United States. *J. Clin. Microbiol.* 38, 1283-1285.
- Calcagno, A.M., Niño-Vega, G., San-Blas, F., San-Blas, G., 1998. Geographic Discrimination of *Paracoccidioides brasiliensis* Strains by Randomly Amplified Polymorphic DNA Analysis. *J. Clin. Microbiol.* 36, 1733-1736.
- Carrero LL, Niño-Vega G, Teixeira MM, Carvalho MJA, Soares CMA, Pereira M, Jesuino RSA, McEween JG, Mendoza L, Taylor J W, Felipe MS, San-Blas G 2008. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. *Fungal Genet. Biol.* 45(5), 605-612.
- Cassens, I., Vicario, S. Waddell, V.G., Balchowsky, H., Belle, D.V., Ding, W., Fan, C., Mohan, R.S.L., Simoes-Lopes, P.C., Batisda, R., Meyer, A., Stanhope, M.J., Milinkovitch, M.C., 2000. Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages. *PNAS* 97, 11343-11347.
- Clement, M. Posada, D., Crandall, K.A., 2000. TCS: A computer program to estimate genes genealogies. *Mol. Ecol.* 9(10), 1657-1659.
- Corredor, G. G., Castaño, J. H., Peralta, A., Díez, S., Arango, M., McEwen, J., Restrepo, A., 1999. Isolation of *Paracoccidioides brasiliensis* from the nine-banded armadillo *Dasyus novemcinctus*, in na endemic area for paracoccidioidomycosis in Colombia. *Rev. Iberoam. Micol.* 16(4), 216-220.
- Corredor, G.G., Peralta, L.A., Castaño, J.H., Zuluaga, J.S., Henao, B., Arango, M., Tabares, A.M., Matute, D.R., McEwen, J.G., Restrepo, A., 2005. The naked-tailed armadillo *Cabassous centralis* (Miller 1899): a new host to *Paracoccidioides brasiliensis*. Molecular identification of the isolate. *Med. Mycol.* 43(3), 275-280.

- Cowan, D.F., 1993. Lobo's disease in a bottle-nosed dolphin (*Tursiops truncatus*) from Matagorda Bay, Texas. *J. Wildl. Dis.* 29, 488-489.
- Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer (Ed), Sunderland, MA, USA.
- De Vries, G.A., Laarman, J.J., 1973. A case of Lobo's disease in the dolphin *Sotalia guianensis*. *Aquat. Mamm.*, 1, 1-8.
- Farias, M.R., Werner, J., Muro, M. D., Marques, S.A., Marques, M.E.A., Franco, M.F., Ribeiro, M.G., Custodio, C.C., Condas, L.A.Z., Bosco, S.M.G., Bagagli, E., 2005. Canine paracoccidioidomycosis: case report of generalized lymphadenitis. *Rev. Inst. Med. Trop.* 47(14), 64.
- Fisher, M.C., Koenig, G.L., White, T.J., Taylor, J.W., 2002. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. *Mycologia*, 94 (1), 73-84.
- Franco, M., 1987. Host-parasite relationships in paracoccidioidomycosis. *J. Med. Vet. Mycol.* 25, 5-18.
- Fournier, E., Giraud, T., 2008. Sympatric genetic differentiation of a generalist pathogenic fungus, *Botrytis cinerea*, on two different host plants, grapevine and bramble. *J. Evol. Biol.* 21, 122-132.
- Hamilton, H., Caballero, S., Collins, A.G., Brownell Jr, R.L., 2001. *Proc R Soc Lond B.* 268, 549-556.
- Hebeler-Barbosa, F., Morais, F. V., Montenegro, M. R., Kuramae, E. E., Taylor, J. W., Montes, B., McEwen, J. G., Puccia, R., Bagagli, E., 2003. Sequence comparison of the internal transcribed spacer regions and gp 43 in *Paracoccidioides brasiliensis* for patients and armadillos *Dasypus novemcinctus*. *J. Clin. Microbiol.* 41, 5735-5737.
- Herr, R.A., Tarcha, E.J., Taborda, P.R., Taylor, J.W., Ajello, L., Mendoza, L., 2001. Phylogenetic Analysis of *Lacazia loboi* Places This Previously Uncharacterized Pathogen within the Dimorphic Onygenales. *J. Clin. Microbiol.* 39, 309-314.
- Hoorn, C., 1993. Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: Results of a palynostratigraphic study. *Palaeogeogr Palaeoclimatol* 105, 267-309.
- Hoorn, C., Guerrero, J., Sarmiento, G.A., Lorente, M.A., 1995. Andean tectonics as a cause of changing drainage patterns in Miocene northern South America. *Geology* 23, 237-240.
- Kasuga, T., White, T.J., Koenig, G., McEwen, J., Restrepo, A., Castañeda E., Lacaz, C.S., Heins-Vaccari, E.M., Freitas R.S., Zancopé-Oliveira, R.M., Qin, Z., Negroni, R., Carter, D.A., Mikami, Y., Tamura, M., Taylor, M.L., Miller, G.F., Poonwan, N., Taylor, J.W., 2003. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol. Ecol.* 12, 3383-3401.
- Keely, S.P., Stringer, J.R., 2005. Nomenclature and genetic variation of *Pneumocystis*. In: Walzer, P.D., Cushion, M.T. (Eds), *Pneumocystis Pneumonia*, third ed. Marcel Dekker, New York, pp. 39-59.

- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5, 150-163.
- Kwon-Chung, K.J., 1973. Studies on *Emmonsia capsulata* I. Heterothalism and development of the ascocarp. *Mycologia* 65, 109-121.
- Lovejoy, N.R., Bermingham, E., Martin, A.P., 1998. Marine incursion into South America. *Nature* 396, 421-422.
- Lundberg, J.G., Marshall, L.G., Guerrero, J., Horton, B., Lalabarba, M.C.S.L., Wesselingh, F., 1998. The Stage for Neotropical fish diversification: A history of tropical South American rivers. In *Phylogeny and Classification of Neotropical Fishes*. Malabarba, L.R., Reis, R.E., Vari, R.P., Lucena, Z.M., Lucena, C.A.S. (eds). Porto Alegre, Edipucrs, 603p.
- Macedo, R.C.L., Lazera, M.S., Trilles, L., Bulcão, A.S., Silva Jr, N.J., Oliveira, N.A., Wanke, B., 1999. *Paracoccidioides brasiliensis*- In: VII International Meeting On Paracoccidioidomycosis, Campos do Jordão. Proceedings São Paulo: Brasil, p.128.
- Matute, D.R., McEwen, J.G., Montes, B.A., San-Blas, G., Bagagli, E., Rauscher, J.T., Restrepo, A., Morais, F., Nino-Veja, G., Taylor, J.W., 2006a. Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Mol. Biol. Evol.* 23, 65-73.
- Matute, D.R., Sepulveda, E.V., Quesada, L.M., Goldman, G.H., Taylor, J.W., Restrepo, A., McEwen, J.G., 2006b. Microsatellite Analysis of three phylogenetic species of *Paracoccidioides brasiliensis*. *J. Clin. Microbiol.*, 44, 2153-2157.
- McCullough, M.J., Disalvo, A.F., Clemons, K.V., Park, P., Stevens, D.A., 2000. Molecular epidemiology of *Blastomyces dermatitidis*. *Clin. Infec. Dis.* 30, 328-335.
- McEwen, J.G., Bedoya, V., Patino, M.M., Salazar, M.E., Restrepo, A., 1987. Experimental murine paracoccidioidomycosis induced by the inhalation of conidia. *J. Med. Vet. Mycol.* 25, 165-175.
- Migaki, G., Valerio, M.G., Irvine, B., Gardner, F.M., 1971. Lobo's disease in an Atlantic bottle-nosed dolphin. *J. Amer. Vet. Med. Assoc.* 159, 578-582.
- Naiff, R. D., Ferreira, L. C. P., Barreto, T. V., Naiff, M. F. and Arias, J. R., 1986. Paracoccidioidomycose enzoótica em tatus (*Dasypus novemcinctus*) no Estado do Pará. *Rev. Inst. Med. Trop. São Paulo* 28, 19-27.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Panchal, M., 2007. The automation of Nested Clade Phylogeographic Analysis. *Bioinformatics*, 23, 509-510.

- Posada, D., Crandall, K.A., Templeton, A.R., 2000. GeoDis: A program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9(4), 487-488.
- Purtilo, D.T., Walsh, G.P., Storrs, E.E, Gannon, C., 1975. The immune system of the nine-banded armadillo (*Dasypus novemcinctus* Linn). *Anat. Rec. Philad.* 181, 725-734.
- Ramos-e-Silva, M., Vilela, F.A.S., Brito, A.C., Carneiro, S.C., 2009. Lobomycosis. Literature review and future perspectives. *Actas Dermosifiliogr.* 100, 92-100.
- Ricci, G., Mota, F.T., Wakamatsu, A., Serafim, R.C., Borra, R.C., Franco, M., 2004. Canine paracoccidioidomycosis. *Med. Mycol.* 42, 379-83.
- Richini-Pereira, V. B. ; Bosco, S.M.G. ; Griese, J ; Theodoro, R. C. ; Macoris, S.A.G.; da Silva, RJ ; Barrozo, L. ; Tavares, PM ; Zancoppe-Oliveira, RM ; Bagagli, E., 2007. Molecular detection of *Paracoccidioides brasiliensis* in road-killed wild animals. *Med. Mycol.* 18, 1-6.
- Richini-Pereira, V.B.; Bosco, S.M.G. ; Theodoro, R. C. ; Macoris, S.A.G.; Bagagli, E., 2009. Molecular approaches for eco-epidemiological studies of *Paracoccidioides brasiliensis*. *Mem. Inst. Oswaldo Cruz* 104, 636-643.
- Ronquist, F., 1996. DIVA version 1.1 Computer Program and Manual Available by Anonymous FTP from Uppsala University. <http://www.ebc.uu.se/systzoo/research/diva/diva.html>.
- Salgado-Salazar, C., Jones, L.R., Restrepo, A., McEwen, J.G., 2010. The human fungal pathogen *Paracoccidioides brasiliensis* (Onygenales: Ajellomycetaceae) is a complex of two species: phylogenetic evidence from five mitochondrial markers. *Cladistics* 26, 1-12.
- Sheff, K.W., York, E.R., Driebe, E.M., Barker, B.M., Rounsley, S.D., Waddell, V.G., Beckstrom-Sternberg, S.M., Beckstrom-Sternberg, J.S., Keim, P.S., Engelthaler, D.M., 2010. Development of a rapid, cost-effective TaqMan Real-Time PCR Assay for identification and differentiation of *Coccidioides immitis* and *Coccidioides posadasii*, *Medical Mycology*, 48, 466–469.
- Silva-Vergara, M.L., Martinez, R., Camargo, Z.P., 2000. Isolation of *Paracoccidioides brasiliensis* from armadillos (*Dasypus novemcinctus*) in an area where the fungus was recently isolated from soil. *Med. Mycol.* 38, 193-199.
- Symmers, W.S.T.C., 1983. A possible case of Lobo's disease acquired in Europe from a bottle-nosed dolphin (*Tursiops truncatus*). *Bull Soc. Pathol. Exot. Filiales* 76, 777-784.
- Taborda, P.R., Taborda, V.A., McGinnis, M.R., 1999. *Lacazia loboi*, the etiologic agent of lobomycosis. *J. Clin. Microbiol.* 37, 2031-2033.
- Takayama, A., Itano, E.N., Sano, A., Ono, M.A., Kamei, K., 2009. An atypical *Paracoccidioides brasiliensis* clinical isolate based on multiple gene analysis. *Med. Mycol.*, 99999:1, DOI: 10.1080/13693780902718065

- Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.*, 10, 512-526.
- Taylor, J. W., Jacobson, D.J., Kroken, S., Kasuga, T., Gêiser, D.M., Hibbett, D.S., Fisher, M.C., 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Gen. Biol.* 31(1), 21-32.
- 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., 2009. Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides* genus. *Mol Phylogenet Evol.* 52, 273-283.
- Terçarioli, G.R., Theodoro, R.C., Bosco, S.M.G., REIS, G.M., Macoris, A.S.G., Simões, L.B., Bagagli, E., 2007. Ecological study of *Paracoccidioides brasiliensis* in soil: growth ability, conidia production and molecular detection.. *BMC Microbiology (Online)*, 7:92.
- Theodoro, R.C., Bagagli, E., Oliveira, C., 2008a. Phylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complex. *Fungal Genet. Biol.* 45: 1284-1291.
- Theodoro, R. C., Bosco, S. M. G., Araújo, J. P. JR., Candeias, J. M., Macoris S. A. G., Trinca, L. A., Bagagli, E., 2008b. Dimorphism, Thermal Tolerance, Virulence and Heat Shock Protein 70 Transcription in different isolates of *Paracoccidioides brasiliensis*. *Mycopathologia.* 165, 355-365.
- Ulrich, M., Convit, J., Centeno, M., Rapetti, M., 1976. Immunological characteristics of the armadillo, *Dasypus sabanicola*. *Clin. Exp. Immunol.* 25, 170-176.
- Untereiner, W.A, Scott, J.A., Naveau, F.A., Sigler, L., Bachewish, J., Angus, A., 2004. The Ajellomycetaceae, a new family of vertebrate-associate Onygenales. *Mycol.* 96(4), 812-821.
- Velazco, P.M., Patterson, B.D., 2008. Phylogenetics and biogeography of the broad-nosed bats, genus *Platyrrhinus* (Chiroptera: Phyllostomidae). *Mol. Phylogenet. Evol.* 49, 749-759.
- Vilela, R., Rosa P.S., Belone, A.F., Taylor, J.W., Diorio, S.M., Mendoza, L., 2009. Molecular phylogeny of animal pathogen *Lacazia loboi* inferred from rDNA and DNA coding sequences. *Mycol. Res.* 113, 851-857.
- Weir, J.T., 2006. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. *Evolution* 60, 842–855.
- Wesselingh, F.P., Räsänen, M.E., Irion, G., Vonhof, H.B., Kaandorp, R., Renema, W., Romero Pittman, L., Gingras, M., 2002. Lake Pebas: A paleoecological reconstruction of a Miocene, long-lived lake complex in western Amazonia. *Cainozoic Res*, 1, 35-81.
- Wesselingh, F.P., 2006. Miocene long-lived lake Pebas as a stage of mollusc radiations, with implications for landscape evolution in western Amazonia. *Scripta Geol.* 133, 1–17.

Wesselingh, F.P., Salo, J.A., 2006. A Miocene perspective on the evolution of the Amazonian biota. *Scripta Geol.* 133, 439–458.

Figure Legends

Figure 1: Dispersal and vicariance analysis by DIVA

The ancestral areas are represented in the nodes of the tree constructed by Neighbor Joining in Mega 4.0v software using *ARF* sequences from the different species of *Paracoccidioides* genus (haplotypes 1-7) and from other Ajellomycetaceae (8-10).

Figure 2: Real Time PCR with TaqMan probes for SNP detection in isolates belonging to the different species from *Paracoccidioides* genus.

End point fluorescence intensity graphics for: SNP3-*ARF* (G for S1, PS3 and *P. lutzii* isolates, and A for PS2 isolates), SNP4-*GP43*, exon 2 (G for S1, PS2 and *P. lutzii*, and C for PS3) and SNP5-*PRP8* intein (A for S1, PS2 and PS3, and G for *P. lutzii*). The green dots represent “heterozygote” additional controls, using two different DNA samples together, the black squares are negative controls and the “X” means undetermined samples.

Figure 3: Geographic distribution of *Paracoccidioides* genus and *L. loboii*

Current distribution of *Paracoccidioides* species S1, PS2, PS3 and *P. lutzii*, and their sister species *L. loboii*, in South America (modified from Richini et al., 2009).

Figure 4: Neighbor Joining tree of the ten haplotypes used in DIVA analysis.

Phylogenetic Neighbor Joining tree with the divergence time calculated under Tamura Nei distance model and *ARF* substitution rate of 0.86×10^{-9} estimated by Kasuga et al., 2003.

Legends of figures from Supplementary Material:

SM 3: Yeast morphology

Aspects of yeast cells from isolates belonging to *P. lutzii* (A), PS2 (B), PS3 (C) and S1 (D and D') species.

SM 4: Conidia morphology

Production of conidia and its morphological aspects in isolates belonging to S1 (A and A'), PS2 (B) and *P. lutzii* species (C).

Tables:

Table 1: isolates from *Paracoccidioides* genus used in this work for morphological and molecular analysis and their species identification.

Species	Isolate	Origin	Host	Isolation source	Method for species identification*		
					MLST	SNaPshot	Real Time PCR - TaqMan
S1	T1F1	Pratânea, SP, Brazil	<i>D. novencinctus</i>	Liver	x		
	T3B6	Pratânea, SP, Brazil	<i>D. novencinctus</i>	Spleen	x	x	x
	T4B17	Manduri, SP, Brazil	<i>D. novencinctus</i>	Spleen	x		x
	T5LN1	Botucatu, SP, Brazil	<i>D. novencinctus</i>	Mesenteric lymph node	x		
	T7F6	Manduri, SP, Brazil	<i>D. novencinctus</i>	Liver	x		
	T8LN2	Botucatu, SP, Brazil	<i>D. novencinctus</i>	Mesenteric lymph node	x		
	T9B1	Botucatu, SP, Brazil	<i>D. novencinctus</i>	Spleen	x		
	T13LN2	Manduri, SP, Brazil	<i>D. novencinctus</i>	Mesenteric lymph node	x		
	T15LN1	Manduri, SP, Brazil	<i>D. novencinctus</i>	Mesenteric lymph node	x		
	Bt60	Botucatu, SP, Brazil	Human	NK	x		
	Pb18	NK	Human	NK	x		
	Pb339	Brazil	Human	NK	x		
	S1	São Manoel, SP, Brazil	<i>D. novencinctus</i>	Spleen		x	x
	ILSL57	Bauru, SP, Brazil	<i>D. novencinctus</i>	Spleen		x	x
	Pb-baby	Ribeirão Preto, SP, Brazil	Human	NK		x	x
	Bt85	Botucatu, SP, Brazil	Human	NK		x	x
	Pb265	Botucatu, SP, Brazil	Human	NK		x	x
	D01	Avaré, SP, Brazil	Human	cutaneous lesion		x	x
	D02	Laranjal Paulista, SP, Brazil	Human	cutaneous lesion		x	x
	D03	Piracicaba, SP, Brazil	Human	cutaneous lesion		x	x
	D04	Pereiras, SP, Brazil	Human	cutaneous lesion		x	x
	D05	Botucatu, SP, Brazil	Human	cutaneous lesion		x	x
	D06	Elias Fausto, SP, Brazil	Human	cutaneous lesion		x	x
	D07	Assis, SP, Brazil	Human	cutaneous lesion		x	x
	D09	Limeira, SP, Brazil	Human	cutaneous lesion		x	x
	D10	Riversul, SP, Brazil	Human	cutaneous lesion		x	x
	D11	Iperó, SP, Brazil	Human	cutaneous lesion		x	x
	EPM04	Pará, Brazil	<i>D. novencinctus</i>	NK	x	x	x
	EPM46	Argentina	Human	NK		x	x
	EPM48	Argentina	Human	NK		x	x
	EPM85	Peru	Human	Lips	x	x	x
	EPM75	Venezuela (1994)	Human	mouth		x	x
	EPM101	Ibiá, MG, Brazil	<i>D. novencinctus</i>	NK		x	x
	Pb21	Bauru, SP, Brazil, 1964	Human	NK		x	x
	Pb534	Barquissimeto, Venezuela	Human	NK		x	x
	Pb113	Manaus-AM, Brazil, 1971	Human	NK		x	x
	Pb157	Uruguay	Human	NK		x	x
	Pb94	Belo Horizonte, MG, 1971	Human	NK		x	x
	Pb135	PA, Brazil	NK	NK		x	x
	Pb25	MG, Brazil	Human 19 anos	NK		x	x
Pb09	Uruguay, 1925	Human	NK		x	x	
Pb728	Venezuela, 1982	Human	NK		x	x	
Pb698	Venezuela, 1982	Human	NK		x	x	
TCC	Cerqueira César, SP, Brazil	<i>D. novencinctus</i>	Mesenteric lymph node		x	x	
PS2	T10B1	Botucatu, SP, Brazil	<i>D. novencinctus</i>	Spleen	x	x	x
	Pb927 ***	Antarctica/ Uruguay	Pinguin	feces	x	x	x
	Bt84	Botucatu, SP, Brazil	Human	NK	x	x	x
	Pb262	Uberlândia, MG, Brazil	-	Dog food	x	x	x
	Pb02	Caracas, Venezuela	Human	NK	x		
	Pb03	São Paulo, SP, Brazil	Human	NK	x		
	Pb04	São Paulo, SP, Brazil	Human	NK	x		x
	Pbdog	Curitiba, PR, Brazil	<i>Canis familiaris</i>	Popliteal lymph node		x	x
	1430	Santa Maria, RS, Brazil	Human	Tongue		x	
	Pb106	Piraju, SP, Brazil, 1970	Human, 17 anos	NK		x	x
Pb22	Bauru, SP, Brazil, 1964	Human	NK		x	x	

Table 1, Continuation

Species	Isolate	Origin	Host	Isolation source	Method for species identification*		
					MLST	SNaPshot	Real Time PCR - TaqMan
PS3	EPM83	Colombia (CIB)	Human	NK	x	x	x
	EPM77	Colombia (CIB)	NK	NK	x		x
	60855	Colombia (CIB)	NK	NK	x	x	
<i>P. lutzii</i>	Pb01	Goiania, GO, Brazil	Human	NK	x	x	x
	Pb66	Goiania, GO, Brazil	Human	NK	x	x	x
	8334	Goiania, GO, Brazil	Human	NK	x	x	x
	JAL**	Cuiabá, MT, Brazil	Human	NK	x	x	x
	EE	Cuiabá, MT, Brazil	Human	NK	x	x	x

In blue: isolates used for morphological analysis. * The MLST was previously carried out by Matute et al (2006) and Teixeira et al (2009). The SNPs detection was performed in this work. **Isolate whose morphometry of yeast cell was not studied due to the slow mycelia-yeast transition. *** Isolate identified as S1 by Matute et al (2006), but clustered with PS2 isolates according to the SNPs analysis and sequencing of the intein PRP8 (Theodoro et al., 2008). NK: Not Known.

Table 2: SNPs evaluated by SNaPshot and Real Time PCR techniques; primers and probes used and expected SNP for each species.

Gene	SNP position*	SNaPshot Primer	Real Time PCR - Taqman				SNPs/ species			
			VIC probe	FAM probe	Forward primer (5'-3')	Reverse Primer (5'-3')	S1	PS2	PS3	<i>P. lutzii</i>
GP43-exon2	344	4(C)GTATCTAAAATG GGCGTGGCA	-	-	-	-	T	T	T	G
	494	-	CATCCAGCCGGACCT	CATCCAGCCCGACCT	GCGCTTGTGTGCCGA	CTCCGICTTCCATGTCCA GAA	G	G	C	G
ARF**	298	2(C)CATCCAATTTA CCGTGTGGGA	-	-	-	-	C	T	C	C
		-	CCACCCACGTCCACACA	CCACCCACATCCACACA	GCAGGATTTAACGTCTG AGACTGT	GCTGTCAACCACCGAAAAT GATACCT	G	A	G	G
PRP8 intein	1146	-	TCCTCCTCACCCAGGACA	CTCCTCGCCAGGACA	GCGAACTGGTTGTCGA ACGTA	GGCCAACGATAATCTTCAA GGATTTC	A	A	A	G

*According to the sequences deposited by Matute et al (2006) and Theodoro et al (2008) in GenBank. ** The same SNP was, in this case, studied by using the SNaPshot and Real Time PCR (TaqMan) methodologies, in the two DNA strands.

Table 3: Conidia production by isolates belonging to the four species of *Paracoccidioides* genus

Species	Isolate	Conidia production			
		Adhesive tape	PDA Slide Culture	Adhesive tape	SEA Slide Culture
S1	T1F1	-	-	-	-
	T3B6	-	+	++++	++++
	T4B17	-	+	++++	++++
	T5LN1	-	-	-	-
	T7F6	-	++	++++	+++
	T8LN2	-	++	++	++
	T9B1	-	++	++++	++++
	T13LN2	-	-	-	-
	T15LN1	-	++	++++	++++
	BT60	-	-	++	+
	PB18	-	-	-	-
	PB339	-	-	-	-
	S1	+	+	++	+++
	ILSL57	-	+	++++	++++
	BT85	-	-	-	+
	PB265	-	-	-	-
	D01	-	-	-	+
	D02	-	-	-	-
	D03	-	-	+++	+++
	D04	-	-	++	++
	D05	-	-	++	+
	D06	-	-	-	-
	D07	-	-	+	++++
	D09	-	-	+	+
	D10	-	-	++++	+++
	D11	-	-	++++	++
	T10B1	-	-	-	-
	PB927	-	-	-	-
BT84	-	-	-	-	
PB262	-	-	-	-	
PS2	PB02	-	-	+	-
	PB03	-	-	-	-
	PB04	-	-	-	-
	PBdog	+++	++	++++	++++
	1430	-	-	-	-
PS3	EPM83	-	-	-	-
	EPM77	-	-	-	-
<i>P. lutzii</i>	PB01	-	-	-	+
	EE	-	-	+++	+++
	Jal	-	-	-	++
	Pb8334	-	-	-	-
	Pb66	-	-	-	-

- 0
 + up to 5
 ++ from 5 to 20
 +++ from 20 to 50
 ++++ >50 conidia per microscopic field

Table 4: SNPs detected by SNaPshot and Real Time PCR (TaqMan) for different isolates from *Paracoccidioides* genus.

Species	Isolate	SNaPshot		Real Time PCR (TaqMan)		
		SNP1 <i>ARF</i>	SNP2 <i>Exon2-GP43</i>	SNP3 <i>ARF</i>	SNP4 <i>Exon2-GP43</i>	SNP5 <i>PRP8 intein</i>
S1	T1F1*	C	T	G	NA	A
	T3B6*	C	T	NA	G	NA
	T4B17*	C	T	NA	G	NA
	T15LN1*	NA	NA	G	NA	A
	S1	C	T	G	G	A
	ILSL57	C	T	G	G	A
	Pb-baby	C	T	G	G	A
	Bt85	C	T	G	G	A
	Pb265	C	T	G	G	A
	D01	C	T	G	G	A
	D02	C	T	G	G	A
	D03	C	T	G	G	A
	D04	C	T	G	G	A
	D05	C	T	G	G	A
	D06	C	T	G	G	A
	D07	C	T	G	G	A
	D09	C	T	NA	G	NA
	D10	C	T	G	G	A
	D11	C	T	G	G	A
	EPM04*	C	T	G	G	A
	EPM46	C	T	G	G	A
	EPM48	C	T	G	G	A
	EPM85*	C	T	G	G	A
	EPM75	C	T	G	G	A
	EPM101	C	T	G	G	A
	Pb21	C	T	G	G	A
	Pb534	C	T	G	G	A
	Pb113	C	T	G	NS	A
	Pb157	C	T	G	G	A
	Pb94	C	T	G	G	A
	Pb135	C	T	G	G	A
	Pb25	C	T	G	G	A
	Pb09	C	T	NA	G	NA
Pb728	C	T	G	G	A	
Pb698	C	T	G	G	A	
TCC	C	T	G	G	A	
PS2	Pb927 **	T	T	A	NA	NA
	T10B1*	T	T	A	NS	A
	Bt84*	T	T	A	G	NA
	Pb262*	T	T	A	NA	NA
	Pb04*	T	T	A	NA	A
	Pbdog	T	T	A	NA	A
	1430	T	T	A	NA	A
	Pb106	T	T	A	NA	NA
	Pb22	T	T	A	NA	NA
PS3	EPM83*	C	T	G	C	A
	EPM77*	C	T	G	C	A
	60855*	C	T	NA	NA	NA
<i>P. lutzii</i>	Pb01*	C	G	NA	G	NA
	Pb66*	C	G	NA	G	NA
	8334*	C	G	G	G	G
	JAL *	C	G	NA	G	NA
	EE *	C	G	G	G	G

* Isolates used as positive controls. ** Isolate previously grouped in S1 species (Matute et al., 2006), but recognized as PS2 species in this study, corroborating the data found by PRP8 intein sequencing (Theodoro et al., 2008b). NA= Not Analyzed ND= Not Determined by the method.

Table 5: Historical events associated with *Paracoccidioides* genus biogeography, according to Nested Clade Analysis.

Locus	Historical events with statistical significance
	<u>Within S1/PS3</u>
<i>ARF</i>	- Simultaneous geographic expansions - Restricted gene flow with isolation by distance/ Restricted dispersal by distance in non sexual species (PS3).
<i>TUB</i>	- Gene flow/ Restricted dispersal, but with some long distance dispersal events over intermediate areas. - Long distance colonization and/ or fragmentation event separating Venezuelan isolates from others S1.
<i>GP43</i>	- Restricted gene flow with isolation by distance / Restricted dispersal by distance in non sexual species (PS3).
	<u>Between PS2 and S1/ PS3</u>
<i>ARF</i>	No historical event detected
<i>TUB</i>	No historical event detected
<i>GP43</i>	Inconclusive data
	<u>Between <i>P. lutzii</i> and S1, PS2, PS3</u>
<i>ARF</i>	Inconclusive data
<i>TUB</i>	No historical event detected
<i>GP43</i>	Inconclusive data

Figures

Figure 1

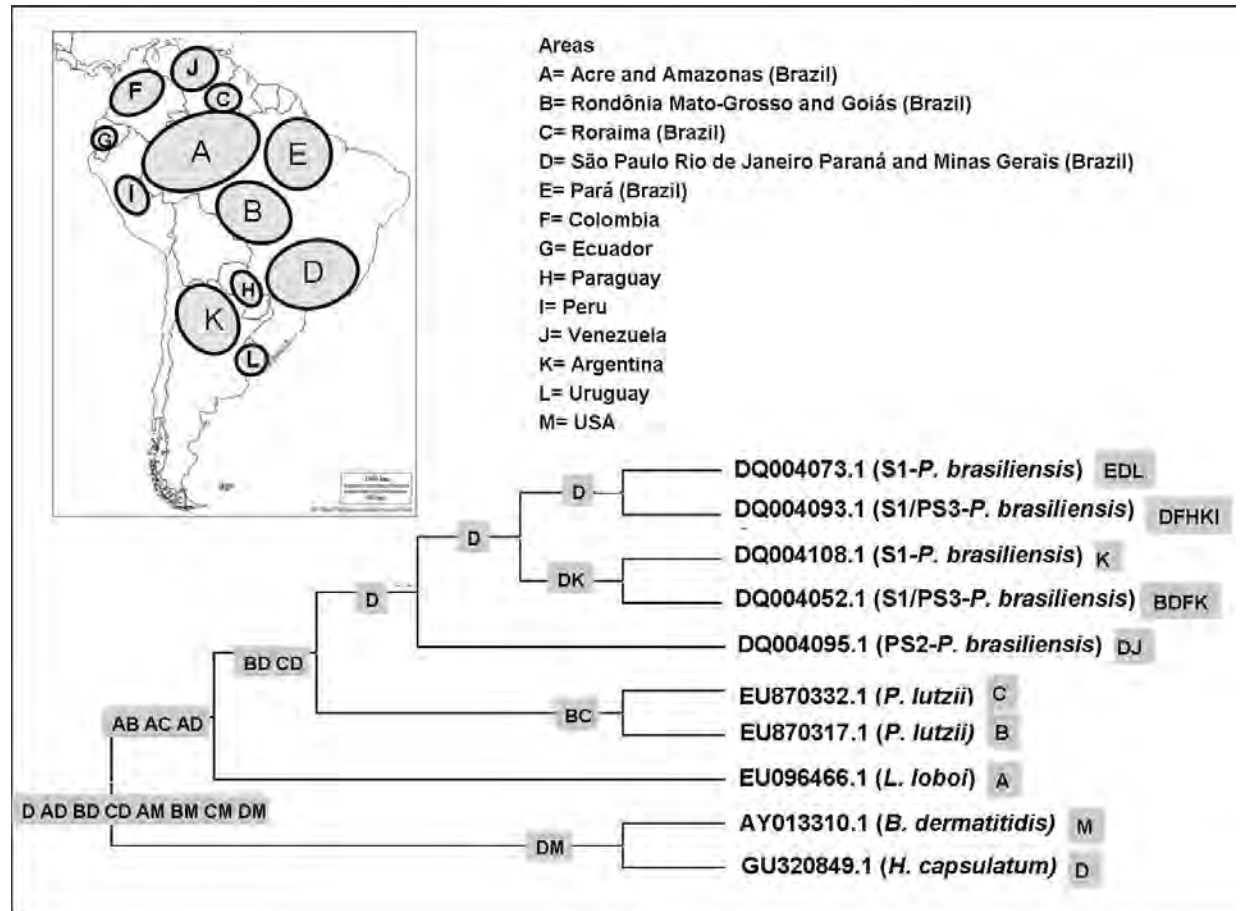


Figure 2

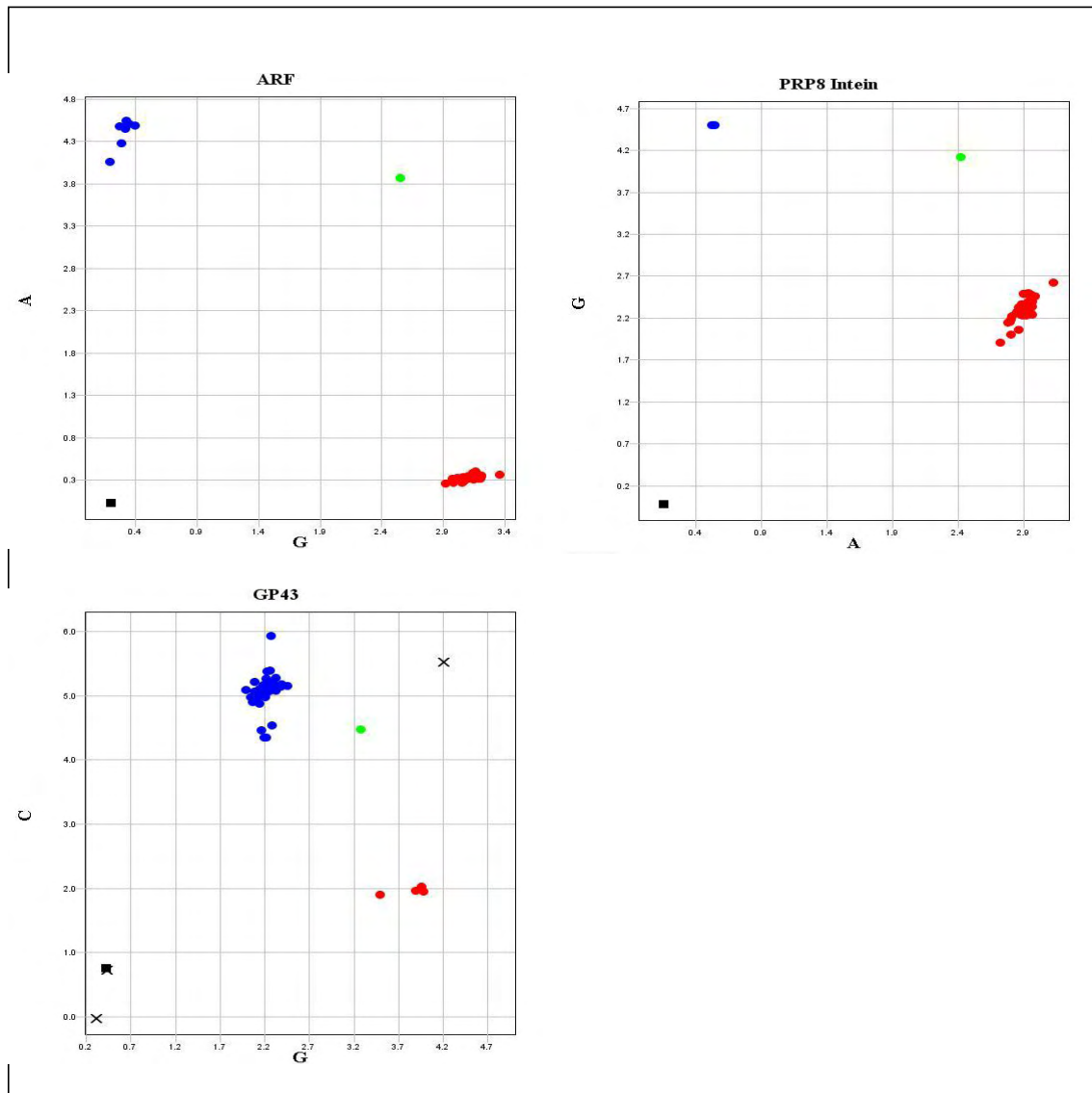


Figure 3

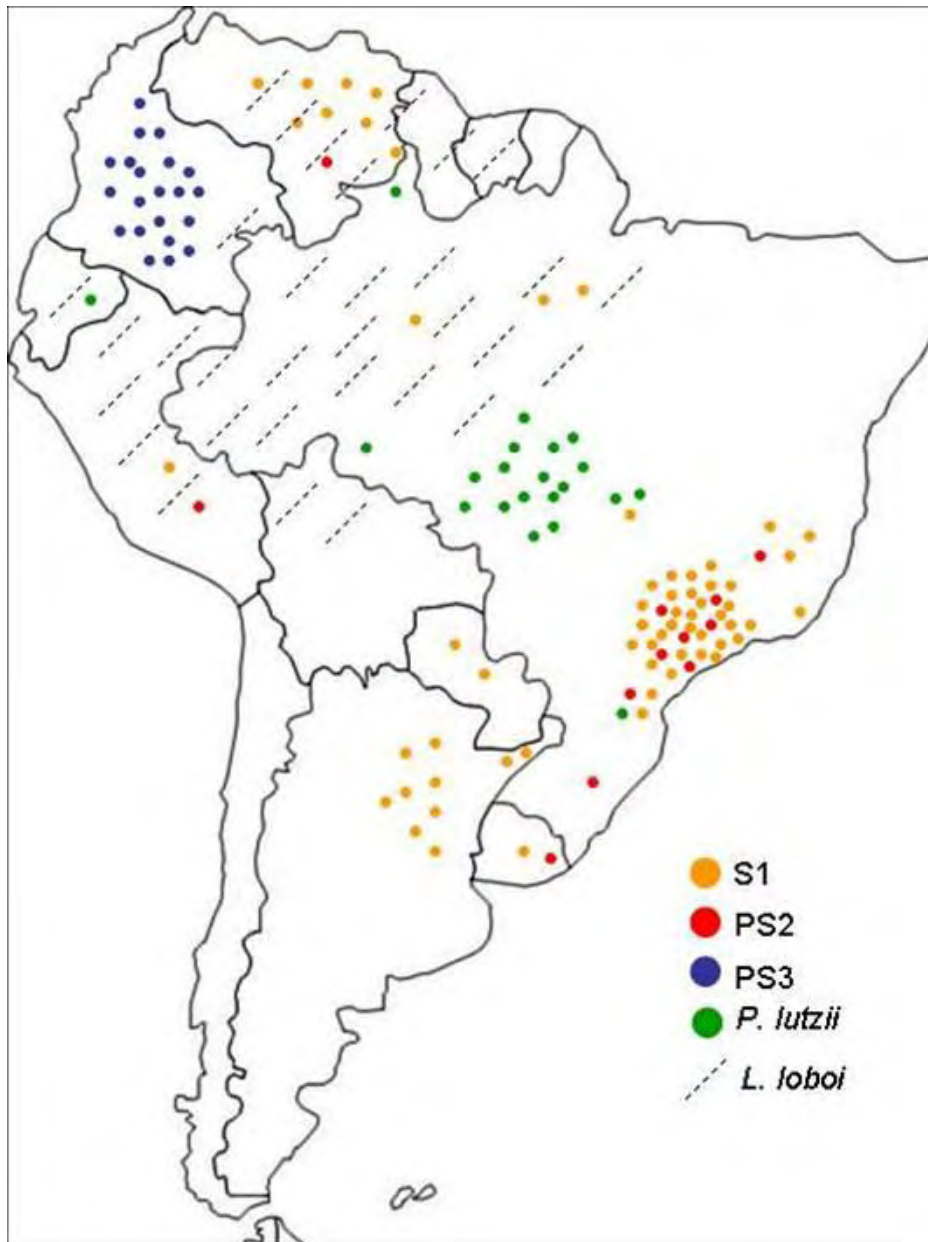
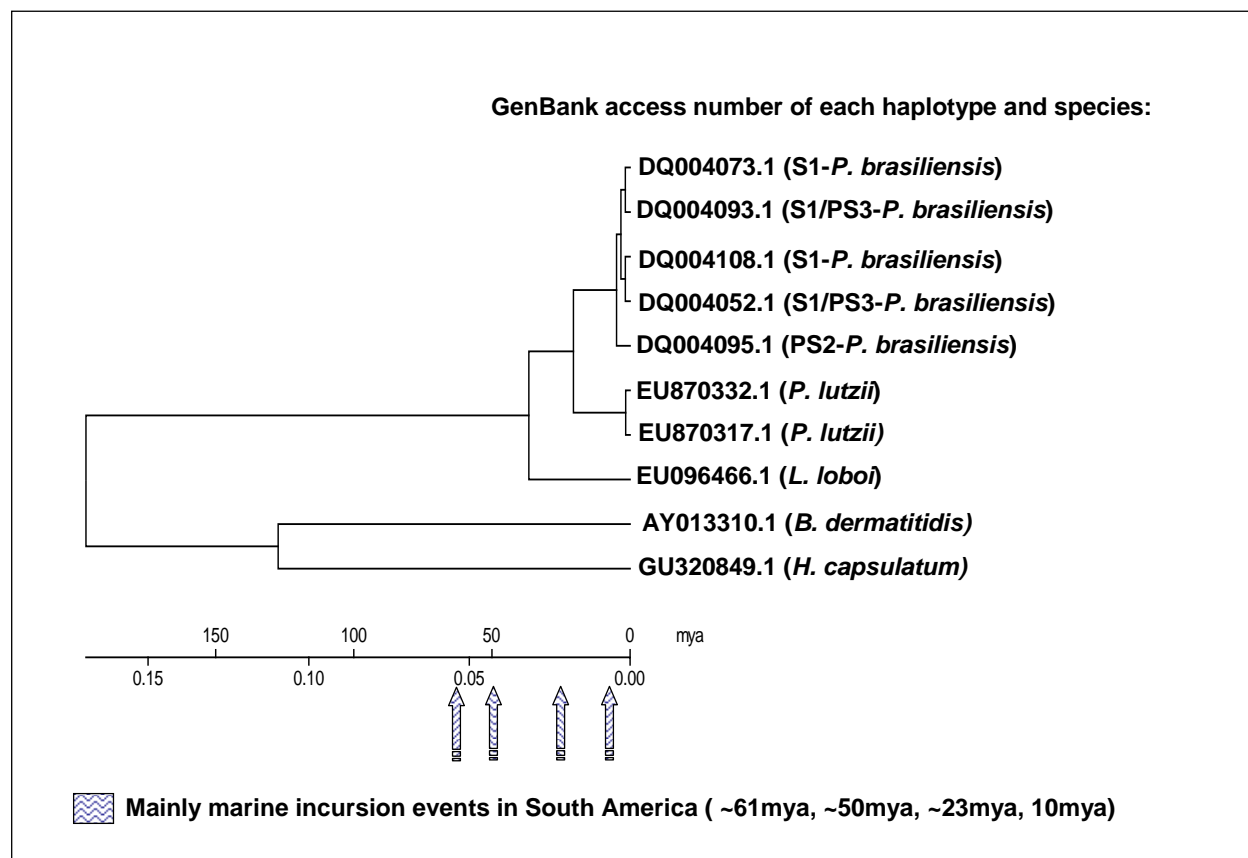


Figure 4



Supplementary Material

SM1: primers used to amplify the genes containing the SNPs evaluated by SNaPshot technique.

Gene	Primer	Sequence 5'-3'	amplicom	Reference
<i>GP43</i>	<i>GP43</i> -E2-fwd	CTAGAATATCTCACTCCCAG	723pb	This work
	<i>GP43</i> -E2-rev	GCCCCCTCCGTCTTCCATGTCC		Matute et al., 2006
<i>ARF</i>	<i>ARF</i> – fwd	TTCATGGTTGGCCTCGATGCTGCC	407pb	Matute et al., 2006
	<i>ARF</i> -rev	GAGCCTCGACGACACGGTCACGATC		

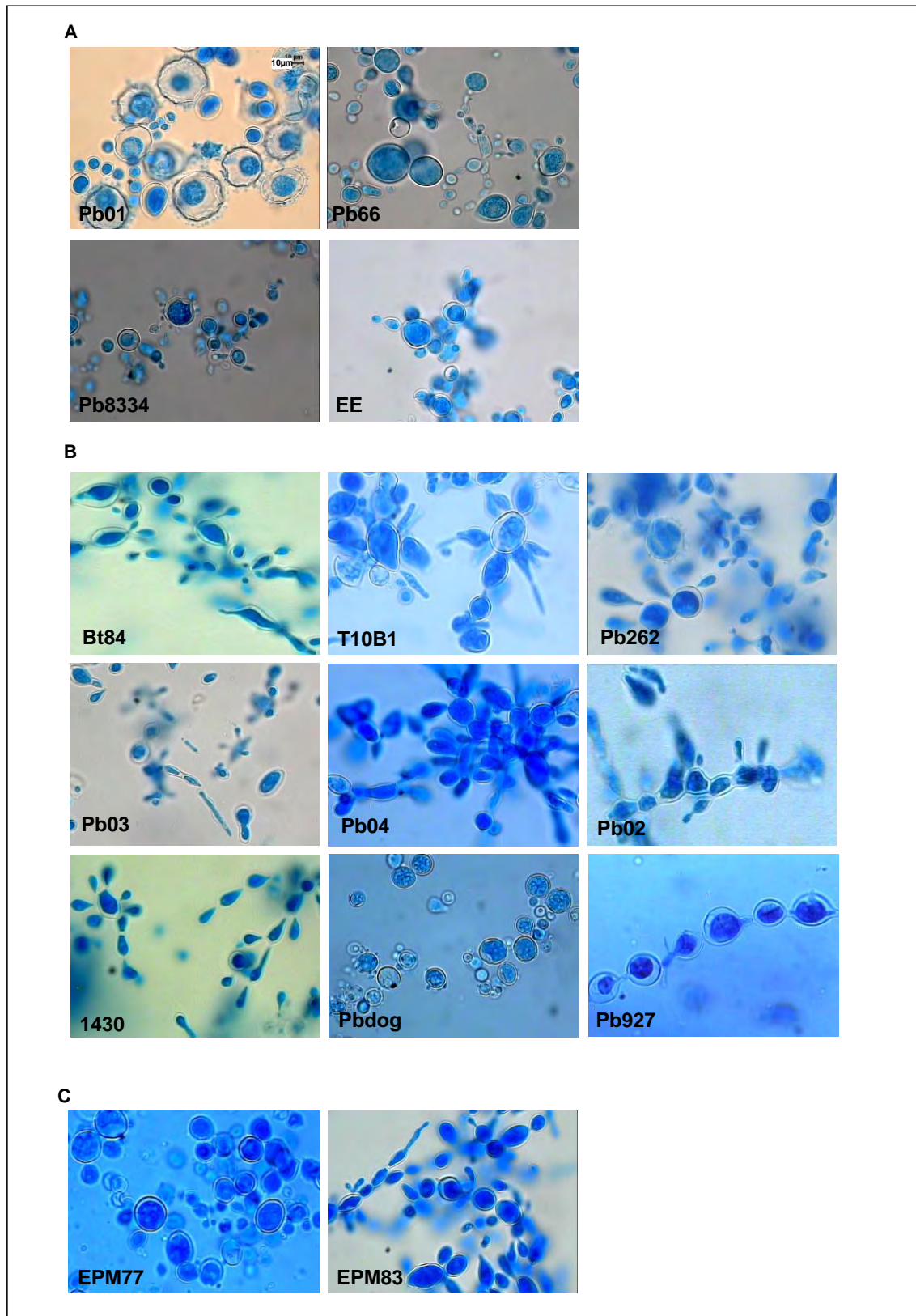
SM2: Paracoccidioides isolates used in Nested Clade Analysis and their respective geographic areas.

Isolate	Reference	Area	Latitude***	Longitude***
pbB1	Matute et al (2006)	São Paulo/ Brazil	-23.55	-46.64
pbB2	Matute et al (2006)			
pbB4	Matute et al (2006)			
pbB3	Matute et al (2006)			
pbB5	Matute et al (2006)			
pbB6	Matute et al (2006)			
pbB7	Matute et al (2006)			
pbB8	Matute et al (2006)			
pbB9	Matute et al (2006)			
pbB10	Matute et al (2006)			
pbB14	Matute et al (2006)			
pbB15	Matute et al (2006)			
pbB17	Matute et al (2006)			
pbB20	Matute et al (2006)			
pbB22	Matute et al (2006)			
pbB23	Matute et al (2006)			
pbB25	Matute et al (2006)			
pbB26	Matute et al (2006)			
pbC1	Matute et al (2006)	Colômbia	4.60	-74.08
pbC2	Matute et al (2006)			
pbC3	Matute et al (2006)			
pbC4	Matute et al (2006)			
pbC5	Matute et al (2006)			
pbC6	Matute et al (2006)			
pbC7	Matute et al (2006)			
pbC8	Matute et al (2006)			
pbC9	Matute et al (2006)			
pbC10	Matute et al (2006)			
pbC11	Matute et al (2006)			
pbC12	Matute et al (2006)			
pbC13	Matute et al (2006)			
pbC14	Matute et al (2006)			
pbC15	Matute et al (2006)			
pbC16	Matute et al (2006)			
pbC17	Matute et al (2006)			
pbC18	Matute et al (2006)			
pbC19	Matute et al (2006)			
pbC21	Matute et al (2006)			
pbC20	Matute et al (2006)			
pbV1	Matute et al (2006)	Venezuela	10.48	-66.91
pbV2	Matute et al (2006)			
pbV3	Matute et al (2006)			
pbV4	Matute et al (2006)			
pbV5	Matute et al (2006)			
pbV6	Matute et al (2006)			
pbA1	Matute et al (2006)	Argentina	-27.42	-58.98
pbA2	Matute et al (2006)			
pbA4	Matute et al (2006)			
pbA5	Matute et al (2006)			
pbA6	Matute et al (2006)			
pbA7	Matute et al (2006)			
pbA8	Matute et al (2006)			
pbA3	Matute et al (2006)			
pbB21	Matute et al (2006)	Paraná/ Brazil	-25.43	-49.23
pbB24	Matute et al (2006)			
pbPbdog	Bosco et al (not published)	Minas Gerais/ Brazil	-19.82	-43.96
pbB12	Matute et al (2006)			
pbB13	Matute et al (2006)			
pbP2	Matute et al (2006)	Paraguay	-25.41	-57.12

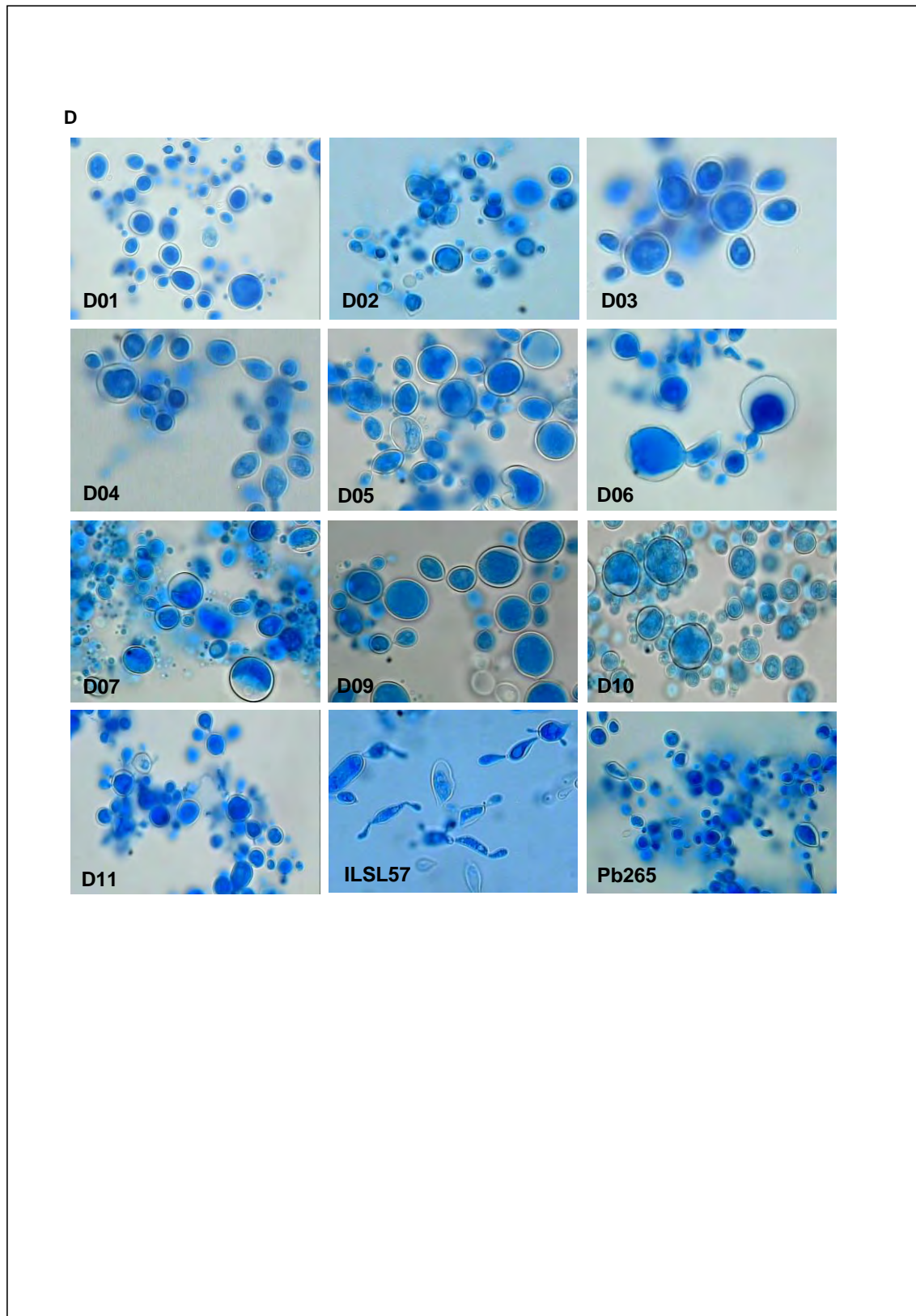
pbP1	Matute et al (2006)			
pbB11	Matute et al (2006)	Pará/ Brazil	-1.46	-48.50
pbB16	Matute et al (2006)	Rio de Janeiro/ Brazil	-22.90	-43.21
pbPE1	Matute et al (2006)	Peru	-8.35	-74.54
pbU1	Matute et al (2006)	Uruguay	-34.94	-55.51
pl7455	Teixeira et al (2009)	Ecuador	-2.21	-79.93
plPIRO1**	Teixeira et al (not published)	Rondônia/ Brazil	-11.44	-61.45
pl769	Teixeira et al (2009)	Roraima/ Brazil	2.74	-60.58
plRaj2	Teixeira et al (2009)	Mato-Grosso/ Brazil	-15.60	-56.10
pl717	Teixeira et al (2009)			
pl133	Teixeira et al (2009)			
pl84	Teixeira et al (2009)			
pl189	Teixeira et al (2009)			
pl61*	Teixeira et al (not published)			
pl206	Teixeira et al (2009)			
pl694*	Teixeira et al (not published)			
pl397	Teixeira et al (2009)			
pl57	Teixeira et al (2009)			
pl351*	Teixeira et al (not published)			
pl218	Teixeira et al (2009)			
plIE**	Teixeira et al (not published)			
plMF**	Teixeira et al (not published)			
plJHS**	Teixeira et al (not published)			
pl1578**	Teixeira et al (not published)	Goiás/ Brazil	-16.68	-49.25
plPb01	Teixeira et al (2009)			
pbB19	Matute et al (2006)			

S1, PS2, PS3 and *P. lutzii* isolates are listed in yellow, red, blue and green, respectively. * Isolates not evaluated for *GP43* gene. ** Isolates not evaluated for both *GP43* and *TUB* gene. ***Geographic coordinates of the main cities for PCM occurrence.

SM3

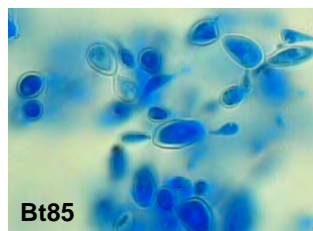
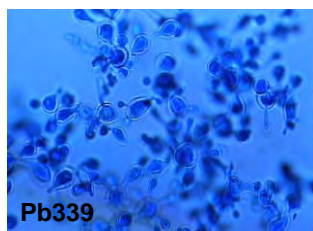
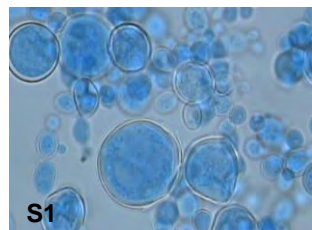
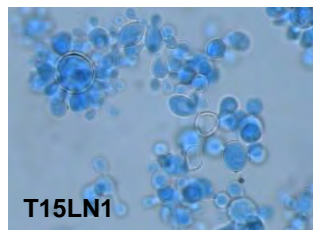
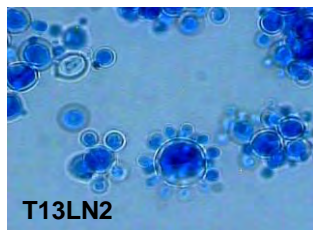
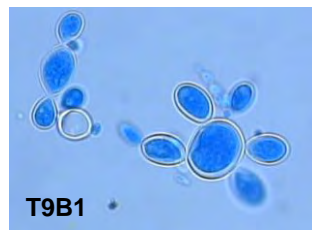
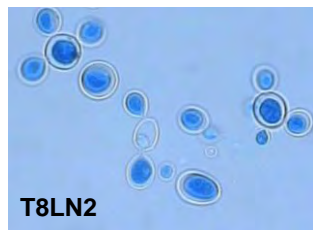
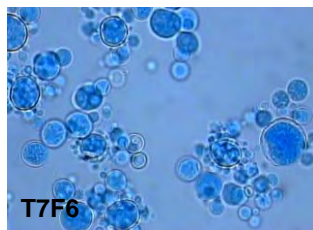
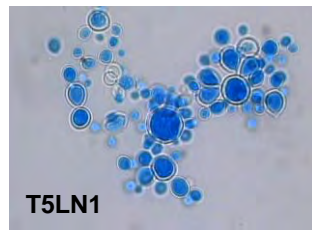
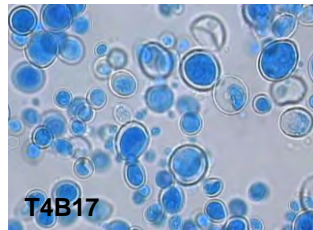
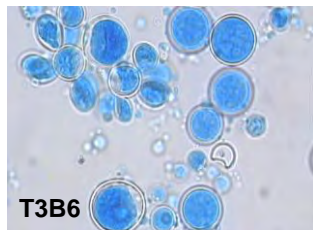
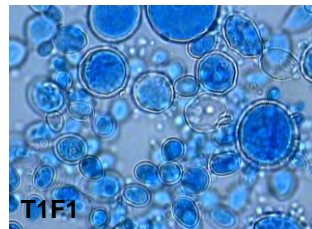
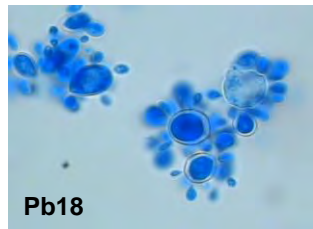
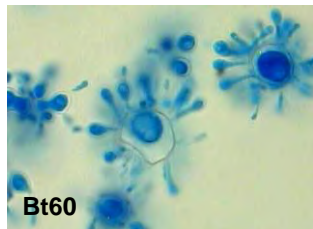


SM3 continuation



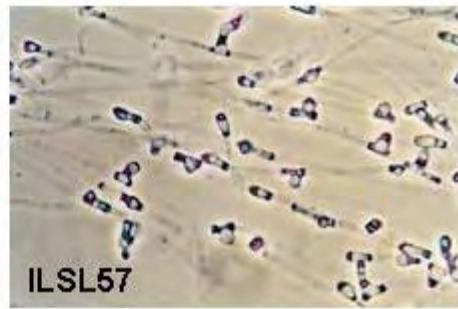
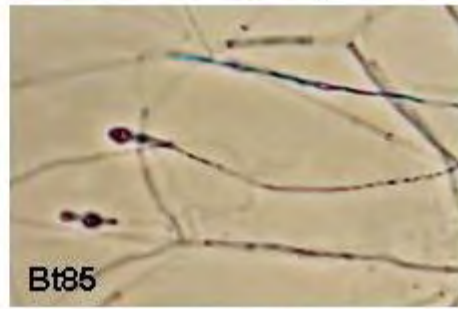
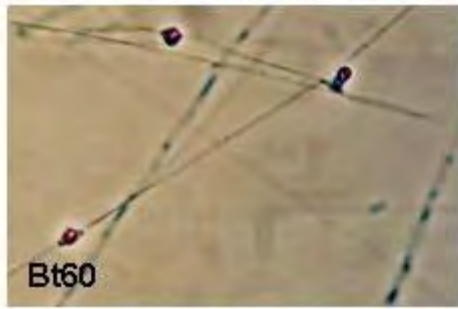
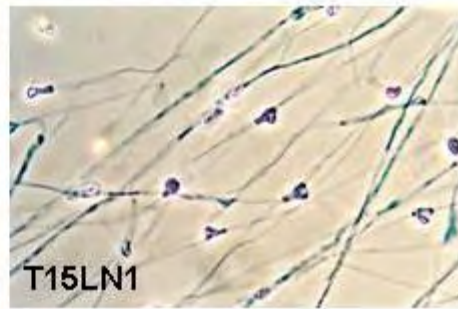
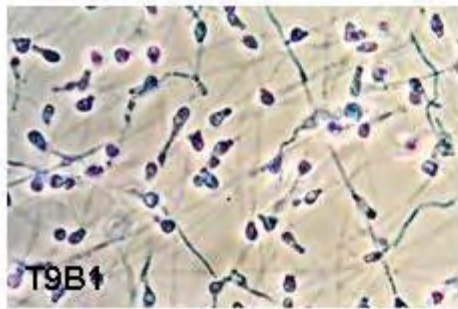
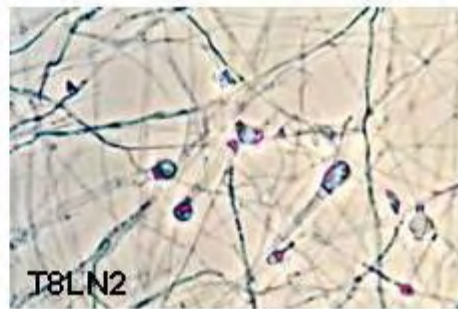
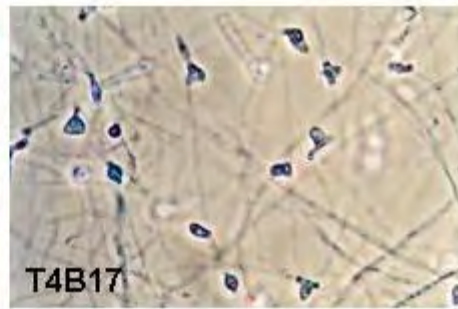
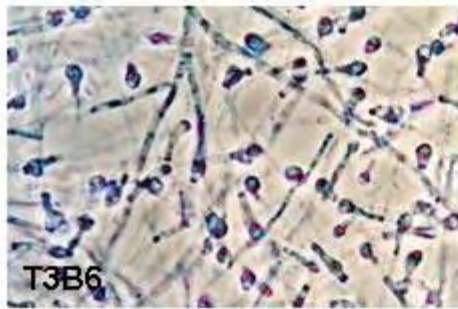
SM3 continuation

D'

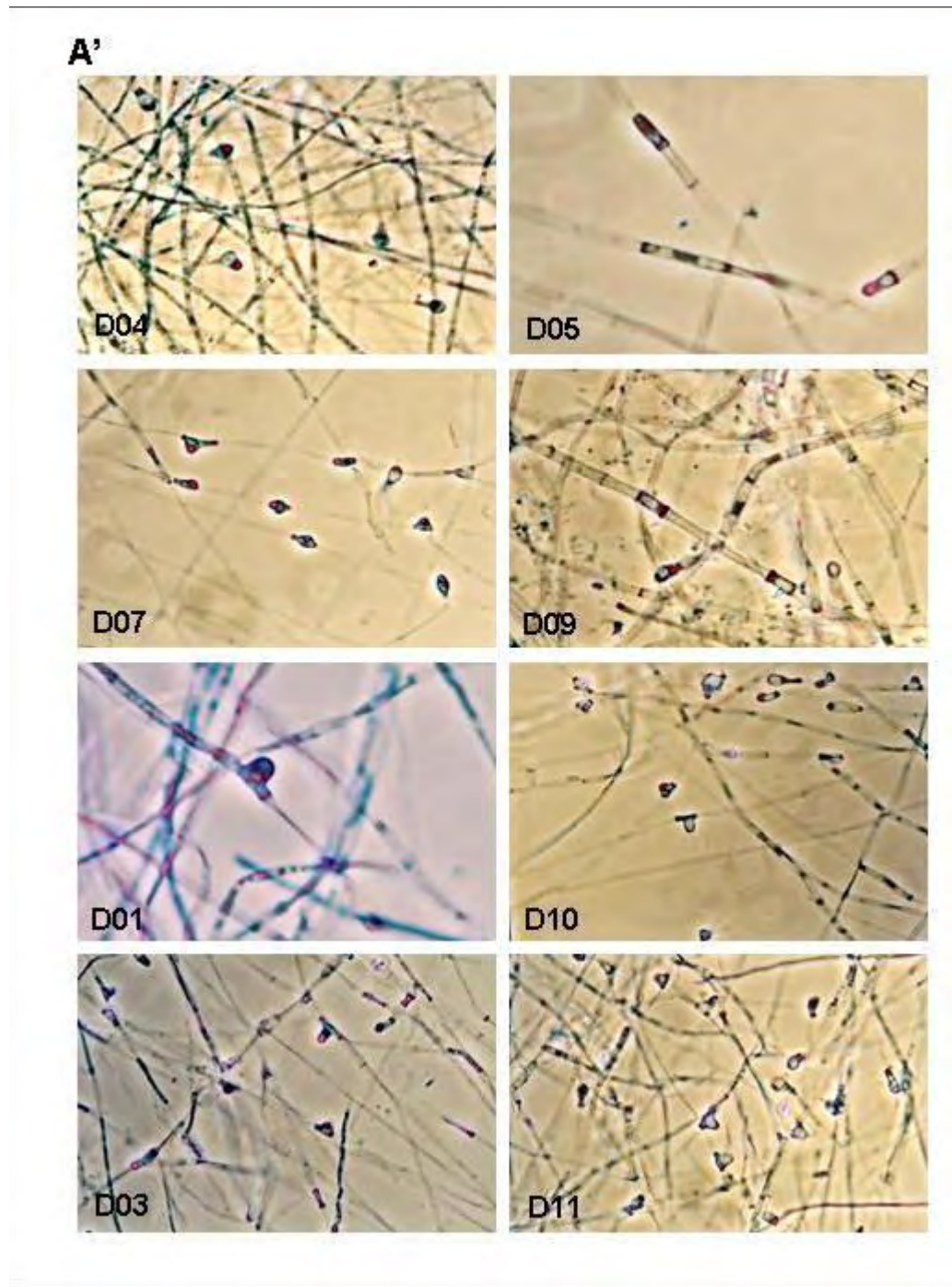


SM4

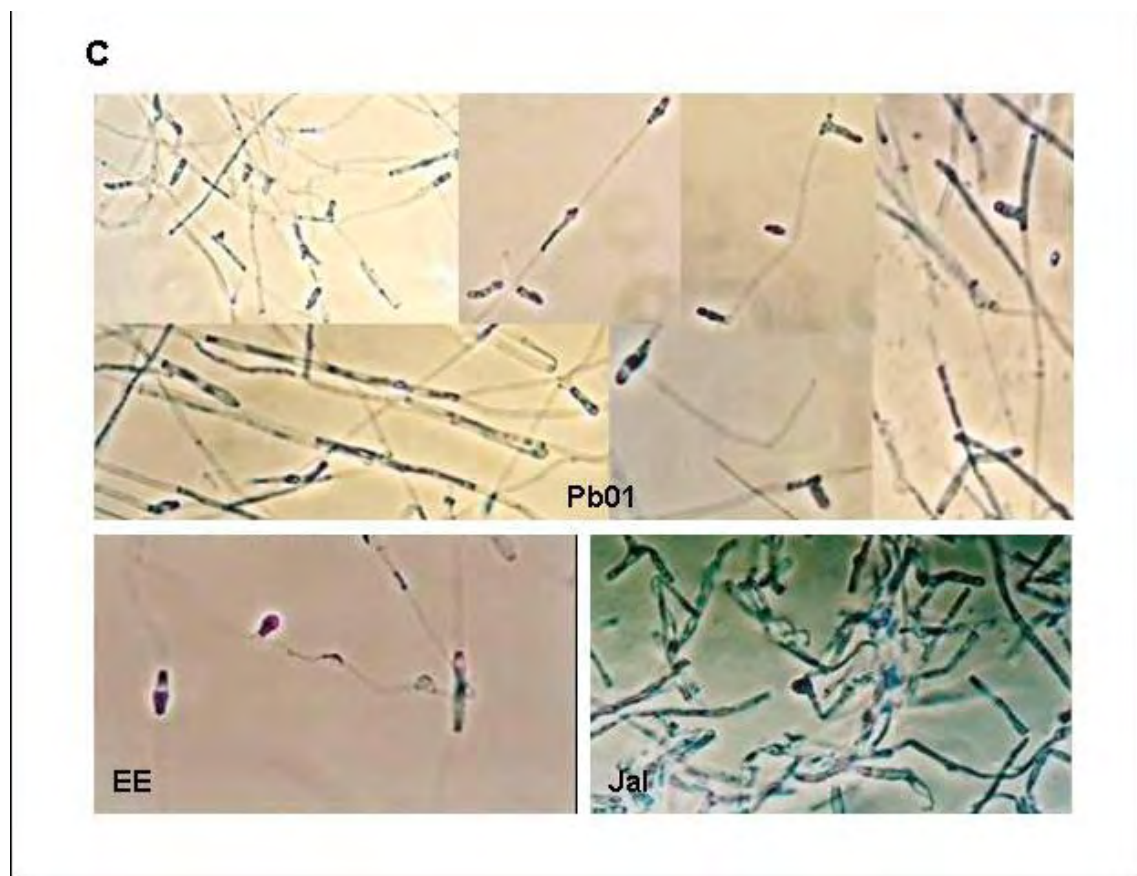
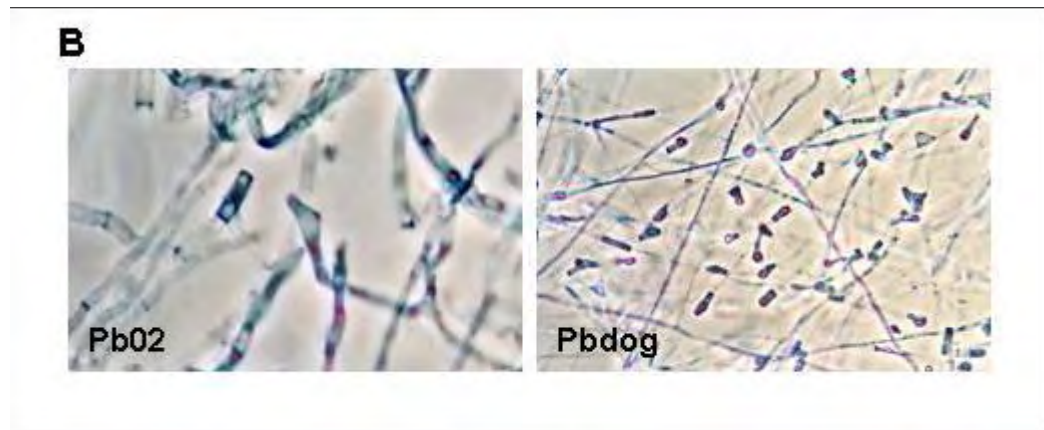
A



SM4 continuation



SM4 continuation



Capítulo 2:

Intein PRP8

*“... os genes são cidadãos do tempo geológico: os genes são
eternos.”*

*“Disse um gene egoísta de passagem,
Tantos corpos já vivi que perdi a contagem
Te julgas tão inteligente
Mas és apenas minha aparelhagem
Para viver eternamente”.*

Richard Dawkins, A grande história da Evolução (*The ancestor tale*).

***Intein*: um elemento genético parasita**

Durante toda a história da biologia molecular, cada avanço ou descoberta científica vem colocando em prova antigos dogmas ou conceitos. Um exemplo disso é o próprio conceito de gene. Sabe-se que descobertas como RNA de interferência, regiões regulatórias localizadas longe da região a ser transcrita, importância do intron na transcrição e *splicing* alternativo, entre outras, inviabilizam a criação de um significado global para “gene”. Com os *inteins* a história não é diferente. Descritos há pouco mais de vinte anos, os *inteins* 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 sequências codificadoras e, diferente dos introns, são transcritos e traduzidos juntamente com o fragmento hospedeiro sofrendo uma excisão auto catalítica pós-traducional, permitindo a funcionalidade normal da proteína hospedeira ou *extein*. Nesta revisão será dada uma introdução sobre *inteins*, no que se refere a sua descoberta, distribuição nos diferentes seres vivos, dinâmica populacional, aquisição de função e perspectivas de aplicações biotecnológicas.

***Intein*: uma nova descoberta da biologia molecular**

De uma forma geral os genes de eucariotos apresentam uma região promotora constituída de elementos curtos conservados situados antes do ponto inicial da transcrição, como por exemplo, o TATAbox, sendo reconhecida por fatores de transcrição necessários para que as RNAs polimerases iniciem a síntese de RNA mensageiro, seguida de regiões conhecidas como exons (sequências expressas) e introns (sequências intercalares não codificadas presentes na maioria dos genes eucarióticos e também presentes nas Archaeobacterias) que durante o processo de maturação do RNAm sofrem *splicing* ou corte. A ponta 3' do gene possui segmentos distais (de 11 a 30 nucleotídeos depois de uma região conservada AAUAAA), para a sinalização de clivagem e poliadenilação da molécula de RNAm

que será polimerizada. Em eucariotos, além da região promotora, outras regiões longe do próprio gene podem atuar na regulação de sua expressão, sendo acentuadores ou silenciadores (Figura 1).

Em procariotos, 1/3 dos genes que codificam proteínas estão organizados em operons, ou seja, várias regiões codificadoras (de produtos protéicos distintos), sem introns, com apenas uma única região promotora e outra operadora (sítio de ligação de proteínas regulatórias ativadoras ou repressoras), sendo estas regiões as mais importantes para a regulação da expressão gênica, mediada por proteínas que reagem a sinais ambientais aumentando ou diminuindo as taxas de transcrição de genes específicos (Figura 1). (Snustad & Simmons, 2000).

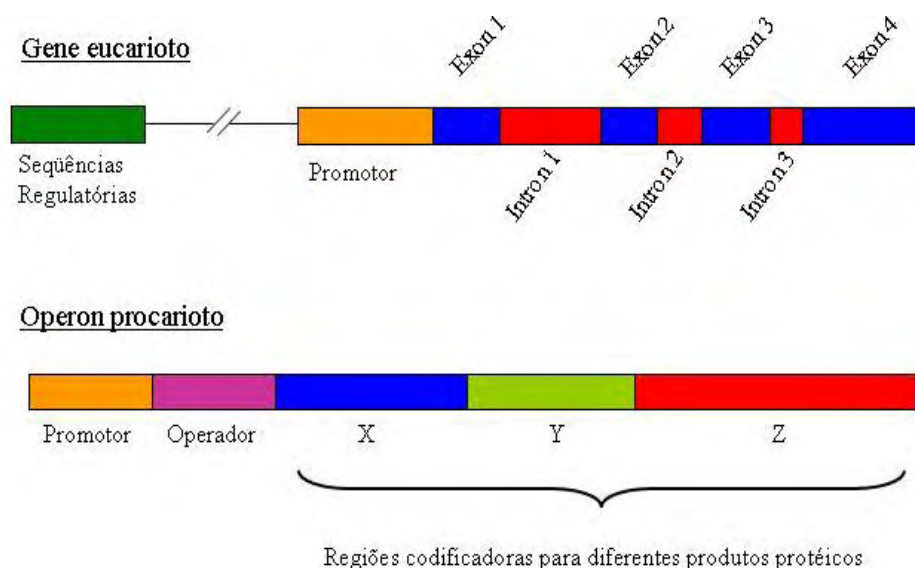


Figura 1: Esquema sobre a estrutura dos genes eucarióticos e procarióticos.

Toda esta teoria a respeito de como funcionam os genes codificadores de proteínas parece deixar bem claro que o RNAm maduro e o produto de sua tradução possuem apenas regiões que de fato codificarão uma proteína funcional. Porém, em 1987, quando a ATPase vacuolar de *Saccharomyces cerevisiae* foi comparada com sequências de outras ATPases, observou-se que sua região central não apresentava homologia com nenhuma outra sequência

ATPase até então depositada nos bancos de dados, mas possuía certa similaridade com algumas endonucleases (Hirata et al., 1990). Curiosamente, a proteína ATPase, em gel de poliacrilamida, apresentava peso molecular de 70KDa, inferior ao esperado, correspondente ao peso da proteína caso não existisse a sequência central, similar a endonuclease. Em 1990, Kane e colaboradores demonstraram que a inserção (correspondente a endonuclease) estava presente no RNAm, que todo este RNA era traduzido e que a proteína de peso de 70KDa era resultado de um processamento pós-traducional (Figura 2).

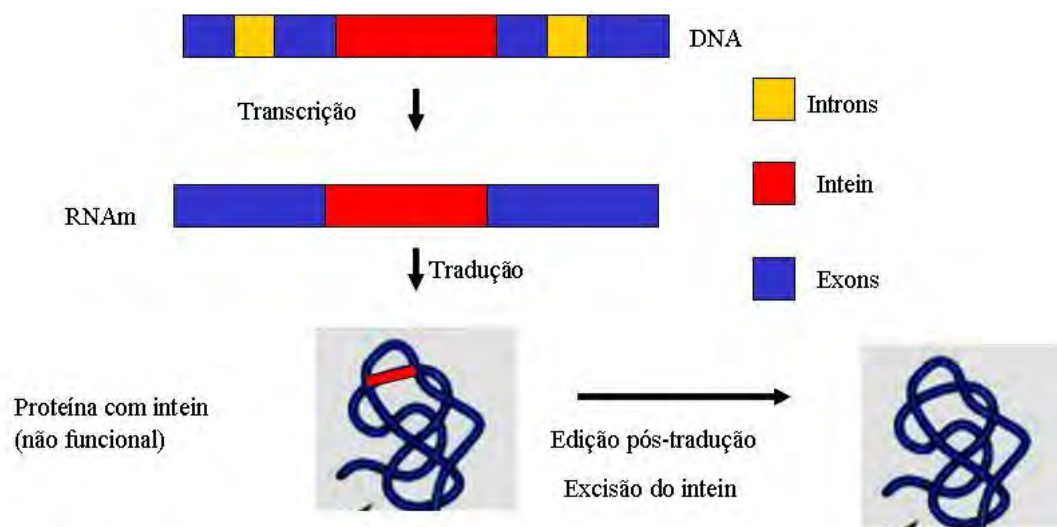


Figura 2: Esquema da excisão do *intein* da proteína hospedeira após o processo da tradução.

Tipos, estrutura e distribuição dos *Inteins*

Os *inteins* ou *Internal Proteins* são sequências codificadoras que são transcritas e traduzidas juntamente com a sequência flanqueadora. Os *inteins*, que podem ter de 134-608 aminoácidos, são excisados e as sequências protéicas flanqueadoras (*exteins*) são unidas por uma ligação peptídica, formando uma proteína funcional (Gogarten et al., 2002).

Existem três tipos de *inteins* na natureza: os *inteins* bi-funcionais (também conhecidos como *large-inteins* ou *full-length inteins*), os *mini-inteins* e os *split inteins* (Figura 3) Os *inteins*

bi-funcionais são constituídos de um domínio responsável pelo *splicing* (auto-corte) protéico (domínio Spl) cujas porções N e C terminais são separadas por um domínio central codificador de uma *homing* endonuclease (domínio HE) que confere mobilidade ao *intein*, resultando na ocupação de alelos vazios e duplicação do elemento genético parasita (Liu 2000). Os *mini-inteins* não possuem o HE tendo, portanto um domínio Spl contínuo. 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. 1998a) (Figura 3).

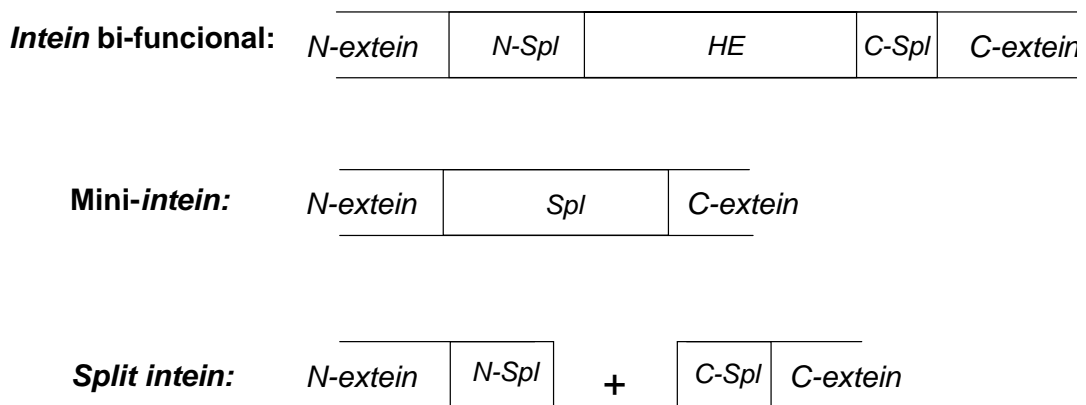


Figura 3: Esquema dos três tipos de *inteins* encontrados na natureza.

Inteins podem ser encontrados nos três grandes domínios de vida: Eucariotos, Bactéria e Archaea, de acordo com a base de dados de *inteins*, o *Inbase* (<http://www.neb.com/neb/inteins.html>). Interessante notar que a razão do tamanho do *intein* em relação ao tamanho da proteína hospedeira pode variar muito. Por exemplo, o tamanho dos três *inteins* (chamados *Mja RCF 1-3*) encontrados no Fator C de Replicação do *Methanococcus jannaschii* são quatro vezes maiores que a proteína hospedeira, enquanto que o *intein Cne*

PRP8 encontrado no Fator de Splicing de RNAm de *Cryptococcus neoformans* é menor do que 1/10 da proteína hospedeira (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 (Liu 2000). Acredita-se que 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 alterarem a função da proteína hospedeira (Swithers et al., 2009).

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). Um *intein* também pode “parasitar” proteínas funcionalmente distintas com alguma similaridade estrutural, como é o caso do *intein* TopA que está presente na topoisomerase I de *Pyrococcus furiosus* e na girase reversa de *Methanococcus jannaschii* (Chute et al. 1998).

Com os avanços gerados pelos 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, *Pbr Prp8* representa o *intein* que ocorre na proteína Prp8 da espécie *Paracoccidoides brasiliensis*. Se mais de um *intein* ocorre em uma mesma proteína, os diferentes *inteins* são designados por números, assim *Tfu Pol-2* denota o segundo *intein* na DNA Polimerase da espécie *Thermococcus fumicolans*. As HEs também possuem uma nomenclatura apropriada, sempre começando com a sigla I (insertion) para HEs 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-*Pbrl* designa o HE presente no *intein* PRP8 de *P. brasiliensis*.

Domínio *Splicing*

A comprovação de que o *splicing* do primeiro *intein* descoberto, o *Sce* VMA, ocorria em nível traducional, e não transcricional, não foi tarefa fácil. A inabilidade em se detectar RNAm cuja sequência do *intein* havia sido cortada não era conclusiva, já que RNAm poderiam ser rapidamente degradados após a tradução. Substituições sinônimas nas junções *splicing* não afetavam o *splicing* do *intein*, indicando que o mesmo não se tratava de um intron, para o qual substituições nucleotídicas nas junções exon-intron são cruciais para a edição do RNAm. Mutações que criavam *stop* codons dentro da sequência do *intein* bloqueavam a síntese da proteína VMA madura e o *splicing* era reativado por uma segunda mutação que restituía a fase aberta de leitura original do *intein*. Embora isso tudo corroborasse a ideia de um *splicing* protéico, e não de RNAm, era possível que estas mutações tivessem sido feitas em sítios cruciais para o funcionamento do intron. Experimentos de *pulse-chase labeling* foram feitos para se estabelecer a cinética da síntese das proteínas precursoras e maduras (sem a sequência espaçadora) expondo células a metionina ou cisteína marcadas com ³⁵S. Além do precursor não ser detectado, a análise foi complicada devido à presença de proteínas adicionais marcadas de tamanho intermediário ao da proteína madura e ao da proteína contendo o *intein*. Estas proteínas eram na verdade produtos resultantes de uma só clivagem (N ou C-terminal). Finalmente Xu e colaboradores (1993) demonstraram que o *splicing* era de fato protéico, portanto pós-traducional, ao clonar o *intein* presente no gene DNA Polimerase de *Pyrococcus* em um *extein* não nativo mostrando a presença de proteína precursora segundo alterações na temperatura e pH (Xu et al., 1993).

O domínio Spl é dividido em blocos de aminoácidos conservados: A ou N1, N2, B ou N3 e N4 para a porção N-Spl (as letras N referem-se à 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, a asparagina da porção C-terminal do *intein* e as duas histidinas conservadas (uma no N-Spl e outra no C-Spl) (Figura 4) (Perler, 2008).



Figura 4: desenho esquemático do domínio *Splicing*, modificado do *Inbase* (Perler, 2000). 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 (Gorgaten et al., 2002) (Figura 5):

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 éster ou tioéster. Conseqüentemente a porção N-*extein* se liga ao oxigênio da serina ou ao enxofre da cisteína do N-Spl.

2) Transesterificação: a clivagem do éster 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 éster. 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.

Uma vez comprovado que o *splicing* era protéico, o próximo passo foi estudar qual seria o elemento mínimo para a ocorrência do mesmo, a fim de se descobrir se o *intein* mediava seu próprio *splicing* ou se dependia das sequências flanqueadoras do *extein*. Para isso, muitos grupos de pesquisadores clonaram *inteins* com cada vez menos nucleotídeos pertencentes ao *extein*, até se descobrir que o *intein* mais o primeiro aminoácido do C-*extein* eram geralmente suficientes para a ocorrência do *splicing* protéico. Contudo, a reação de *splicing* era melhorada quando o *intein* era inserido em um sítio que mimetizava seu sítio de inserção nativo (Davis et al., 1992; Cooper et al., 1993; Xu et al., 1993). Um sistema de avaliação de *splicing* protéico *in vitro* demonstrou que nenhum fator acessório, como fonte de energia ou enzimas, é necessário (Xu et al., 1993).

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, inibiria o funcionamento normal da mesma podendo levar o organismo hospedeiro a morte ou inibindo seu crescimento ou reprodução (Butler et al., 2006).

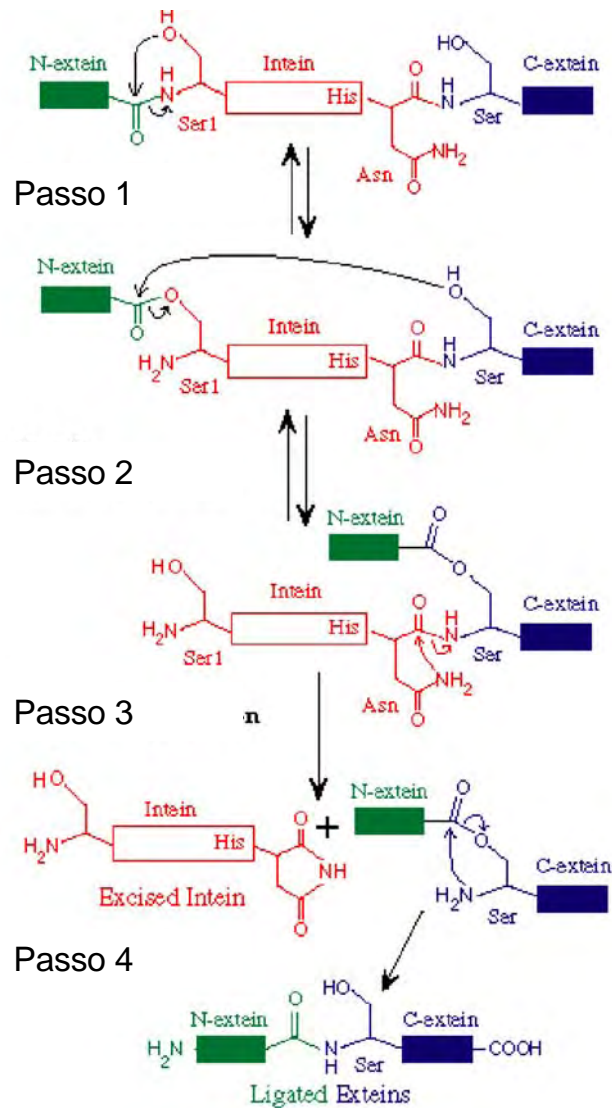


Figura 5: passos envolvidos no *splicing* protéico do *intein*, segundo *Inbase* (Perler, 2000).

Domínio *Homing* Endonuclease e o ciclo *homing*

Homing Endonucleases são enzimas raras que cortam a fita de DNA. Elas estão presentes nos três grandes domínios de vida, em diferentes compartimentos celulares, podendo ocorrer de forma livre no genoma, porém a maioria delas possui sua fase aberta de leitura 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. Assim, a *homing* endonuclease faz destes introns e *inteins*

elementos genéticos móveis. As endonucleases codificadas por introns do grupo II são mais complexas, pois formam partículas de ribonucleoproteínas (RNP) com o intron do RNA. O RNP reconhece o sítio de reconhecimento de um alelo sem o intron II via pareamento entre o seu RNA e a sequência sense de DNA (aquela que é similar ao RNAm). A endonuclease faz um corte *nick* na fita de DNA e transcreve novo DNA usando o seu RNA como molde. Este processo é conhecido como *retrohoming* (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. Todas as endonucleases reconhecem sequências de DNA assimétricas ou pseudo-simétricas variando de 14pb (para a endonuclease I-Dmol da família LAGLIDADG) a 40pb (para a endonuclease I-TevI da família GIY-YIG) (Belfort, 2008).

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 do ATPase vacuolar) de 12 espécies de leveduras do gênero *Saccharomyces*, observaram que somente três endonucleases eram ativas e destas, a PI-Scal (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.

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 4). 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 e Burt, 2005)

A presença de *inteins*, ao que tudo indica, não traz nenhum aumento do valor adaptativo aos organismos que o possuem, mas se trata apenas de um elemento genético autorremovível que, no caso dos *inteins* bi-funcionais, possuem a capacidade de *homing*, responsável por sua rápida dispersão na população. Mas se estes genes são meros elementos parasitas, por que a Seleção Natural não os elimina, já que os manter na população é um gasto energético desnecessário?

A Síntese Moderna da Seleção Natural com a Teoria da Herança Mendeliana, explica como a Seleção Natural interfere nas frequências alélicas ao longo das gerações, porém, a forma segundo a qual esta seleção opera se em grupos, indivíduos ou genes vem sendo tema de intrigantes discussões. A hipótese de que a Seleção Natural opera em um nível molecular foi muito bem defendida por Richard Dawkins (1989) em seu livro “The selfish gene” (“O gene egoísta”), no qual o autor propõe que todos os seres vivos são “máquinas” usadas para a sobrevivência e propagação de genes. Assim, a Seleção Natural obviamente favorecerá aqueles genes que proporcionam fenótipos de alto valor adaptativo para seus hospedeiros. Portanto, um gene egoísta é aquele que aumenta a sua chance de perpetuação em consequência do aumento do valor adaptativo do seu hospedeiro e, no caso dos *inteins*, sua fixação na população não está relacionada a um aumento do *fitness* do organismo que os possui, mas sim ao processo de *homing* e de transferência horizontal. Sendo assim, podemos considerar os *inteins* como genes “mais egoístas” que os demais, pois se proliferam na população sem trazer benefício algum e, por não diminuírem o *fitness* do organismo hospedeiro, já que se retiram da proteína hospedeira pelo processo de *splicing* protéico, “driblam” a Seleção Natural atuante no fenótipo. Por isso, são chamados de elementos genéticos parasitas (Gorgaten et al., 2002).

É notável que os *inteins* ocorram em regiões conservadas de proteínas 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. 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*. Além disso, alguns dos genes ocupados pelos *inteins* são frequentemente transferidos horizontalmente. Por exemplo, genomas virais frequentemente codificam suas próprias enzimas usadas na replicação do DNA. Um *intein* nestes genes possui maior chance de passar de uma espécie para outra, invadindo novas populações (Derbyshire & Belfort, 1998; Perler et al., 1997; Gorgaten et al., 2002). Por fim, sabe-se que grande parte dos *inteins* reside em proteínas responsáveis pela replicação do DNA e, como o *intein* é traduzido simultaneamente com a proteína hospedeira, conclui-se que a endonuclease do domínio HE é produzida durante a replicação do DNA. Isto parece ser vantajoso para a perpetuação do *intein*, já que o processo de *homing* se baseia no reparo da quebra da dupla fita, usando como molde o alelo que já possui o *intein* (Gorgaten et al., 2002).

Após excisada da proteína hospedeira, a HE reconhece 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. Ou seja, o domínio HE é capaz de transferir um elemento genético parasita (no caso o *intein*) para um alelo cognato que não o possui, transformando um organismo heterozigoto para a presença do *intein* em homozigoto (Figura 6), possibilitando

uma herança Super-Mendeliana (Burt & Koufopanou, 2004). No caso do *intein* VMA de *S. cerevisiae*, por exemplo, a partir de um cruzamento $HE^+ \times HE^-$, ao invés de 50% da F1 ser HE^+ , como esperado em uma herança Mendeliana clássica, esta porcentagem chega a 75% de HE^+ (Gimble & Thorner, 1992), sendo, portanto notável a rapidez com a qual os “large-*inteins*” fixam-se nas populações.

Muitos *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 (mais de 130) 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). Já o *intein* TopA ocorre em proteínas funcionalmente distintas, uma é a topoisomerase I e outra é a girase reversa. A semelhança estrutural entre a girase reversa e a topoisomerase poderia explicar esta ocupação em genes distintos (Chute et al., 1998). Quanto à transferência vertical, alelos com e sem 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).

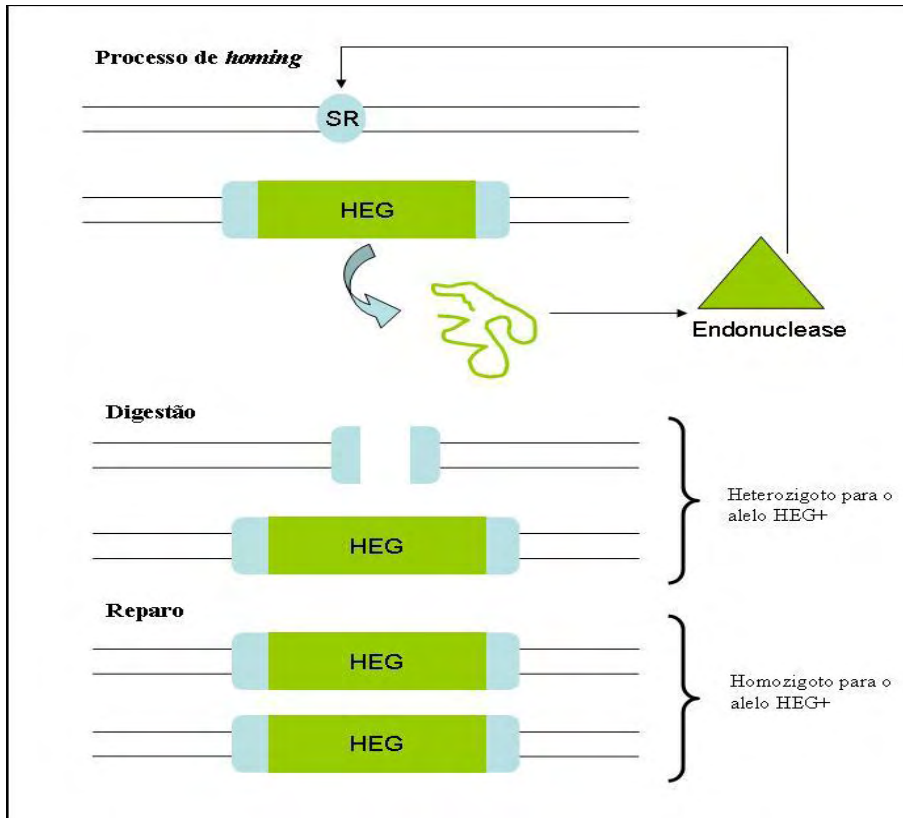


Figura 6: Processo de *homing*, segundo Burt & Koufopanou (2004). SR: sítio de reconhecimento específico da endonuclease, HEG: *homing endonuclease gene*.

A distribuição e dinâmica populacional dos *inteins* parecem seguir um modelo de invasão, fixação, degeneração, perda e re-invasão. Esta dinâmica populacional foi estudada em detalhes para o HE do intron mitocondrial do grupo I presente no gene codificador do RNAr da grande subunidade ribossômica, conhecido como ω ou I-Scel, de *S. cerevisiae*. Quando ocorre reprodução sexuada nesta levedura, ambas as células haplóides contribuem com sua mitocôndria, sendo esta uma herança biparental. Se uma célula haplóide possui o *intein* e a outra não, a frequência do mesmo pode chegar a 99% na próxima geração, devido ao processo de *homing*. Porém, como a reprodução assexuada é mais frequente nesta espécie, a frequência com a qual os alelos I-Scel⁺ e I-Scel⁻ se encontram é muito pequena. Neste estudo da dinâmica evolutiva, Goddard & Burt (1999) analisaram 20 espécies proximamente relacionadas à *S. cerevisiae* e, consistente com o modelo de fixação seguido de degeneração, observaram um número significativo de mutações que inviabilizavam a atividade do HE,

restando apenas a atividade *splicing* do intron. Sendo assim, como o I-Scel poderia ter permanecido ao longo do tempo evolutivo?

Uma possível resposta para esta questão é o processo de invasão em novas espécies por transferência horizontal, aumento da frequência do *intein* e/ou intron do grupo I, seguido de fixação e novo declínio (Figura 7). A ideia de transferência horizontal para explicar novas invasões vem do fato de que a filogenia das espécies hospedeiras não coincide com a das HEs I-Scel de *S. cerevisiae* (Goddard & Burt, 1999).

De fato, a frequência desses domínios HE, seja de introns ou de *inteins*, parece aumentar e diminuir de forma cíclica nas populações de reprodução sexuada. Neste tipo de reprodução, é mais fácil o encontro de alelos com e sem o *intein*. A partir do momento que o *intein* ou intron do grupo I 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 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 (Burt & Koufopanou, 2004).

Entretanto, uma recente análise de *inteins* bi-funcionais sugere 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, ou 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, como esquematizado na figura 7, sendo positivamente selecionados devido a um eventual benefício trazido pela HE.

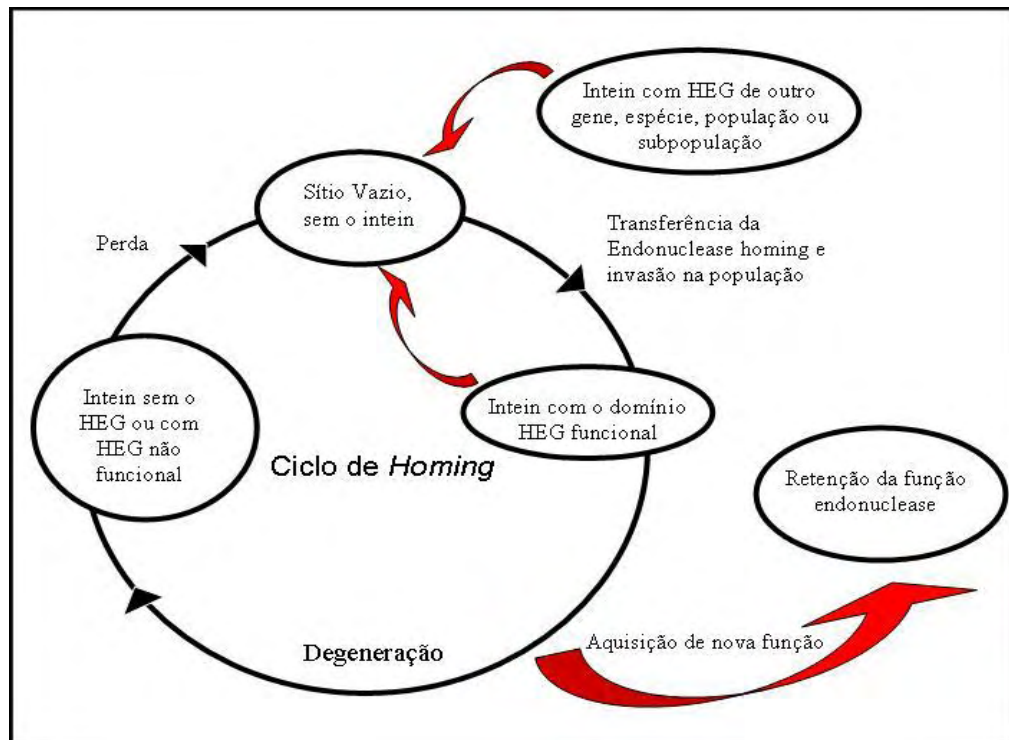


Figura 7: Esquema de invasão, fixação, degeneração e re-invasão do *intein* na população segundo Burt & Koufopanou (2004).

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. Esta endonuclease possui grande similaridade com o *intein* VMA1 de *S. cerevisiae*, um dos *inteins* mais estudados desta espécie, incluindo regiões conservadas relacionadas à atividade de *splicing*, embora a HO não possua de fato a capacidade de se autorremover. 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* (Perler, 1997; Pietrokovski, 2001)

A endonuclease HO inicia 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* ($Y\alpha$ e $Y\alpha$) 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 8) (Haber & Wolfe., 2005)

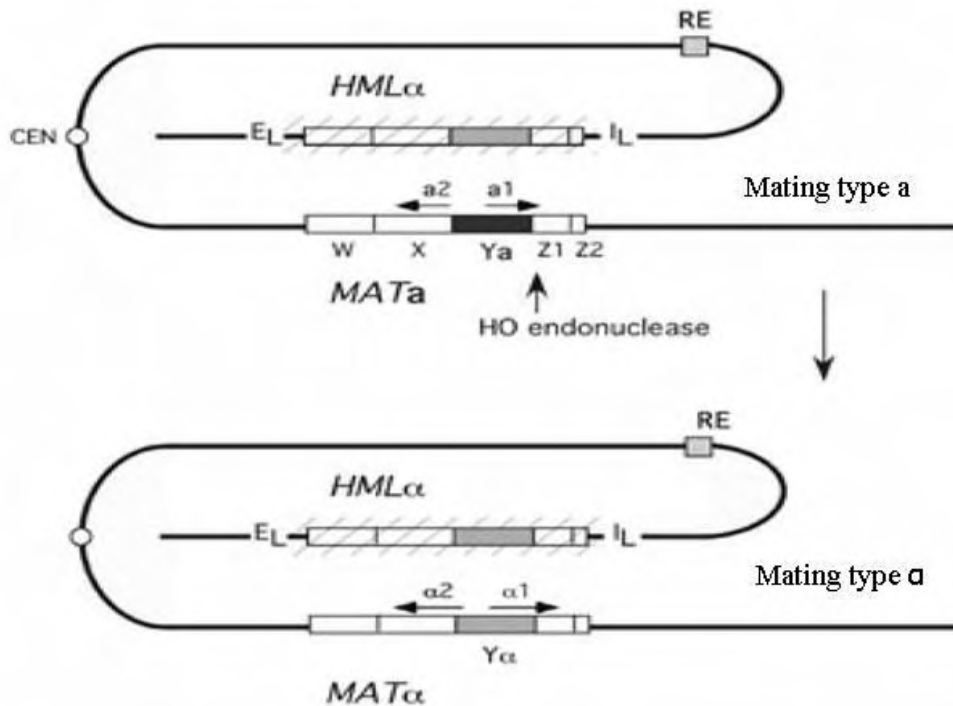


Figura 8: troca do mating type em *S. cerevisiae*, catalisada pela endonuclease HO (Haber & Wolfe, 2005).

Origem dos *inteins*

Sabe-se que o domínio Spl dos *inteins* é similar aos domínios de auto-processamento de proteínas regulatórias que sofrem excisão catalítica, porém não se sabe o que veio primeiro, se o elemento genético parasita ou o domínio de *splicing* autocatalítico (Petrokovski, 2001).

Existe uma hipótese de que a *Homing* Endonuclease tenha invadido introns do grupo I e *inteins* depois que eles se originaram. Esta hipótese parece ser consistente uma vez que diferentes tipos de endonucleases ocupam diferentes *inteins* e o tipo mais comum, que é a LAGLIDADG, também é encontrada em introns do grupo I e também como *open reading frames* sem conexão alguma com elementos parasitas. Ora, se endonucleases são móveis nos genomas, elementos genéticos que se autorremovem do produto gênico constituiriam sítios ideais de integração, pois, nestes locais, as funções codificadas pelo DNA vizinho não seriam interrompidas (Derbyshire & Belfort, 1998).

Embora o domínio Spl do *intein* VMA de *S. cerevisiae* parece colaborar com o HE no reconhecimento do sítio *homing* do DNA, experimentos de mutagênese sítio-dirigida, em muitos *inteins*, comprovam que os domínios Spl e HE são autônomos, ou seja, uma vez retirado o domínio HE de um *intein* bi-funcional, a atividade *splicing* é mantida (Hodges et al., 1995, Gimble & Stephens, 1995, Derbyshire et al., 1997). Isso sustenta a ideia de que os dois domínios, Spl e HE, tiveram origens independentes e que endonucleases provavelmente tenham invadido sítios capazes de realizar auto-*splicing*.

Observa-se que o domínio HE tem tendência a degradar. Muitos *inteins* já foram descritos sem este domínio (“mini-*inteins*”) ou com um HE não funcional. É razoável, portanto inferir que a maioria dos “mini-*inteins*” tenham evoluído a partir de *inteins* com HE funcional. Isto é reforçado pelo fato de que os “mini-*inteins*” não formam um grupo coerente em reconstruções filogenéticas, além disso, sua ampla distribuição entre os mais diferentes organismos indica que um dia eles foram *inteins* móveis (com HE). Considerar os “mini-*inteins*” como derivados de *inteins* ancestrais sem o HE implicaria em várias origens ao longo do tempo evolutivo, nos mais diversos organismos, e isso não é nada parcimonioso. A maior

degeneração do domínio HE quando comparado ao Spl é compreensível uma vez que um domínio Spl degenerado e não funcional acarretaria em morte da célula (como já mencionado, a maioria dos *inteins* ocupam genes essenciais para a sobrevivência celular) diminuindo drasticamente seu valor adaptativo, e, portanto tendo contribuição para a próxima geração cada vez menor. Já o domínio HE, uma vez dentro da proteína hospedeira, não sofre mais pressão seletiva para a manutenção desta atividade (Gorgaten et al., 2002).

Inteins em fungos

Dentre todos os *inteins* descritos, cerca de 60 ocorrem no domínio Eucarioto, dentre estes, podemos encontrar algumas algas e amebas, mas a maioria dos eucariotos nos quais *inteins* foram encontrados até hoje, pertence ao reino Fungi (cerca de 50 espécies), sendo provenientes dos diferentes filos (Chitridiomycota, Zigomicota, Basidiomicota e Ascomicota) (Tabela 1, segundo *Inbase*). A maioria dos *inteins* fúngicos estão presentes nos genes ATPase vacuolar, subunidade A (*intein VMA*) e *PRP8* (*intein PRP8*). O *intein VMA* é amplamente disperso entre leveduras, como *Saccharomyces*, *Zygosacchromyces* e algumas espécies de *Candida* (todos Ascomicotas), enquanto o *intein PRP8* está presente, de forma esporádica em todos os filos do reino Fungi (Perler 2000), sendo mais abundantes entre os Ascomicetos, em algumas espécies das ordens Eurotiales, Onigenales e Helotiales e em poucos basidiomicetos, da ordem Tremellales e no chitridiomiceto *Batrachochytrium dendrobatidis*, atualmente conhecido por dizimar populações de anuros (*Inbase*). Algumas espécies patogênicas importantes, como *Histoplasma capsulatum*, *Paracoccidioides brasiliensis* (Ascomicota, Ordem Onygenales), *Cryptococcus neoformans* e *Cryptococcus gattii* (Basidiomicota, Ordem Tremellales) possuem o *intein PRP8* (Butler et al. 2006).

Tabela 1: *inteins* no reino Fungi

Filo	Espécie	Gene hospedeiro (<i>Extein</i>)	Tipo de <i>intein</i> **
Chitridiomycota	<i>Batrachomyces dendrobatidis</i>	RNA polymerase II, second largest subunit	bi-funcional
	<i>Coelomomyces stegomyiae</i>	PRP8, pre-mRNA <i>splicing</i> factor Eukaryotic translation initiation factor 5B	bi-funcional bi-funcional
	<i>Spizellomyces punctatus</i>	RNA polymerase II, second largest subunit PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional bi-funcional
Zigomicota	<i>Phycomyces blakesleeenanus</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
	<i>Spiromyces spiralis</i>	RNA polymerase II, second largest subunit	bi-funcional
Ascomycota	<i>Histoplasma capsulatum</i> [†]	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Aspergillus fumigatus</i> var. <i>ellipticus</i>	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Aspergillus fumigatus</i>	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Aspergillus nidulans</i>	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Aspergillus giganteus</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
	<i>Aspergillus viridinutans</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
	<i>Aspergillus brevipes</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
	<i>Botrytis cinerea</i>	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Candida glabrata</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Candida parapsilosis</i>	Threonyl-tRNA synthetase	Mini- <i>intein</i>
	<i>Candida tropicalis</i>	Threonyl-tRNA synthetase	Mini- <i>intein</i>
		Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Candida famata</i>	Glutamate synthase	bi-funcional
		Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Kazachstania exígua</i> , antigo	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Saccharomyces exiguus</i> ,	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Kluyveromyces lactis</i> ,	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Lodderomyces elongisporus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Neosartorya aurata</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
	<i>Neosartorya fennelliae</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
	<i>Neosartorya fischeri</i>	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Neosartorya glabra</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
	<i>Neosartorya quadricincta</i> ,	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>
<i>Neosartorya spinosa</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>	
<i>Podospora anserina</i>	Chitin synthase class II	bi-funcional	
	Glutamate synthase	bi-funcional	
<i>Paracoccidioides brasiliensis</i>	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional	
<i>Penicillium chrysogenum</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>	
<i>Penicillium expansum</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>	
<i>Pichia guilliermondii</i> (<i>Candida</i>)	Glutamate synthase	bi-funcional	
<i>Phaeosphaeria nodorum</i>	Glutamate synthase	bi-funcional	
	RNA polymerase I	Mini- <i>intein</i>	
<i>Pichia stipitis</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional	
<i>Pyrenophora tritici-repentis</i> Pt-1C-BF	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional	
<i>Penicillium vulpinum</i> (antigo <i>P. claviforme</i>)	PRP8, pre-mRNA <i>splicing</i> factor	Mini- <i>intein</i>	

Tabela 1: continuação

Filo	Espécie	Gene hospedeiro (Extein)	Tipo de intein**
Ascomycota	<i>Pyrenophora tritici-repentis</i> Pt-1C-BF	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Penicillium vulpinum</i> (antigo <i>P. claviforme</i>)	PRP8, pre-mRNA <i>splicing</i> factor	Mini-intein
	<i>Blastomyces dermatitidis</i> *	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Emmonsia parva</i> *	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Saccharomyces castellii</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Saccharomyces cariocanus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Saccharomyces cerevisiae</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Saccharomyces dairenensis</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Saccharomyces exiguus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Schizosaccharomyces japonicus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Saccharomyces pastorianus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Saccharomyces unisporus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Torulaspora globosa</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Torulaspora pretoriensis</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Vanderwaltozyma polyspora</i> , formerly <i>Kluyveromyces polysporus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
	<i>Zygosaccharomyces bailii</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional
<i>Zygosaccharomyces bisporus</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional	
<i>Zygosaccharomyces rouxii</i>	Vacuolar ATPase (H ⁺ -transporting ATP synthase), subunit A	bi-funcional	
<i>Uncinocarpus reesii</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini-intein	
Basidiomycota	<i>Cryptococcus gattii</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini-intein
	<i>Cryptococcus laurentii</i>	PRP8, pre-mRNA <i>splicing</i> factor	bi-funcional
	<i>Cryptococcus neoformans</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini-intein
	<i>Cryptococcus bacillisporus</i>	PRP8, pre-mRNA <i>splicing</i> factor	Mini-intein

*Descritos nesta tese (artigo 3, Theodoro et al., 2010, *in oress*).

** Mini-inteins só possuem os blocos A, B, F e G e os bi-funcionais são aqueles que possuem o domínio HE íntegro ou apenas parte dele.

Intein PRP8

Este *intein* ocorre no gene *PRP8*, o qual codifica a proteína Prp8, uma das maiores e mais conservadas proteínas nucleares, apresentando cerca de 60% de similaridade entre leveduras e humanos. Esta proteína ocupa uma posição central na porção catalítica do spliceossomo, sendo um dos componentes da U5 do snRNP (*small nuclear Ribonucleoprotein*) envolvido em vários rearranjos moleculares cruciais no *splicing* do RNA pré-mensageiro, passo essencial para a expressão gênica (Grainger & Beggs 2005).

O *intein PRP8* foi inicialmente descrito em isolados de *C. neoformans* (variedades *grubii* e *neoformans*) e de *C. gattii* como mini-*inteins* (sem o domínio HE) (Butler et al. 2001 e Butler & Poulter 2005). Além destas espécies de *Cryptococcus*, o *intein PRP8* também foi encontrado, como um *intein* bi-funcional, em *Cryptococcus laurentii*. Uma vez que esta espécie é considerada moderadamente distante de *C. neoformans*, Butler & Poulter (2005) sugeriram que o *intein* encontrado em *C. neoformans* e *C. gattii* receberam este *intein* por transferência horizontal, na qual o *C. laurentii* ou outra espécie tenha sido a fonte doadora.

Butler et al. (2006) procuraram *inteins PRP8* na base de dados de projetos genomas e descobriram que ele está presente em muitas espécies de Ascomicetos. Os autores também compararam o *intein PRP8* com o *intein VMA*, em relação aos domínios Spl e HE. Curiosamente o *intein PRP8* pode ser encontrado tanto na sua forma de mini-*intein* como também na sua forma bi-funcional, dependendo da espécie fúngica (Tabela 1), enquanto todos os *inteins VMA* são bi-funcionais, isto é, apresentam aminoácidos referentes ao domínio HE ainda que, em algumas espécies, este domínio se apresente degenerado, ou seja, com deleções e/ou substituições e, portanto, provavelmente inativo.

Razões dS/dN revelaram que sequências HE do *intein PRP8* são conservadas por seleção sugerindo que os *inteins* estão ativos ou estiveram ativos durante a maior parte do tempo de divergência. Comparações de *inteins PRP8* de algumas espécies indicam que as sequências estão saturadas de substituições sinônimas, sugerindo que eles vêm divergindo por

um período maior do que os *inteins* VMA, cujas *homing* endonucleases são extremamente divergentes, fato que vem sendo atribuído à perda de função das mesmas (Butler et al., 2006).

Visto que o *intein* PRP8 tem uma origem antiga e apresenta transferência vertical dentro dos Euscomycetes (a filogenia do *intein* e do gene *PRP8* são as mesmas), atribui-se sua esporádica ocorrência às sucessivas perdas ao longo do tempo evolutivo. Este é o caso de *Coccidioides immitis* e de *Coccidioides posadasii*, duas espécies patogênicas sem *intein* da ordem Onygenales, filogeneticamente próximas de *Uncinocarpus reesii*, o qual apresenta um mini-*intein* PRP8. Por outro lado, a presença do *intein* PRP8 em *Cryptococcus*, um gênero de Basidiomicetos, sugere a ocorrência de transferência horizontal, uma vez que a sua alta similaridade com *inteins* de Euscomycetes não reflete o tempo de divergência entre os organismos hospedeiros (Butler et al., 2006).

No artigo 3 deste capítulo, as sequências de aminoácidos do *intein* PRP8 de *P. brasiliensis*, *H. capsulatum*, *E. parva* e *B. dermatitidis* foram avaliadas tanto em relação à razão dS/dN, para ambos os domínios, Spl e HE, bem como em relação à presença dos dois resíduos de aspartato ditos essenciais para o funcionamento da endonuclease. Como já esperado, o domínio Spl se mostrou mais conservado em relação ao HE e, corroborando dados da literatura, observamos que o valor dS/dN para o HE destes *inteins* foi maior que 1. Isto poderia indicar que o domínio HE está ativo nestes patógenos, contudo, apenas as espécies *H. capsulatum* e *B. dermatitidis* apresentam os dois aspartatos essenciais. As espécies S1, PS2 e PS3 do complexo *P. brasiliensis* e a espécie *P. lutzii* apresentam uma glicina e uma serina, respectivamente, no lugar do segundo aspartato, enquanto *E. parva* apresenta substituições em ambos aspartatos.

Ainda no artigo 3, experimentos *in vitro* foram conduzidos a fim de se averiguar a atividade do domínio HE (PI-*Pbr1*) do *intein Pbr* PRP8. Uma reação de PCR inverso foi realizada para restaurar o segundo aspartato, originalmente substituído por uma glicina, no bloco E do *intein* PRP8 de um isolado de *P. brasiliensis* pertencente à espécie PS2. Assim, tanto a endonuclease do tipo original (PI-*Pbr1*/WT), como a mutante (PI-*Pbr1*/G451D) foram

clonadas, produzidas em *E. coli*, purificadas e testadas contra um substrato contendo dois sítios de reconhecimento, o plasmídeo pDHES2.

Para construir o pDHES2 modificamos o gene *PRP8* de *S. cerevisiae*, o qual não possui *intein*, substituindo os 52pb que estariam flanqueando o *intein* pelos 52pb que de fato flanqueiam o *intein* PRP8 em *P. brasiliensis*. Esta região, correspondente ao sítio de reconhecimento da PI-*Pbr1*, inserida em 455pb do gene *PRP8* de *S. cerevisiae*, foi clonada em um plasmídeo pJET, resultando no pJHES1, do qual esta região foi digerida e subclonada em um plasmídeo pDrive, resultando no pDHES1. O substrato, pDHES2, foi obtido pela ligação do fragmento de 814pb digeridos a partir do pDHES1 com *HindIII/NotI* (contendo o sítio de reconhecimento) com um plasmídeo pDHES1 digerido com as mesmas enzimas.

O fragmento esperado, cerca de 800pb, resultante da digestão do pDHES2 não foi observado nem com a proteína selvagem PI-*Pbr1*/WT, nem com a mutante PI-*Pbr1*/G451D, reforçando a ideia de que provavelmente o domínio HE de *P. brasiliensis* se encontra em fase de degeneração e, portanto, inativo.

Logicamente, as condições dos experimentos *in vitro* não recriam completamente o ambiente intracelular. Porém o estudo *in vivo* da atividade de *homing* endonuclease, ou seja, a cópia do *intein* para um alelo vazio seria um tanto trabalhosa no caso de *P. brasiliensis*, por dois motivos: i) aparentemente não existem indivíduos sem o *intein* nas populações deste complexo de espécies e ii) apesar da reprodução sexuada ser evidente, pelo menos em análises filogenéticas e das sequências do locus MAT, ainda não foi possível obter ascocarpos e ascósporos em cruzamentos entre *mating types* diferentes para *P. brasiliensis* (Torres et al., 2009; Teixeira et al, em preparação).

O único *intein* PRP8 que teve a atividade de seu HE comprovada experimentalmente, *in vivo*, foi o do fungo fitopatógeno *Botrytis cinerea*, da ordem Helotiales, família Sclerotiniaceae, filo Ascomycota. O processo de *homing* foi descrito durante a meiose de um heterozigoto *Intein⁺/Intein⁻*. Isto foi possível, pois, no caso desta espécie, além da fase sexuada ser facilmente obtida em laboratório, existem alelos com e sem o *intein* na população. Os autores deste trabalho, Bokor et al (2010), também evidenciaram que o processo de *homing* do *intein*

Bci PRP8 desencadeia um fenômeno de conversão gênica, pois também copia parte do *extein* que flanqueia o *intein*, substituindo as regiões cognatas do alelo vazio que recebe o *intein*.

Ao compararmos a sequência de aminoácidos do *intein* de *B. cinera* com o de *P. brasiliensis*, observamos que o primeiro possui os dois aspartatos essenciais e alguns resíduos a mais no domínio HE. Interessante notar que estes *inteins* PRP8 parecem estar em diferentes fases do ciclo *homing*. Enquanto o *Pbr* PRP8 e também o *Epa* PRP8 parecem estar em um início de um processo de degeneração devido às substituições dos aspartatos essenciais (início, pois o fato do dS/dN do HE ser maior que 1 pode indicar que até recentemente este domínio estava ativo), o *Bci* PRP8 parece estar em fase de invasão e expansão, na qual a presença de alelos sem o *intein* na população é considerada um fator importante para a manutenção de endonucleases ativas no domínio HE.

Como observado na Tabela 1, além das espécies de *Cryptococcus*, pertencentes ao filo Basidiomicota, *inteins* no gene *PRP8* também já foram descritos em muitos Ascomicetos, como *A. nidulans*, *A. fumigatus*, *P. brasiliensis*, *H. capsulatum*, *B. dermatitidis*, *E. parva* e *B. cinera*. Neste grupo, porém, não há evidência de transferência horizontal, uma vez que análises filogenéticas destes *inteins* são concordantes com aquelas realizadas com outras sequências nucleares como, por exemplo, genes codificadores de RNAr, sugerindo assim uma transmissão vertical dentro do grupo dos Euascomicetos (Butler et al. 2006). Já que poucos basidiomicetos (somente as espécies de *Cryptococcus*, da ordem Tremalles) apresentam o *intein* PRP8 *intein*, parece ser plausível atribuir esta ocorrência a um evento de transmissão horizontal entre ascomicetos como doadores e basidiomicetos como receptores. Segundo Poulter et al (2007), esta transferência poderia ter ocorrido durante a fase parasítica destes fungos, em casos de co-infecção ou durante a fase saprofítica destes patógenos, quando diferentes células fúngicas são fagocitadas por microrganismos do solo. De fato, a interação entre fungos patogênicos e amebas fagocíticas do solo tem sido tema de muitos estudos que demonstram similaridades entre as interações fungo-macrófago e fungo-ameba e sugerem que os tratos de virulência tenham na verdade se originado em um ambiente saprobiótico, muito antes do aparecimento do hospedeiro vertebrado. Hoje, estes fatores de virulência teriam um duplo uso, considerando

ambos os ambientes, solo e organismo hospedeiro (Casadevall et al. 2003; Steenbergen et al. 2003, 2004).

Uso de *inteins* em análises filogenéticas

O conceito de espécie, como já discutido com maiores detalhes no capítulo 1 desta tese, é um dos temas centrais da disciplina de evolução e, embora espécie seja considerada um agrupamento concreto e não artificial como gênero, ordem, classe, família, filo e reino, sua definição e seu reconhecimento são considerados grandes desafios, mesmo atualmente, na “era” da genômica e da proteômica.

Como também já discutido no capítulo 1, um dos objetivos principais desta tese foi o de pesquisar marcadores morfológicos e moleculares capazes de diferenciar as espécies crípticas de *P. brasiliensis*. Dentre os candidatos a marcadores moleculares estão os SNPs dos genes *gp43* e *arf*, e a sequência de nucleotídeos do *intein PRP8*.

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 ativo (Gorgaten & Hilário, 2006). No caso dos *inteins PRP8*, ainda que os domínios HE pareçam ser conservados pela seleção (seu dS/dN é um pouco maior que 1), indicando possível atividade, eles são muito diversos, provavelmente devido a uma origem antiga e uma longa história evolutiva (Butler et al. 2006).

Estudos filogenéticos usando o *intein PRP8* são facilitados não somente devido a sua transmissão vertical dentro do grupo dos Euscomycetos, mas também devido à localização deste elemento parasita em um gene altamente conservado nos eucariotos, o *PRP8*, possibilitando assim o desenho de *primers* para sua amplificação por PCR seguido de seu sequenciamento. Desta forma, Butler e 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* pode se tratar

de um grupo em especiação. Estas leveduras encapsuladas são capazes de causar sérias infecções tanto em pessoas imunocomprometidas como também em imunocompetentes. 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).

Aplicando uma abordagem semelhante, realizamos a análise filogenética do *intein* PRP8 em diferentes isolados de *P. brasiliensis*, pertencentes às quatro diferentes espécies: S1, PS2, PS3 e *P. lutzii* (artigo número 1 deste capítulo, Theodoro et al., 2008). Neste trabalho, o *intein* PRP8 foi sequenciado em 22 isolados e as análises filogenéticas realizadas por Máxima Parcimônia, Máxima Verossimilhança e análise Bayesiana (para detalhes sobre estes métodos, veja capítulo 1 desta tese) separaram claramente, com altos valores de *bootstrap*, as quatro espécies revelando uma diferença significativa entre o *P. lutzii* (na época do artigo, conhecido como Pb01-like) e as demais espécies, corroborando com as análises filogenéticas realizadas por Carrero et al (2008) e Teixeira et al (2009).

Todos os isolados de *P. brasiliensis* avaliados apresentaram um *intein* bi-funcional no gene *PRP8*. O domínio HE destes *inteins* está aparentemente inativo, pois o segundo aspartato considerado essencial para a atividade endonuclease foi substituído por uma glicina nas espécies S1, PS2 e PS3 e por uma serina no *P. lutzii*. Por outro lado, a razão dS/dN para o domínio HE foi maior que 1, indicando que esta região estaria sob seleção conservadora e provavelmente se encontrava ativa até recentemente. Embora o número de sítios polimórficas no *intein* PRP8 de *P. brasiliensis* não foi alto como geralmente se observa em outras regiões nucleares codificadoras, este elemento parasita contém informação genética suficiente para separar as espécies de *P. brasiliensis*, sendo, portanto um marcador genético adicional para este patógeno.

Como foi ressaltado, como nota, na Tabela 1, nesta tese também foram descritos mais dois *inteins* PRP8, um em *B. dermatitidis* e outro em *E. parva* (artigo número 3 deste capítulo).

Juntamente com *H. capsulatum* e *P. brasiliensis*, estas duas espécies fazem parte da família Ajellomycetaceae, ordem Onygenales (Untereiner et al., 2004) e são agentes etiológicos de importantes micoses sistêmicas. *B. dermatitidis* é restrito à América do Norte, enquanto *H. capsulatum* e *E. parva* apresentam uma distribuição geográfica um pouco mais ampla, sendo que o primeiro ocorre em quase todos os continentes e o segundo no sudoeste dos Estados Unidos, Austrália e leste Europeu (Hoog et al., 2000).

Segundo análises filogenéticas da sequência codificadora da grande subunidade ribossomal (LSU) e da região ITS1, Untereiner et al (2004) propuseram a família Ajellomycetaceae, a qual também inclui o fungo patogênico não cultivável *Lacazia loboi* (Vilela et al., 2009). Este fungo é agente etiológico da lacaziose, micose subcutânea comumente observada em golfinhos e humanos, principalmente provenientes da bacia Amazônica, Brasil (Taborda et al., 1999). Na tentativa de se detectar o *intein PRP8* nesta espécie, em amostras de DNA extraídas de camundongos artificialmente infectados (amostras cedidas pelo Dr Leonel Mendoza, Universidade de Michigan, USA) diferentes combinações de primers degenerados foram desenhadas (com base em sequências do gene *PRP8* de Ascomicetos depositadas no GenBank). Porém, os produtos de PCR obtidos foram inespecíficos, ora pertencendo a outras regiões genômicas, ora provenientes do DNA do hospedeiro (camundongo). Espera-se que isto seja em breve resolvido, principalmente através dos dados gerados pelo projeto genoma deste patógeno, possibilitando também um melhor entendimento das características peculiares deste patógeno quando comparado aos demais fungos da família Ajellomycetaceae.

Além destas espécies, a família Ajellomycetaceae também inclui as espécies *Emmonsia crescens*, *Emmonsia. ssp* e *Ajellomyces grisea* (a única não patogênica deste grupo). Em geral os membros desta família se encontram associados a vertebrados e compartilham substratos e características fisiológicas similares, como crescimento em fezes e/ou solo e capacidade de crescimento a 36°C.

O estudo documentado no artigo 3 deste capítulo reforça a aplicação do *intein PRP8* para estudos filogenéticos no nível de espécie, uma vez que a análise realizada por Neighbour Joining para *B. dermatitidis*, *H. capsulatum*, *E. parva* e *P. brasiliensis*, incluindo as quatro

espécies crípticas, corrobora os dados previamente obtidos por Untereiner et al (2004) com genes codificadores de RNAr.

Contudo, vale aqui ressaltar que o uso do *intein* PRP8 para análises filogenéticas acima do nível taxonômico de espécie (gênero, família, ordem...) não seria adequado, dado a distribuição esporádica deste *intein* (Tabela 1).

***Intein* PRP8 como alvo terapêutico**

O *intein* PRP8 das espécies patogênicas da família Ajellomycetaceae é, obviamente, um potencial alvo terapêutico, já que se encontra em um dos genes mais conservados dentre os eucariotos e essencial para a sobrevivência da célula hospedeira. Uma vez inibido o *splicing* do *intein*, a proteína Prp8 resultante não seria mais funcional, interrompendo assim a sobrevivência e proliferação da célula fúngica. Como o sítio de inserção do *intein* PRP8 é conservado entre estes patógenos, a descoberta de uma droga que iniba o *splicing* protéico poderia ter um amplo campo de ação (Liu & Yang, 2004). Sendo assim, o estudo do funcionamento do domínio Spl deste *intein* é de extrema importância e este foi um dos objetivos principais do artigo 3 deste capítulo.

Liu & Yang (2004) já haviam anteriormente avaliado a atividade *splicing* do *intein* PRP8 em uma proteína hospedeira não-nativa, em *E. coli*, para as espécies *Aspergillus fumigatus*, *A. nidulans*, *H. capsulatum* e *C. neoformans*. O *intein* foi inserido no plasmídeo de expressão pMST entre os sítios *XhoI* e *AgeI*, 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. O *splicing* do *intein* foi avaliado a 25 e a 37°C. Em *H. capsulatum* a eficiência do *splicing* a 37°C foi somente 25%, enquanto que a 25°C foi de 100%.

O trabalho referente ao artigo 3 foi realizado em cooperação com o Prof. Dr. Xiang-Qin Liu na universidade Dalhousie, Canadá. Neste trabalho, 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. lutzii*) foi avaliada no mesmo modelo de *extein* não nativo, sendo inserido entre uma MBP e uma tireodoxina, em *E. coli*.

Foi possível observar o *splicing* protéico em todos os isolados fúngicos avaliados neste artigo tanto a 25 como a 37°C. Dentre eles, o *Epa PRP8* apresentou a menor eficiência de *splicing* a 37°C. Essa baixa eficiência, como mencionado acima, já havia sido documentado em *H. capsulatum* por Liu e Yang (2004). Vale ressaltar que como fungos dimórficos todas as espécies aqui estudadas precisam converter o micélio (M) para levedura (L), seja durante a infecção ou sob altas temperaturas (36-37°C). Se a eficiência do *splicing* do *intein PRP8* a 37°C for tão baixa como se mostrou em uma proteína não nativa, a transição M-L poderia ser prejudicada já que muitos genes devem ser transcritos e corretamente traduzidos para que o fungo assuma a fisiologia da fase L. Neste caso, o *intein PRP8* poderia diminuir o *fitness* do patógeno (Theodoro & Bagagli 2009, artigo 2 deste capítulo). Levando em consideração que o *intein PRP8* de *H. capsulatum* é um *intein* bi-funcional, provavelmente com um HE ativo (Poulter et al., 2007), é intrigante o fato de seu HE estar aparentemente mantido por uma seleção conservadora, lembrando que o *Hca PRP8* possui ambos aspartatos essenciais além de um dS/dN maior que 1. Neste caso, o balanço entre a queda do *fitness* do hospedeiro e o contínuo processo de *homing* do elemento genético parasita poderiam levar a uma longa permanência do HE, especialmente quando a desvantagem seletiva do hospedeiro e a mobilidade do *intein* não são as mesmas em diferentes subpopulações (Gogarten and Hilario 2006). Esta hipótese parece ser razoável para a manutenção do *intein PRP8* em *H. capsulatum*, considerando sua distribuição quase que mundial devido a sua propagação por hospedeiros voadores como morcegos e aves, resultando assim em muitas populações, incluindo espécies crípticas (Kasuga et al., 2003).

O mesmo não parece ser válido para *E. parva*, apesar de sua ampla distribuição quando comparado a *P. brasiliensis* e *B. dermatitidis*. Apesar do dS/dN maior que 1, o HE do *Epa PRP8* parece estar degenerado devido a substituições em ambos aspartatos essenciais. Além disso, a baixa eficiência de *splicing* deste *intein* a 37°C poderia contribuir para

o aspecto não-dispersivo e assintomático da adiaspiromicose (micose sistêmica causada pelo *E. parva*), uma vez que os adiasporos ficam nos pulmões e não se disseminam para outros órgãos internos. Curiosamente, é também interessante notar que tanto *H. capsulatum* como *E. parva* necessitam de meios de cultura mais enriquecidos, como BHI (*Brain Heart Infusion*) para transição M-L, quando comparados com *P. brasiliensis* e *B. dermatitidis*, para os quais a transição M-L é facilmente obtida em meios mais simples como BDA (Batata Dextrose Agar) ou GPYA (Dextrose, Peptona, Extrato de Levedura e Agar). A baixa eficiência do *splicing* do *Intein PRP8* a 37°C observada para *E. parva* e *H. capsulatum* poderia estar de alguma forma relacionada à lenta transição M-L e a maior necessidade de nutrientes? Para responder esta questão, mais isolados destas espécies devem ser avaliados, tanto no que diz respeito a suas características micológicas básicas bem como ao funcionamento do domínio Spl de seus *inteins*.

A habilidade dos *inteins* de se autorremoverem de *exteins* não-nativos vem, na verdade, sendo estudada desde que se comprovou que o *splicing* era de fato protéico e provou ser de fato um sistema robusto e ao mesmo tempo flexível, sugerindo muitas aplicações biotecnológicas como será abordado mais a frente.

Os modelos de *exteins* não nativos facilitaram estudos de mutagênese do domínio Spl. Assim, Pearl et al (2007) determinaram os resíduos de aminoácidos essenciais e a sequência mínima necessária para a clivagem completa do *intein PRP8* de *C. neoformans*. O mesmo foi realizado para o mini-*intein PRP8* de *Penicillium chrysogenum*, o menor *intein* conhecido até hoje. Após a remoção de 14 aminoácidos, o *intein* contendo apenas 143 aa se mostrou funcional quando expresso em *E. coli* (Elleuche et al. 2008). Este mesmo *intein* foi testado em um ensaio de *trans-splicing* ao ser partido artificialmente. O *trans-splicing* protéico envolve a junção de duas metades de um *intein*, devido a sua alta afinidade, seguida da ligação das duas sequências de *extein*. Portanto este *intein*, mesmo com suas partes N-Spl e C-Spl separadas, é capaz de se re-associar e realizar atividade de *splicing* em *trans* (Tavassoli et al. 2008).

A metodologia aplicada no artigo 3 deste capítulo é ideal para se avaliar, inicialmente, o potencial *splicing* de um *intein* em um contexto não nativo, porém seu uso em sistemas de

seleção de drogas que inibam o *splicing* seria um tanto laborioso. 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 *LACZα* 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 o crescimento da levedura em pH 5,0, mas não em pH 7,5. Deste modo, 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 tetrahidrofolato, é responsável pela conversão de dUMP em dTMP, este é posteriormente convertido em dTTP que é incorporado nas novas cadeias de DNA sintetizadas. Ao inserirem o *intein Mtu* RecA no gene T4 *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 dihidrofolato redutase, diminuindo assim a síntese de tetrahidrofolato, 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 tetrahidrofolato e, por causa do Trimetoprim, 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 (no caso de sistemas de seleção positiva), 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 ureia, purificados e restituídos em um precursor estável na presença de zinco, elemento químico conhecido por inibir o *splicing* (Mills & Paulus, 2001; Nichols et al., 2003). O *splicing* pode ser induzido pela adição de EDTA. Este sistema permitiu a avaliação de mais de 85 mil compostos químicos para a inibição do *splicing* do *Mtu RecA* e alguns compostos eletrofílicos se mostraram eficientes devido a sua ligação com o primeiro resíduo de cisteína do domínio Spl, indispensável ao *splicing* protéico (Paulus, 2007).

Aplicações biotecnológicas

“It is a profound and necessary truth that the deep things in science are not found because they are useful, they are found because it was possible to find them”

Robert Oppenheimer

“Inteins....Nature’s Gift to the Protein Chemist”.

Tom Muir

Estas duas frases, sabiamente citadas por Belfort (2008) no primeiro capítulo do livro “Homing endonucleases and *Inteins*”, resumem bem o caminho percorrido por quase tudo descoberto pela ciência, ou seja, a transformação do conhecimento básico em aplicado. Com

os *inteins* a história não foi diferente. Com pouco mais de vinte anos após sua descoberta, eles hoje são considerados ferramentas biotecnológicas, sendo sua função *splicing* extremamente útil no estudo estrutural e engenharia de proteínas e sua função *homing* endonuclease um potencial recurso para terapia gênica e controle de pragas.

Apesar do trabalho desenvolvido nesta tese de doutorado abranger a biologia básica, e não aplicável, dos *inteins*, vale a pena ressaltar nesta revisão bibliográfica um pouco do que vem sendo feito em termos biotecnológicos com estes elementos genéticos parasitas.

O estudo detalhado do domínio Spl tornou possível o total controle da reação de *splicing*, sendo que *inteins* podem ser “ligados” ou “desligados” dependendo de mutagênese sítio-dirigida, alterações na temperatura, pH e sítios de inserção. Vale aqui ressaltar que alguns *inteins* são exigentes quanto aos *exteins* que os flanqueiam, em alguns casos devem ser inseridos e clonados juntamente com alguns nucleotídeos de seu *extein* nativo, quando testados em um contexto não nativo. O número de aminoácidos antes e depois do *intein* que leva à eficiência de *splicing* máxima varia dependendo do *intein* (Perler, 2008).

Uma vez que a clivagem pode ser realizada e controlada ligando-se o *intein* a um *extein* não nativo, logo se cogitou o seu uso para purificação de proteínas. Tradicionalmente a proteína de interesse é expressa juntamente com uma sequência de afinidade *tag*, como a cauda poli-histidina (His *tag*), por exemplo. Porém a remoção da cauda *tag* por proteases e os demais passos de purificação são caros e laboriosos. Além disso, estes métodos são inconvenientes, pois é necessária uma boa purificação da proteína recombinante, o que pode comprometer o rendimento, estabilidade, solubilidade e atividade protéica. Esse procedimento pode ser facilitado por um sistema no qual a proteína é fundida ou na região N ou na C-terminal de um *intein* modificado (o qual só processa clivagem N ou C-terminal), ligado a uma sequência de afinidade *tag*, esta faz com que o precursor, proteína alvo + *intein*, fique imobilizado na coluna e após a reação de clivagem a proteína alvo é liberada, podendo assim ser eluída (Chong & Xu, 2008; Gorgaten et al., 2002). Esta tecnologia, conhecida como *intein-mediated protein ligation* (IPL) ou *expressed protein ligation* (EPL) envolve *splicing* protéico, auto-proteólise e biosíntese de polipeptídios. *Inteins* modificados para processarem uma clivagem

em apenas uma das suas duas junções (N ou C-terminais) em condições pré-determinadas (como por exemplo, adição de thiol ou mudanças na temperatura ou pH) vêm sendo bastante utilizados para isolar proteínas livres de resíduos derivados de vetores (Perler & Adam, 2000).

Outra aplicação dos *inteins* muito estudada atualmente é o uso de *split-inteins* na marcação (*labeling*) sítio específica de proteínas. A marcação de proteínas é extremamente importante para o estudo de sua atividade biológica tanto *in vitro* quanto *in vivo*, além de facilitar a caracterização da proteína alvo quanto a sua estrutura tri-dimensional, dobramento e interação com outras proteínas. Porém, métodos químicos de marcação tradicionais criam na verdade populações mistas de proteínas marcadas, uma vez que as cadeias de aminoácidos a serem marcadas (em seus grupos thiol, carboxil e aminos) estão presentes mais de uma vez na proteína de interesse (Volkman & Liu, 2009).

Os *split-inteins*, como já mencionado, são capazes de processarem um *trans-splicing* protéico. Vários grupos já haviam demonstrado isto tanto *in vivo* como *in vitro* separando as porções N e C terminais de *inteins* (I_N e I_C) e ligando-as a *exteins* separados, ou seja, criando *split-inteins* artificiais (Mills et al., 1998; Wu et al., 1998b; Southworth et al., 1998), quando o primeiro, e até hoje único, *split-intein* natural foi descrito no gene DnaE de *Synechocystis sp* (Wu et al., 1998a). A reação de *trans-splicing* protéico catalisada por estes tipos de *inteins* é considerada uma alternativa promissora para ligar peptídeos sintéticos às proteínas alvo (proteínas recombinantes).

Porém, uma dificuldade neste processo é o tamanho do *intein* (geralmente 100aa no I_N e 40aa no I_C), considerado grande para sua síntese e modificação com os elementos de marcação protéica (Evans et al., 2000). A fim de solucionar este problema, novos *split-inteins* vêm sendo construídos de modo a reduzir suas porções N ou C terminais. Assim, Appleby et al (2009) construíram um *split-intein* (*Ssp* GyrB S11) contendo 150 aa na porção I_N e apenas 6 aa na I_C . A reação de *trans-splicing* foi comprovada *in vivo* (em *E. coli*) entre duas proteínas recombinantes (MBP e tireodoxina) e *in vitro* entre uma MBP e um peptídeo sintético (uma cauda de 6 histidinas). No mesmo ano, Volkman & Liu incorporaram duas marcações, um fluoróforo e uma biotina, na porção C-terminal de uma proteína recombinante (MBP) através da

reação de *trans-splicing* do *Ssp GyrB S11*. Basicamente os autores produziram separadamente dois sistemas de fusão com *intein*, um era constituído por MBP+ I_N+Chitin binding protein (CBP) e o outro por I_C+marcador (L) (fluoróforo ou biotina). Após incubação das duas proteínas a temperatura ambiente por cerca de 16h, foi possível a visualização dos complexos MI_NC (precursor que não sofreu o *trans-splicing*), I_NCBP (*intein* removido) e M^L (MBP ligada ao marcador), por SDS PAGE corado com Coomassie e pela emissão de fluorescência da banda de tamanho correspondente ao M^L.

Outra aplicação dos *split-inteins* é a produção de proteínas cíclicas, área de grande interesse atualmente, uma vez que a conformação cíclica em muitos casos confere maior potência e estabilidade *in vivo* à proteína, resultando em uma maior resistência às proteases. Peptídeos cíclicos podem ocorrer naturalmente, como por exemplo, no caso dos antibióticos ciclosporina A e gramicidina S (Tavassoli et al., 2008). Scott et al (1999) demonstraram a ligação circular mediada por um *split-intein in vivo* usando o *split-intein* natural do *DnaE* de *Synechocystis sp.* Os autores inverteram a ordem dos fragmentos de *intein* da proteína tripartida, assim, ao invés de construir um sistema do tipo Extein1 + I_N e Extein2 + I_C, eles construíram um sistema do tipo I_C + proteína alvo + I_N. Ao se ligarem, as porções I_N e I_C do *intein* sofrem o *splicing* protéico e as porções N e C-terminais da proteína alvo são ligadas, transformando uma proteína antes linear em uma proteína circular, a qual demonstrou total funcionalidade *in vivo* e maior termo estabilidade *in vitro*.

A produção de proteínas circulares também pode ser importante no estudo do funcionamento de uma dada proteína. Por exemplo, sabe-se que a proteína ACP (acyl carrier protein), responsável por doar grupos acil para enzimas responsáveis pela formação de fosfolipídios e outros produtos lipídicos essenciais para a sobrevivência da célula bacteriana, possui o seu centro hidrofóbico, doador de grupos acil no meio de sua conformação tri-dimensional. Assim uma pergunta básica a respeito do funcionamento desta proteína seria se ela mudaria para uma conformação mais linear permitindo a liberação destes grupos acil. A fim de responder esta questão, Volkmann et al (2010) criaram uma ACP cíclica usando *split-inteins Ssp GyrB*, porém a funcionalidade desta proteína não foi comprometida por sua circularização.

Portanto os autores sugerem que a ACP não necessita de um desdobraimento completo ou de suas pontas N e C-terminais livres para desempenhar seu papel dentro da célula, o que corrobora outras hipóteses que sugerem alta flexibilidade e dinamismo desta proteína ao interagir com várias enzimas.

Homing Endonucleases também apresentam potencial biotecnológico. Estas enzimas reconhecem de 12 a 40pb, podendo tolerar algumas substituições ao longo do sítio de reconhecimento, mas ainda assim são consideradas enzimas de alta especificidade e seguras, ou seja, não trazem grandes riscos ao hospedeiro devido à raridade de sítios alvo no genoma. Além disso, a função *homing* (cópia da endonuclease em um sítio antes vazio) corrompe o sítio de reconhecimento, não sendo necessárias metilações para proteção do DNA hospedeiro contra clivagens sucessivas. Acredita-se que o entendimento estrutural e funcional da interação destas enzimas com o DNA alvo pode abrir novas possibilidades de terapia gênica através do desenho e produção de enzimas customizadas, cujo alvo será uma sequência de DNA de interesse como, por exemplo, genes defeituosos (Marcaida et al., 2009).

Basicamente, a terapia gênica consiste em reparo ou compensação de um gene defeituoso por um alelo funcional. Entretanto, a maioria das estratégias usa vetores virais para a incorporação do gene normal e, muitas vezes a integração semi-randômica destes vetores no genoma trazem limitações, como silenciamento e rompimento de genes endógenos e ativação transcricional de genes próximos. Uma alternativa seria inserir na célula (abordagem *ex vivo*) uma *homing* endonuclease modificada (que reconheça o gene defeituoso como alvo para o corte) juntamente com uma cópia normal do gene (que será usada como DNA molde). Após o corte (*double break strand*, DBS) e recombinação homóloga (usando a cópia normal como molde), o gene é reparado (Gimble et al., 2007; Marcaida et al., 2009).

A recombinação homóloga após o DBS foi demonstrada usando a endonuclease I-SceI, codificada por um intron mitocondrial de *S. cerevisiae*, em um alvo cromossômico de células de mamífero, no qual o DBS foi reparado com uma sequência homóloga doadora da região que flanqueava o sítio de reconhecimento da endonuclease (Choulika et al., 1995).

Genes parasitas, como os HE sítio-específicos, estão recebendo especial atenção para seu uso como ferramentas para o controle e engenharia de populações naturais consideradas prejudiciais à saúde humana (como por exemplo, mosquitos transmissores da malária), ou à produção agropecuária.

Neste sentido, Burt (2003) revisou aspectos potenciais básicos do uso desta tecnologia, bem como suas vantagens sobre os métodos tradicionais já conhecidos como controle químico, biológico e genético (usando elementos transponíveis). Segundo Burt (2003) três aspectos estariam envolvidos neste processo de controle de pragas. Primeiramente, um gene HE seria produzido de forma a reconhecer e cortar uma sequência no meio de um gene essencial, sendo então copiado via *homing* para este sítio, interrompendo o gene e ao mesmo tempo evitando novos cortes. Isso obviamente não ocorre naturalmente, já que as HEs, quando em genes conservados e importantes, ocorrem sendo flanqueadas por elementos autorremovíveis como introns do grupo I e II e *inteins*. Em segundo lugar, o gene alvo seria escolhido de forma que a mutação teria pouco ou nenhum efeito fenotípico quando em heterozigose, mas um efeito letal em homozigose. Finalmente o gene HE estaria sob controle de um promotor específico de meiose. Assim, os zigotos heterozigotos se desenvolveriam normalmente, mas transmitiriam o gene HE em uma razão desproporcional para os seus gametas devido à herança Super-Mendeliana gerada pelo processo *homing*. Uma vez que este gene HE é introduzido na população, inicialmente apareceria em heterozigose e logo apresentaria uma distorção na razão de transmissão (TRD). O HE aumentaria em frequência até que os efeitos deletérios e o TRD entrassem em equilíbrio. Portanto, em uma população idealizada, onde os cruzamentos são aleatórios, o HE só é letal em homozigose e TRD é o mesmo para machos e fêmeas, a frequência de equilíbrio do gene HE (q) é $q=e$ onde e é a probabilidade de um alelo HE^- virar HE^+ ($e=0$ em heranças Mendelianas). Assim, o número de homozigotos resultante de cruzamentos onde o esforço reprodutivo será em vão (já que a homozigose de HE é letal) será $L= q^2$, sendo o valor adaptativo médio da população $w=1-e^2$. Por exemplo, genes HE de leveduras possuem TRD alto, $e=0,9$, ou seja, de um cruzamento $HE^+ \times HE^-$, onde a prole esperada seria 50% de HE^+ e 50 % de HE^- em uma herança mendeliana, observa-se cerca de

90% de HE⁺. Se, um HE artificialmente produzido possuísse este TRD e fosse inserido em uma população, após 12 gerações, o *w* médio da população seria apenas 0,19 e o gene HE estaria fixado em cerca de 90% dos alelos.

Segundo Burt (2003), este sistema possui certa estabilidade evolutiva, uma vez que formas mutantes irão se espalhar na população, mas depois serão negativamente selecionadas e, portanto perdidas. O que também aconteceria com um HE mutante que perdesse a capacidade de reconhecer o sítio de corte na sequência alvo, pois apesar de ter o TRD reduzido, não diminuiria o dano causado ao hospedeiro sendo, portanto, eliminada por seleção. Além disso, o sistema pode ser reversível pela inserção de alelos resistentes ao HE na população (já que o código genético é degenerado, o sistema poderia ser revertido inserindo sequências diferentes, sem o sítio de reconhecimento, mas com a codificação para os mesmos aminoácidos).

Vale ressaltar que estes estudos estão em fase inicial e, portanto ainda longe de serem de fato aplicados. Muitos parâmetros devem ainda ser estudados cautelosamente, principalmente quando se considera a liberação de transgenes em populações naturais. De qualquer forma, são ideias criativas e inovadoras que devem primeiramente passar por uma longa fase de experimentação laboratorial antes de serem testadas em campo.

Muitos outros estudos de aplicações biotecnológicas usando endonucleases e atividades de *splicing* de *inteins* vêm sendo continuamente, publicados em revistas de alto impacto. Todos estes artigos podem ser encontrados na base de dados sobre *inteins*, o *Inbase* (<http://www.neb.com/neb/inteins.html>).

Referências Bibliográficas

- Appleby, J.H., Zhou, K., Volkmann, G., Liu, X.Q., 2009. Novel Split *Intein* for trans-Splicing Synthetic Peptide onto C Terminus of Protein. *J. Biol. Chem.*, 284, 6194–6199.
- Belfort, M. ., Derbyshire, V., Stoddard, B.L., Wood, D.W. (eds). 2008. Homing Endonucleases and *Inteins*. 2ed Springer – Verlag Berlin Heidelberg.
- Belfort, M., Roberts, R.J., 1997. Homing endonucleases – keeping the house in order. *Nucleic acids Res.* 25, 3379-3388.

- Boekhout T, Theelen B, Diaz M, Fell JW, Hop WC, Abeln EC, Dromer F, Meyer W 2001. Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. *Microbiol.* 147, 891-907.
- Bokor, A.A.M., Kan, J.A.L.V., Poulter R.T.M., 2010. Sexual mating of *Botrytis cinerea* illustrates PRP8 *intein* HEG activity. *Fungal Genet. Biol.* doi: 10.1016/j.fgb.2010.01.003.
- Burt, A., 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. B.* 270, 921-928.
- 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.L., Goodwin, T.J.D., Poulter, R.T.M., 2001. A nuclear-encoded *intein* in the fungal pathogen *Cryptococcus neoformans*. *Yeast* 18, 1365-1370.
- 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.
- Carrero LL, Niño-Vega G, Teixeira MM, Carvalho MJA, Soares CMA, Pereira M, Jesuino RSA, McEween JG, Mendoza L, Taylor J W, Felipe MS, San-Blas G 2008. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. *Fungal Genet. Biol.* 45(5), 605-612.
- Casadevall, A., Steenbergen, J.N., Nosanchuk, J.D., 2003. "Ready – made" virulence and "dual use" virulence factors in pathogenic environmental fungi – The *Cryptococcus neoformans* paradigm. *Curr. Opin. Microbiol.* 6, 332-7.
- Chong, S., Xu, M.Q., 2008. Harnessing *Inteins* for protein purification and characterization. In: Belfort, M., Derbyshire, V., Stoddard, B.L., Wood, D.W. (eds). *Homing Endonucleases and Inteins*. 2ed Springer – Verlag Berlin Heidelberg.
- Choulika, A., Perrin, A., Dujon, B., Nicolas, J.F., 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 15, 1968-1973.
- 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.
- Cooper, A.A., Chen, Y.J., Lindorfer, M.A., Stevens, T.H., 1993. Protein *splicing* of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J.* 12, 2575-2583.
- Dalgaard, J.Z., Klar, A.J., Moser, M.J., Holley, W.R., Chatterjee, A., Mian, I.S., 1997. 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.

- Davis, E.O., Jenner, P.J., Brooks, P.C., Colston, M.J., Sedgwick, S.G., 1992. Protein *splicing* in the maturation of *M. tuberculosis* recA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* 71, 201-210.
- Dawkins R 1989. *The selfish gene*. Oxford Univ. Press: Oxford, UK, 540pp.
- Derbyshire, V., Belfort, M., 1998. Lightning strikes twice: intron-*intein* coincidence. *Proc. Natl. Acad. Sci. USA* 95, 1356-1357.
- Derbyshire, V., Wood, D.W., Wu, W., Dansereau, J., Dalgaard, J.Z., Belfort, M., 1997. 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.
- Elleuche S, Döring K, Pöggeler S 2008. Minimization of a eukaryotic mini-*intein*. *Biochem Biophys. Res. Commun.* 366, 239-243.
- Ellis, D., Marriott, D., Hajjeh, R.A., Warnock, D., Meyer, W., Barton, R., 2000. Epidemiology, surveillance of fungal infections. *Med. Mycol.* 38(1), 173–182.
- Evans, T.C.Jr, Martin, D., Kolly, R., Panne, D., Sun, L., et al 2000. Protein trans-*splicing* and cyclization by a naturally split *intein* from the dnaE gene of *Synechocystis* species PCC6803. *J. Biol. Chem.*, 275, 9091–9094.
- Franzot, S.P., Salkin, I.F., Casadevall, A., 1999. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. *J. Clin. Microbiol.* 37, 833–840.
- Gangopadhyay, J.P., Jiang, S.Q., Paulus, H., 2003. An in vitro screening system for protein *splicing* inhibitors based on green fluorescent protein as an indicator. *Anal. Chem.* 75, 2456-2462.
- Gimble, F. S. (2007). Engineering *homing* endonucleases to modify complex genomes. *Gene Therapy & Regulation*, 3(1), 33-50.
- 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.
- Gimble, F.S., Thorner, J., 1992. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, 357, 301-306.
- Goddard, M.R., Burt, A., 1999. Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. USA* 96, 13880-13885.
- Gogarten, J.P., Hilario, E., 2006. *Inteins*, introns, and *homing* endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* 6, 1-5.
- Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L., Hilario, E., 2002. *Inteins*: Structure, Function and Evolution. *Ann. Rev. Microbiol.* 56, 263-287.
- Grainger, R.J., Beggs, J.D., 2005. Prp8 protein: at the heart of the spliceosome. *RNA* 11, 533-557.
- Haber, J.E., Wolfe, K.H., 2005. Function and Evolution of HO and VDE Endonucleases in Fungi. *Nucleic Acids Mol. Biol.* 16, 161-175.

- 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.
- Hodges, R.A., Perler, F.B., Noren, J.N., Jack, W.E., 1992. Protein *splicing* removes intervening sequences in an archaea DNA polymerase. Nucleic Acids Res. 20, 6153-6157.
- Hoog, G.S., Guarro, J., Gené, J., Figueras, M.J., 2000. Atlas of Clinical Fungi. Centraalbureau voor Schimmelcultures, Utrecht.
- 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.
- Kasuga T, White TJ, Koenig G, Mcewen J, Restrepo A, Castañeda E, Lacaz CS, Heins-Vaccari, Em, Freitas RS, Zancopé-Oliveira RM, Qin Z, Negroni R, Carter DA, Mikami Y, Tamura M, Taylor MI, Miller GF, Poonwan N, Taylor JW 2003. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. Mol. Ecol. 12, 3383-3401.
- 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.
- Kwon-Chung, K.J., Boekhout, T., Fell, J.W., Diaz, M 2002. Proposal to conserve the name *Cryptococcus gattii* against *C. hondurianus* and *C. bacillisporus*, Basidiomycota, Hymenomycetes, Tremellomycetiadae. Taxon 51: 804-806.
- Lambowitz, A.M., Zimmerly, S., 2004. Mobile group II introns. Annu. Rev. Genet. 38, 1-35.
- Liu, X.Q., 2000. Protein-Splicing *Intein*: Genetic Mobility, Origin and Evolution. Ann. Rev. Genet. 34, 61-76.
- Liu, X.Q., Hu, Z.A., 1997. DnaB *intein* in *Rhodothermus marinus*: indication of recent *intein* *homing* across remotely related organisms. Proc. Natl. Acad. Sci. USA 94, 7851-7856.
- Liu, X.Q., Yang, J., 2004. Prp8 *intein* in fungal pathogens: target for potential antifungal drugs. FEBS Lett. 572, 46-50.
- Marcaida, M.J., Munoz, I.G., Blanco, F.J., Prieto, J., Montoya, J.M., 2009. Homing endonucleases: from basics to therapeutic applications. Cell. Mol. Life Sci., doi 10.1007/s00018-009-0188-y.
- 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., 1999. 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.
- Mills, K.V., Lew, B.M., Jiang, S., Paulus, H., 1998. Protein *splicing* in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA *intein*. Proc. Natl. Acad. Sci. USA 95, 3543-3548.

- Mills, K.V., Paulus, H., 2001. Reversible Inhibition of Protein Splicing by Zinc Ion. *J. Biol. Chem.* 276, 10832-10838.
- Nichols, N.M., Benner, J.S., Martin, D.D., Evans, T.C. Jr., 2003. Zinc Ion Effects on Individual *Ssp DnaE Intein* Splicing Steps: Regulating Pathway Progression. *Biochemistry* 42, 5301-5311.
- Paulus, H., 2007. Protein *splicing* inhibitors as a new class of antimycobacterial agents. *Drugs of the Future* 32(11), 973 – 984.
- Pearl EJ, Tyndall JDA, Poulter RTM, Wilbanks SM 2007. Sequence requirements for *splicing* by the Cne PRP8 *intein*. *FEBS letters* 581, 3000-3004.
- Perler, F.B., 2000. InBase, the *Intein* Database. *Nucleic acids res.* 28, 344-345.
- Perler, F.B., Olsen, G.J., Adam, E., 1997. Compilation and analysis of *intein* sequences. *Nucleic acids res.* 25, 1087-1093.
- Perler, F.B. 2008. *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.
- Perler, F.B., Adam, E., 2000. Protein *splicing* and its applications. *Current Opinion in Biotechnology* 11, 377-383.
- Perler, F.B., Xu, M.Q., Paulus, H., 1997. Protein *splicing* and autoproteolysis mechanisms. *Curr. Opin. Chem. Biol.*, 1, 292-299.
- Perler, F.B., Olsen, G.J., Adam, E., 1997. Compilation and analysis of *intein* sequences. *Nucleic acids res.* 25, 1087-1093.
- Petrokovski, S., 2001. *Intein* spread and extinction in evolution. *Trends Genet.* 17, 465–472.
- Petrokovski, S., 2001. *Inteins-Proteins Introns*. <http://bioinfo.weizmann.ac.il/~pietro/inteins>.
- 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.
- Scott, C.P., Abel-Santos, E., Wall, M., Wahnou, D.C., Benkovic, S.J., 1999. Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. USA* 96, 13638-13643.
- Snustad, D.P., Simmons, M.J., 2000. *Fundamentos de Genética* 2ed Guanabara Koogan EDS: Rio de Janeiro RJ.
- Southworth, M.W., Adam, E., Panne, D., Byer, R., Kautz, R., Perler, F.B., 1998. Control of protein *splicing* by *intein* fragment reassembly. *EMBO J.* 17, 918–926.
- Steenbergen, J.N., Casadevall, A., 2003. The origin and maintenance of virulence for the pathogenic fungus *Cryptococcus neoformans*. *Microb. Infect.* 5, 667-75.

- Steenbergen, J.N., Nosanchuk, J.D., Malliaris, S.D., Casadevall, A., 2004. Interaction of *Blastomyces dermatitidis*, *Sporothrix schenckii*, and *Istioplasma capsulatum* with *Acanthamoeba castellanii*. *Infect. Immun.* 72, 3478-88.
- Swithers, K.S., Senejani, A.G., Fournier, G.P., Gogerten, J.P., 2009. Conservation of intron and *intein* insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol. Biol.* 9:303, 1-11.
- Tavassoli, A., Naumann, T.A., Benkovic, S.J., 2008. Production of cyclic Proteins and Peptides. In: Belfort, M., Derbyshire, V., Stoddard, B.L., Wood, D.W. (eds). *Homing Endonucleases and Inteins*. 2ed Springer – Verlag Berlin Heidelberg.
- 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. 2009. Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides* genus. *Mol Phylogenet Evol.* 52, 273-283.
- Theodoro, R.C., Bagagli, E., 2009. *Inteins* in pathogenic fungi: a phylogenetic tool and perspectives for therapeutic applications. *Mem. I. Oswaldo Cruz* 104, 497-504.
- Theodoro, R.C., Bagagli, E., Oliveira, C., 2008. Phylogenetic analysis of PRP8 *intein* in *Paracoccidioides brasiliensis* species complex. *Fungal Genet. Biol.* 45, 1284-1291.
- Torres, I., Garcia, A.M., Hernández, O., González, A., McEwen, J.G., Restrepo, A., Arango, M., 2009. Presence and expression of the mating type locus in *Paracoccidioides brasiliensis* isolates. *Fungal Genet. Biol.*, doi:10.1016/j.fgb.2009.11.005.
- Untereiner, W.A., Scott, J.A., Naveau, F.A., Sigler, L., Bachewish, J., Angus, A., 2004. The *Ajellomycetaceae*, a new family of vertebrate-associate Onygenales. *Mycol.* 96, 812-821.
- Vilela, R., Rosa P.S., Belone, A.F., Taylor, J.W., Diorio, S.M., Mendoza, L., 2009. Molecular phylogeny of animal pathogen *Lacazia loboi* inferred from rDNA and DNA coding sequences. *Mycol. Res.* 113, 851-857.
- Volkman, G., Liu, X.Q., 2009. Protein C-terminal labeling and biotinylation using synthetic peptide and split-*intein*. *PLoS One*, doi: 10.1371/journal.pone.0008381.
- Volkman, G., Murphy, P.W., Rowland, E.E., Cronan, J.E.Jr, Liu, X.Q., Blouin, C., Byers, D.M., 2010. *Intein*-mediated cyclization of bacterial acyl carrier protein stabilizes its folded conformation but does not abolish function. *J. Biol. Chem.*, doi/10.1074/jbc.M109.060863.
- Wu, H., Hu, Z., Liu, X.Q., 1998a. 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.
- Wu H, Xu MQ, Liu XQ., 1998b. Protein trans-*splicing* and functional mini-*inteins* of a cyanobacterial *dnaB intein*. *Biochim. Biophys. Acta* 1387, 422–432.
- 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.

Artigo 1



Contents lists available at ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbiPhylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complexRaquel C. Theodoro^a, Eduardo Bagagli^{a,*}, Claudio Oliveira^b^aDepto. Microbiologia e Imunologia, Instituto de Biociências, UNESP-Botucatu, Distrito de Rubião Junior, s/n 18618 000 Botucatu-Sao Paulo, Brazil^bDepto. Morfologia, Instituto de Biociências, UNESP-Botucatu, Brazil

ARTICLE INFO

Article history:

Received 20 March 2008

Accepted 6 July 2008

Available online 11 July 2008

Keywords:

Fungi

P. brasiliensis

Cryptic species

DNA sequencing

Phylogeny

Pb01 isolate

PRP8 intein

ABSTRACT

A recent species status investigation of the pathogenic fungus *Paracoccidioides brasiliensis* suggested the existence of three cryptic species. In the present study the sequences of the PRP8 intein from *P. brasiliensis* isolates belonging to the three described genetic groups and two unidentified isolates were determined and analyzed in order to check their functionality and usefulness for species identification. All the isolates presented a full-length intein, although the Endonuclease domain seems to be inactive due to substitutions in the second essential aspartic acid residue. Phylogenetic analysis by Maximum-Parsimony, Maximum Likelihood, and Bayesian Analysis clearly separated the isolates from the three species and revealed a significant difference between the Pb01 isolate and the remaining ones. The Pb01 isolate does not belong to any of the groups previously described since it presented a high divergence level compared to the three different genetic groups, corroborating some previous studies that suggested this isolate represents a new species of *Paracoccidioides*.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Paracoccidioides brasiliensis is the etiologic agent of *Paracoccidioidomycosis* (PCM), the most important systemic mycosis in Latin America (Franco, 1987). This pathogen presents a thermal dimorphism growing at 25 °C as mycelia, producing its infective propagula, and at 36 °C as budding yeast cells (Lacaz et al., 1984). While this fungus has been rarely isolated from saprobic sources, it has frequently been recovered from clinical samples and from tissues of the nine-banded armadillo, *Dasypus novemcinctus* (Bagagli et al., 2003). It was also recently detected in and isolated from dogs (Ricci et al., 2004; Farias et al., 2005). Molecular polymorphism analysis suggested that the same "ecopathogenotypes" can infect both human and animals (Hebeler-Barbosa et al., 2003).

Paracoccidioides brasiliensis has long been considered a unique species, mainly because of its peculiar microscopic aspects, both in parasitic and culture conditions. This fact, although well-accepted for practical purposes and medical diagnosis, has been challenged by several pieces of evidence. This fungus presents many variable phenotypes such as growth, mycelial colony, conidia production, mycelia-to-yeast transition, yeast microscopy (budding, size and shape of cells), virulence, thermal tolerance (Macoris et al., 2006; Theodoro et al., 2008) and clinical manifestation (Mendes, 1994).

Recent phylogenetic studies on *P. brasiliensis* detected three genetic groups, suggesting that this fungus consists of at least three distinct, previously unrecognized species, identified as: S1 (species

1 from Brazil, Argentina, Paraguay, Peru, and Venezuela), PS2 (phylogenetic species 2 from Brazil and Venezuela) and PS3 (phylogenetic species 3 from Colombia). It was also observed that while S1 and PS2 were recombining sexual species, the reproduction of PS3 was clonal (Matute et al., 2006a). It is interesting to note that S1 and PS2 are sympatric and, despite the fact that they have sexual reproductive attributes, they are reproductively isolated (Matute et al., 2006a). These data were corroborated by microsatellite analysis of the same isolates used in the previous gene genealogy (Matute et al., 2006b), whereas the divergence time between the species PS2 and PS3 was estimated in 8.04–8.37mya, by analysis of the *chitin synthase* gene (Matute et al., 2007).

The phylogenetic species concept, which detects genetic divergence among groups through multi-locus sequence typing by concordance of gene genealogies, seems to be the most appropriate method for fungal species recognition. However, acquiring DNA sequences from several loci, with sufficient genetic variation, can be extremely laborious, making the use of hypervariable molecular markers very attractive for easier species recognition (Taylor et al., 2000).

Several nuclear protein-coding sequences have been used to differentiate the three cryptic species in *P. brasiliensis* (Matute et al., 2006a, 2007; Carrero et al., 2008). GP43, which codes for the main antigenic component of *P. brasiliensis*, is one of the most polymorphic genes (Morais et al., 2000) and presents a high level of variability among the three cryptic species (Matute et al., 2006a).

Inteins, known as parasitic genetic elements, seem to constitute a promising source of phylogenetic information (Butler and

* Corresponding author. Fax: +55 14 3815 3744.
E-mail address: bagagli@ibb.unesp.br (E. Bagagli).

Poulter, 2005). An intein is a specific insertion within a host protein that is excised during protein maturation, in which the intein itself (“internal protein”) catalyzes its splicing from the host protein (extein) and the subsequent ligation of the flanking host protein segments (Xu et al., 1993; Chong et al., 1996). There are two intein types, the mini-inteins and the full-length inteins, both of which present a splicing domain containing four blocks of nucleotide sequence. Blocks A and B constitute the N-terminal splicing domain and the blocks F and G constitute the C-terminal splicing domain. The full-length inteins present, additionally, an endonuclease domain that can perform a homing function (Homing endonuclease gene or HEG) that makes the intein a mobile genetic element, resulting in occupation of empty alleles and duplication of the parasitic genetic element (Liu, 2000). There are different kinds of endonuclease that are encoded by the HEG domain, but the endonucleases from the LAGLIDADG family are the most frequent ones (Perler, 2000). These large inteins are expected to have more sequence variation in the endonuclease domain than in the splicing domain due to a more relaxed selection, especially if the HEG is no longer active (Butler et al., 2006), constituting, therefore, a promising region for phylogenetic studies.

According to the Intein database, the *InBase*, 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 metabolic enzymes, DNA and RNA polymerases, proteases and others (Liu, 2000). For this reason, any mutation that disrupts the functionality of splicing domain can be fatal for the cell. Although the inteins do not present any improvement in the fitness of the organism, being considered parasitic genetic elements, the location of inteins in important and therefore conserved sites, can be the reason for their survival in genomes, since they recognize specific and conserved sites for the homing process. Furthermore, some of the genes that host inteins are known to be horizontally transferred, increasing the frequency of some inteins in non-related organisms (Liu, 2000; Pietrokovski, 2001).

In fungi, the most studied intein is the VMA intein encoded in the vacuolar ATPase gene of some Hemiascomycetes, such as *Saccharomyces cerevisiae* (Kane et al., 1990; Koufopanou et al., 2002; Okuda et al., 2003) and the PRP8 intein, found in the gene encoding PRP8, a highly conserved protein involved in spliceosome forma-

tion. This intein was first found as a mini-intein in *Cryptococcus gattii* and *Cryptococcus neoformans* varieties *grubii* and *neoformans* and as a full-length intein in *Cryptococcus laurentii* (Butler et al., 2001).

Since the gene *prp8* is highly conserved among the eukaryotes, the primer design for its amplification by PCR is possible for phylogenetic studies of fungi species and varieties. Using this approach, 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).

The PRP8 inteins have also been detected in fungal species from three different orders of Euscomycetes: Eurotiales (*Aspergillus nidulans* and *Aspergillus fumigatus*), Onygenales (*Paracoccidioides brasiliensis* and *Histoplasma capsulatum*) and Helotiales (*Botrytis cinera*). There is no evidence of horizontal transference of PRP8 intein in these groups of fungi. It has been suggested that the PRP8 intein in Euscomycetes had an ancient origin and a vertical transfer (Butler et al., 2006).

The PRP8 intein of *P. brasiliensis* contains both splicing and homing endonuclease domains. Its protein sequence contains 573 amino acids and presents 53% similarity to the PRP8 intein of *H. capsulatum* (Butler et al., 2006). Considering the potential of PRP8 inteins for phylogenetic studies, this paper aimed to describe and analyze the diversity of PRP8 intein in the putative cryptic species within *P. brasiliensis*.

2. Materials and methods

2.1. *Paracoccidioides brasiliensis* isolates and cultures

Twenty-two isolates were used in this study: 13 from S1 group, 5 from PS2 group, 2 from PS3 group and 2 isolates without previous recognition (Table 1). The isolates employed herein were obtained mainly from nine-banded armadillos, which offer the advantage by providing the correct geographic origin of the isolates, because these animals present a small home range and no migration habits (Bagagli et al., 2003). All isolates from armadillos used in this work were previously characterized by Matute et al. (2006a). The isolates were maintained in yeast forms at 36 °C in 2% glucose, 1% peptone, 0.5% yeast extract, and 1% agar (GPYA OX-01D Ltd., Basingstoke, UK).

Table 1
Isolates used in this work and their origin

Isolate	Genetic group	Locality	Host	Origin	References
T1F1	S1	Pratania, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Liver	Hebeler-Barbosa et al. (2003)
T3B6	S1	Pratania, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Spleen	Hebeler-Barbosa et al. (2003)
T4B17	S1	Manduri, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Spleen	Hebeler-Barbosa et al. (2003)
T5LN1	S1	Botucatu, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Mesenteric lymph node	Hebeler-Barbosa et al. (2003)
T7F6	S1	Manduri, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Liver	Hebeler-Barbosa et al. (2003)
T8LN2	S1	Botucatu, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Mesenteric lymph node	Hebeler-Barbosa et al. (2003)
T9B1	S1	Botucatu, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Spleen	Hebeler-Barbosa et al. (2003)
T13LN2	S1	Manduri, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Mesenteric lymph node	Hebeler-Barbosa et al. (2003)
T15LN1	S1	Manduri, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Mesenteric lymph node	Hebeler-Barbosa et al. (2003)
BT60	S1	Manduri, Sao Paulo, Brazil	Human	Sputum	Hebeler-Barbosa et al. (2003)
Pb18	S1	Sao Paulo, Sao Paulo, Brazil	Human	Ganglionar pus	Teixeira et al. (1987)
Pb339	S1	Sao Paulo, Sao Paulo, Brazil	Human	Unknown	Restrepo-Moreno and Schneidau (1967)
Pb927	S1	Antartida/Uruguay	<i>Pygoscelis adeliae</i>	Feces	Garcia et al. (1993)
BT84	PS2	Botucatu, Sao Paulo, Brazil	Human	Sputum	Hebeler-Barbosa et al. (2003)
T10B1	PS2	Botucatu, Sao Paulo, Brazil	<i>Dasybus novemcinctus</i>	Spleen	Hebeler-Barbosa et al. (2003)
Pb262	PS2	Uberlandia, Minas Gerais, Brazil	-	Dog food	Ferreira et al. (1990)
Pb02	PS2	Caracas, Venezuela	Human	Unknown	Morais et al. (2000)
Pb04	PS2	Sao Paulo, Sao Paulo, Brazil	Human	Unknown	Morais et al. (2000)
EPM83	PS3	Colombia	Human	Unknown	Unknown
EPM77	PS3	Colombia	Unknown	Unknown	Unknown
Pbdog	Unknown	Curitiba, Parana, Brazil	<i>Canis familiares</i>	Popliteal lymph node	Farias et al. (2005)
Pb01	Unknown	Brasilia, Distrito Federal, Brazil	Human	Unknown	Unknown

2.2. DNA extraction, PCR and sequencing of PRP8 intein

DNA was extracted according to McCullough et al. (2000), with initial cell disruption with glass beads (425–600 µm, acid washed, Sigma, St. Louis, MO, USA) in solution of 1 M sorbitol and 125 mM EDTA. The DNA was quantified by agarose gel electrophoresis and diluted in ultrapure water to 10 ng/µL.

The primers for PCR amplification were designed with Primer Express software (Applied Biosystems), for the PRP8 intein from *P. brasiliensis* (GenBank Accession No. DQ285419) (Butler et al., 2006). Two primer sets were designed, one for the amplification of the entire intein (for the sequencing of 5' and 3' ends, corresponding to N- and C-terminal splicing domains) and the other for the core amplification of the PRP8 intein (corresponding to the endonuclease domain). The PCRs for the entire PRP8 were carried out in 25 µL of reaction mixture containing 20 ng of genomic DNA, 50 mM KCl, 10 mM Tris-HCl, 1.5 mM MgCl₂, 0.2 mM dNTP, 0.4 µM of each primer: RCTfwd2 (5' GCCTATCGAACCGATGTTATCC 3') and RCTrev2 (5' TGAACCTGGAAGCCAACGTAG 3') and 1 U of Taq polymerase (GE Healthcare), in a thermalcycler (Mastercycler gradient, Eppendorf, Hamburg, Germany). 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, 72 °C for 2 min and a final cycle at 72 °C for 5 min. The same procedure was carried out with the primers FwdIn3 (5' CATCGTGACCATGCTGC 3') and RevIn2 (5' TGAAGCCGAACGTGT TTTCTG 3') for core amplification of the PRP8 intein. The PCR products were identified by 1.5% agarose gel electrophoresis stained with ethidium bromide.

The PCR amplicons were purified by the commercial kit GFX PCR DNA and Gel Band (GE Healthcare), whereas the sequencing reactions were carried out on both strands, in triplicates, using the DYEnamic ET Dye Terminator Kit (with Thermo Sequenase™ II DNA Polymerase) code US81090 (GE Healthcare). The MegaBACE 1000 DNA Analysis System was used for capillary electrophoresis, and the chromatogram visualized by the Chromas program. The sequences were sent to blastn for comparison with the NCBI database (<http://www.ncbi.nlm.nih.gov/BLAST>) and to blastn of the Broad Institute website for comparison with the sequences from the isolates Pb18, Pb01, and Pb03, available in the *P. brasiliensis* genome database (http://www.broad.mit.edu/annotation/genome/paracoccidioides_brasiliensis).

2.3. Sequence analysis

The sequences of each isolate were initially analyzed with the software MEGA version 4.0 (Kumar et al., 2004), where the sequences from 3' and 5' ends and from the core of the PRP8 intein were joined and a consensus sequence of the entire PRP8 intein was obtained for each isolate. After that, all sequences were aligned using the software DAMBE (Xia and Xie, 2001). Nucleotide variation, substitution patterns, and genetic distances were examined using MEGA version 4.0 (Kumar et al., 2004).

The synonymous and non-synonymous substitution rates (dS/dN) were calculated at the syn-Scan web site (Gonzales et al., 2002), according to the Nei-Gojobori method (Nei and Gojobori, 1986) for the splicing and endonuclease domains in order to address the question of endonuclease activity. The translated sequences were compared to the VMA Intein from *S. cerevisiae* (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).

Phylogenetic analyses, based on Maximum-parsimony (MP), were performed using the software PAUP* beta version 4.0b10 (Swofford, 2002) with heuristic searches using random addition of sequences and the Tree Bisection and Reconnection (TBR) algo-

rihm. Parsimony trees were generated using Ti/Tv ratios of 1:1 and 1:2, considering gaps as missing data or as a fifth base. Bootstrap re-sampling (Felsenstein, 1985) was applied to assess support for individual nodes using 1000 replicates with random additions and TBR branch swapping.

Maximum likelihood (ML) phylogenetic analyses were performed using the website software version of PHYML (Guindon and Gascuel, 2003; Guindon et al., 2005). The Kimura 2-parameter (Kimura, 1980) was used as distance model according to Modeltest 3.7 (Posada and Crandall, 1998). The Ti/Tv ratio, gamma shape parameter, and proportion of non-variant sites were estimated by Maximum Likelihood from a Neighbor-joining tree (BIONJ). Bootstrap re-sampling (Felsenstein, 1985) was applied to assess support for individual nodes using 500 replicates with random additions and TBR branch swapping. The best genetic distance model was also computed by Modeltest 3.7 (Posada and Crandall, 1998) including one PRP8 intein sequence from *H. capsulatum* (teleomorphic phase: *Ajellomyces capsulatus*) from GenBank (Accession No. DQ467836).

Bayesian analyses were performed using MrBayes v. 3.0b4 (Ronquist and Huelsenbeck, 2003). Four chains were run simultaneously for 1,000,000 generations using MrBayes analysis. Every 100th generation was sampled and the asymptote of the likelihood score was detected with the SUMP command. The above procedure was repeated two times. All sampled topologies beneath the asymptote (10,000 generations in the present study) were discarded from the tree populations considered in the subsequent majority-rule consensus. The frequency with which a particular clade appeared in the population of retained topologies was interpreted as its posterior probability. Posterior probabilities were interpreted as a measure of the likelihood that the clade would appear in the optimal topology rather than accuracy of the node with respect to isolates relationships or clade stability. Consensus trees were produced with the software TreeExplorer implemented in MEGA version 4.0 (Kumar et al., 2004).

3. Results

3.1. PCR and sequencing of the PRP8 intein

The PCR, with the primers RCTfwd2 and RCTrev2, amplified fragments of 1919 bp whereas the PCR with the primers FwdIn3 and RevIn2 amplified fragments of 857 bp (data not shown). The entire PRP8 intein was sequenced for all the isolates, presenting 95.74–99.8% similarity with a sequence previously deposited in the GenBank accession No. (DQ285419) (Butler et al., 2006). This previously deposited sequence is from a Pb18 isolate and presented three nucleotide differences (in the block/sites B431, C973, and C1215) when compared to the Pb18 used in the present study. Our PRP8 intein sequence from Pb18 was also compared to the sequence from the Pb18 genome project (Supercontig 4, nucleotides 2813781–2815499 at the website http://www.broad.mit.edu/annotation/genome/paracoccidioides_brasiliensis) and presented 4 nucleotide differences (in the block/sites C806, C968, C1067, and D1309). The sequence obtained for Pb01 isolate was also compared by blastn at the same website and it showed 100% similarity to the supercontig 20, nucleotides 241560–243269 of the Pb01 genome. The sequences obtained during the present study have been deposited in the GenBank database under the following Accession Nos. EU561021–EU561042.

The six haplotypes found among the samples analyzed were identified as: (1) T1F1-like (isolates T1F1, T3B6, T4B17, T5LN1, T7F6, T8LN2, T9B1, T13LN2, T15LN1, BT60, Pb18, and Pb339); (2) Bt84-like (isolates Bt84, Pb02, Pb927, Pbdog, and Pb03); (3) EPM77-like (isolates EPM77 and EPM83); (4) T10B1-like (isolates

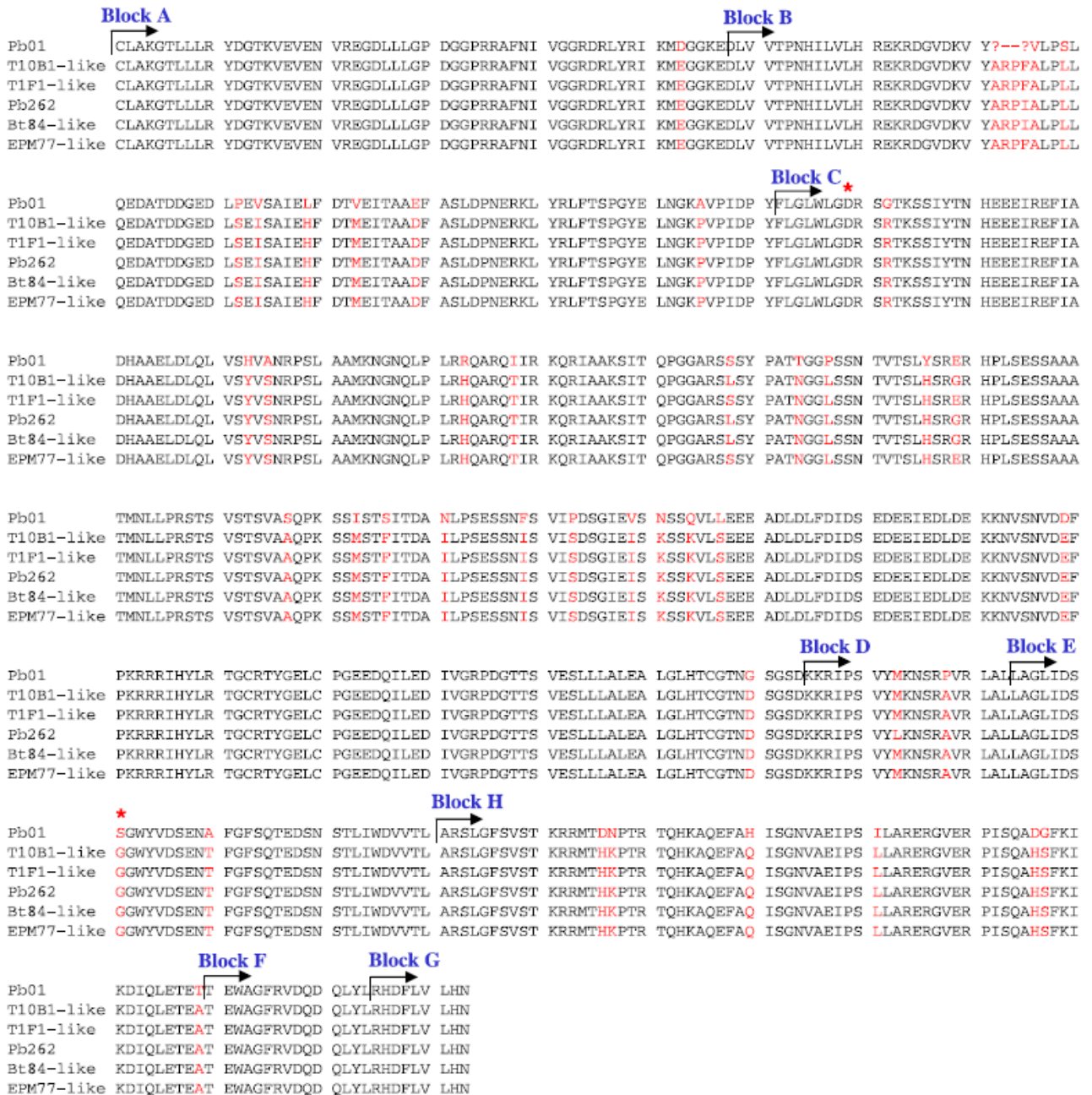


Fig. 1. Deduced amino acid sequences of PRP8 intein from groups of *P. brasiliensis*. The * indicates the position of the two essential aspartic acids according to VMA intein of *S. cerevisiae*. Polymorphic sites are marked in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

length (TL) = 119, consistency index (CI) = 0.9916, homoplasy index (HI) = 0.0084 and retention index (RI) = 0.9286. The Bayesian method resulted in an identical tree with nodes supported by values higher than 80% (Fig. 2). The PRP8 intein sequence of *A. capsulatus* was not included in the phylogenetic analysis because its large divergence from *P. brasiliensis* isolates would misrepresent the phylogenetic relation among the haplotypes of *P. brasiliensis*.

4. Discussion

The recent discovery of possible cryptic species in *P. brasiliensis* increases the importance of correctly recognizing isolates, through practical morphological and/or molecular markers. If the groups

are genetically separated, they are supposed to accumulate some differences, which may result in the occupation of new ecological niches or in different strategies for survival in saprobe or host environment. Thus, the monitoring of the genotypes that are causing PCM disease is extremely important to the comprehension of the different clinical aspects and treatment responses.

The phylogenetic analysis of the sequences of PRP8 intein from *P. brasiliensis* separated, with high bootstrap values, the isolates from the three putative cryptic species, with the exception of the isolate Pb927. According to Matute et al. (2006a), this isolate belongs to the S1 species. However, it has the same haplotype as Bt84, Pb02, and Pbdog that belong to the PS2 species according to its PRP8 intein sequence. Additional molecular analyses are necessary to evaluate the origin and the genotype of this isolate.

Table 4
Nucleotide substitution pattern within the splicing and endonuclease domains of PRP8 intein from *Paracoccidioides brasiliensis* isolates

Domain	Seq 1	Seq 2	S _d	N _d	S	N	p _S	p _N	d _S	d _N	d _S /d _N
Splicing	Pb01	T1F1-like	11.00	9.00	122.67	390.33	0.09	0.02	0.10	0.02	4.08
	Pb01	Bt84-like	12.00	9.00	122.67	390.33	0.10	0.02	0.10	0.02	4.48
	Pb01	EPM77-like	11.00	9.00	122.67	390.33	0.09	0.02	0.10	0.02	4.08
	T1F1-like	Bt84-like	1.00	1.00	125.17	399.83	0.01	0.00	0.01	0.00	3.21
	T1F1-like	EPM77-like	0.00	0.00	125.00	400.00	0.00	0.00	0.00	0.00	
	Bt84-like	EPM77-like	1.00	1.00	125.17	399.83	0.01	0.00	0.01	0.00	3.21
		Mean	6.00	4.83	123.89	395.11	0.05	0.01	0.05	0.01	3.81
	Endonuclease	Pb01	T1F1-like	19.00	30.00	290.50	903.50	0.07	0.03	0.07	0.03
Pb01		Bt84-like	19.00	32.00	291.00	903.00	0.07	0.04	0.07	0.04	1.88
Pb01		EPM77-like	18.00	30.00	290.83	903.17	0.06	0.03	0.06	0.03	1.90
T1F1-like		Bt84-like	4.00	2.00	289.50	904.50	0.01	0.00	0.01	0.00	6.30
T1F1-like		EPM77-like	1.00	0.00	289.33	904.67	0.00	0.00	0.00	-0.00	
Bt84-like		EPM77-like	3.00	2.00	289.83	904.17	0.01	0.00	0.01	0.00	4.71
		Mean	10.67	16.00	290.17	903.83	0.04	0.02	0.04	0.02	3.36

S_d indicates the number of observed synonymous substitutions; N_d is the number of observed non-synonymous substitutions; S denotes the number of potential synonymous substitutions; N is the number of potential non-synonymous substitutions; p_S indicates the proportion of observed synonymous substitutions; p_N is the proportion of observed non-synonymous substitutions; d_S denotes the Jukes-Cantor correction for multiple hits of p_S; d_N is the Jukes-Cantor correction for multiple hits of p_N. The sequences were identified using the isolate names. T1F1-like, Bt84-like and EPM77-like represent the cryptic species S1, PS2, and PS3, respectively.
* The high frequency of synonymous substitutions (greater than 0.74) makes the Jukes-Cantor correction (d_S) unreliable.

The genetic distances of PRP8 intein sequences between the S1 and PS3 groups were very low (similar to the distance found within PS2) when compared to the distances between S1 and PS2 and between the PS3 and PS2 groups. This result corroborates the previous multi-locus phylogenetic analysis (Matute et al., 2006a), in which PS2 was the most strongly supported clade (as it was for the PRP8 intein, with bootstrap values above 95%) and the average pairwise divergence between S1 and PS2 was more than 3 times larger than the average pairwise divergence of isolates within the groups.

Concerning the isolates with no previous species recognition, the isolate Pbdog was grouped with the isolates from the PS2 species (T10B1-like and Bt84-like). The PRP8 intein sequences of isolate Pb18 from GenBank and from the Broad Institute Genome database were not identical to the Pb18 sequence obtained in the present work. As already mentioned all sequencing reactions were carried out in triplicate to generate a reliable consensus sequence for each isolate, and only the Pb18 sequence obtained herein was used in the phylogenetic analysis. Pb18 is the most studied isolate and has been maintained in several laboratories and research institutes for at least 70 years. It is important to clarify the reason for such differences among the Pb18 sequences available, specifically, whether they were caused by simple strain exchanges or by mutations resulting in different Pb18 lineages.

The isolate Pb01 unexpectedly presented many substitutions along the intein sequence (splicing and endonuclease domains) and a deletion of three amino acids in block B of the splicing domain. This isolate, which has been the subject of genome and transcriptome projects (Felipe et al., 2005), also differs substantially from the others in relation to the *hsp70* gene that encodes a conservative heat shock protein (Teixeira et al., 2005; Theodoro et al., 2008). Four large insertions: one of 4 nucleotides, one of 16, a third one of 23 and another of 21 (the last one corresponds to a CT microsatellite) were observed in the *hsp70* gene of Pb01 when it was compared to *P. brasiliensis* isolates from the three cryptic species (Teixeira et al., 2005; Theodoro et al., 2008). Considering that from the 1719 sequenced nucleotides of the PRP8 intein, the isolate Pb01 presented 70, 73, and 69 polymorphic sites compared to S1, PS2, and PS3, respectively, while only 1, 7, and 8 polymorphic sites were observed between PS3 vs. S1, PS3 vs. PS2, and S1 vs. PS2, respectively, it is plausible to deduce that the clade Pb01 has diverged considerably from the other three cryptic species. This result



Fig. 2. Majority rule consensus maximum likelihood tree obtained from the analysis of 500 bootstrap replicates. Parsimony and likelihood bootstrap proportions and Bayesian posterior probabilities (as %) are shown next to the branches, in the order MP/ML/BP. When the bootstrap value was <50%, its percentage was replaced with a “-”.

corroborated other previous phylogenetic analyses with several nuclear encoding regions, such as *GP43*, *CHA*, *Actin*, *ODC*, *URA3*, *CHS2*, *FKS1*, *HSP70*, *Hydrophobin*, *Kex*, *Catalase A*, *Catalase P*, *Formamidase*, and *Glyoxalase* (Carrero et al., 2008) that also showed a significant genetic distance between Pb01 and the remaining genetic groups (S1, PS2, and PS3). The clade containing Pb01-like isolates was well supported and constant with several nuclear coding genes, suggesting the existence of one more cryptic species in *P. brasiliensis* (Teixeira et al., unpublished; Carrero et al., 2008). The genotype Pb01 has been isolated from the central region of Brazil and it is curious that there is no clear geographic barrier that could explain this divergence. Could some specific biotic or abiotic factors be acting as natural barriers in such a manner to explain this high level of genetic divergence, resulting in a sympatric speciation? Morphological assays, such as conidia production in different media and mycelia-yeast transition, which are ongoing in our laboratory, could also clarify some important features about the exploration of the two ecological niches of this pathogen, the host and environmental ones. Furthermore, mycological characteristics, such as virulence and antifungal resistance, are also necessary in order to address whether the different genetic groups present different clinical aspects.

The PRP8 intein from *P. brasiliensis* isolates was compared to the same intein from *A. capsulatus* and they presented a genetic distance of approximately 40%. This intein occurs in other members of the family Ajellomycetaceae and it has already been found in *Blastomyces dermatitidis*, *Emmonsia crescens*, and *Emmonsia parva*

(Rae et al., unpublished, cited in Poulter et al., 2007). We have also evaluated its presence in *P. brasiliensis*, *H. capsulatum*, *B. dermatitidis* and *E. parva* using degenerated primers of PRP8 gene, and we could amplify a fragment of about 1800 bp for all the samples.

All the evaluated isolates presented a full-length intein in the gene *prp8*. The HEG domain of PRP8 intein from *P. brasiliensis* appears to be inactive because the second aspartic acid residue, which is indispensable to its functionality, was replaced by a serine in Pb01 and by a glycine in the rest of the isolates. On the other hand, the ratios of synonymous and non-synonymous substitution (dS/dN) produce values greater than 1, indicating that this region is under conservative selection. The dS/dN values for the splicing domain were slightly higher than the values from the endonuclease domain, corroborating the idea that selection acts more intensively upon the splicing properties. Experimental tests are still necessary to conclude exactly whether the HEG of PRP8 intein of *P. brasiliensis* is, in fact, active.

As mentioned previously, the presence of HEG makes the intein a mobile parasitic genetic element that is potentially able to disperse horizontally as well as vertically, resulting in a large phylogenetic distribution (Perler, 2000). There are more than 100 known inteins and they are distributed among approximately 30 host proteins with distinct functions, some of which occur in different and phylogenetically distant organisms (Liu, 2000). This can explain the presence of PRP8 intein in distant pathogenic fungi, such as *Cryptococcus* species (Basidiomycetes) and different orders of Euascomycetes, Poulter et al. (2007) suggested that the horizontal transference of PRP8 intein between some Ascomycetes and Basidiomycetes could have occurred when cells from the different pathogens were phagocytized by saprophytic microorganisms from the soil (in a saprobic environment), or during a co-infection with two pathogens (in an animal–host environment). The vertical transference occurs during sexual reproduction when the alleles HEG⁺ and HEG⁻ come into contact, thus contributing to the homing process. This was already documented experimentally by the reintroduction of a *vma* allele without the intein in *S. cerevisiae*. After the meiosis all *vma* alleles had the intein (Gimble and Thorner, 1992).

The HEG domain appears to rise and fall cyclically in populations with sexual reproduction. Once the intein is fixated in the population, there is no selection for endonuclease functionality, such that this domain can degenerate by mutations, deletions and genetic drift. These empty sites will be reoccupied again according to the cyclic model (Burt and Koufopanou, 2004). However, recent analysis of HEG inteins suggested that the homing endonuclease domain can persist over long evolutionary times due to several reasons including more recombination, complex population structures preventing the fixation of HEG, a balance between the spreading of HEG and its possible decrease in the host's fitness, or a function of HEG that increases the fitness of its host (Gorgaten and Hilario, 2006).

Although the degree of polymorphism in PRP8 intein from *P. brasiliensis* was not as high as in the commonly used nuclear-coding already mentioned, such as GP43, this sequence clearly contains sufficient phylogenetic signal to separate the genetic groups of *P. brasiliensis*, including the recently discovered species Pb01 (Carrero et al., 2008), indicating that more important than the number of polymorphic sites is their constancy in different groups. Therefore, the PRP8 intein can be considered a reliable additional molecular marker for species recognition in *P. brasiliensis*. Moreover, the knowledge of the diversity of PRP8 inteins in pathogenic fungi can open up new strategies in the treatment of the disease, since these sequences can be used as therapeutic targets (Liu and Yang, 2004). As already observed, the inteins are present in vital proteins and, if their excision is inhibited, the pathogen will not be able to survive or reproduce.

Acknowledgments

We thank Dr. Zoilo Pires de Camargo, Rosely Maria Zancopé Oliveira and Adriana Vicentini Moreira for supplying some of the isolates or DNA samples used in this study. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp Grant Nos. 06/03597-4 and 07/01306-5).

References

- Bagagli, E., Franco, M., Bosco, S.M.G., Hebel-Barbosa, F., Trinca, L.A., Montenegro, M.R., 2003. High frequency of *Paracoccidioides brasiliensis* infection in armadillo (*Dasypus novemcinctus*): an ecological study. *Med. Mycol.* 41, 217–223.
- 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.L., Goodwin, T.J.D., Poulter, R.T.M., 2001. A nuclear encoded intein in the fungal pathogen *Cryptococcus neoformans*. *Yeast* 18, 1365–1370.
- Butler, M.L., Gray, J., Goodwin, T.J.D., Poulter, R.T.M., 2006. The distribution and evolutionary history of the PRP8 intein. *BMC Evol. Biol.* 6, 1–26.
- Butler, M.L., Poulter, R.T.M., 2005. The PRP8 inteins in *Cryptococcus* are a source of phylogenetic and epidemiological information. *Fungal Genet. Biol.* 42, 452–463.
- Carrero, L.L., Niño-Vega, G., Teixeira, M.M., Carvalho, M.J.A., Soares, C.M.A., Pereira, M., Jesuino, R.S.A., McEwen, J.G., Mendoza, L., Taylor, J.W., Felipe, M.S., San-Blas, G., 2008. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. *Fungal Genet. Biol.* 45 (5), 605–612.
- 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.
- Farias, M.R., Werner, J., Muro, M.D., Marques, S.A., Marques, M.E.A., Franco, M.F., et al., 2005. Canine *Paracoccidioidomycosis*: case report of generalized lymphadenitis. *Rev. Inst. Med. Trop. Sao Paulo* 47, 64.
- Felipe, M.S., Andrade, R.V., Arraes, F.B., Nicola, A.M., Maranhão, A.Q., Torres, F.A., et al., 2005. Transcriptional profiles of the human pathogenic fungus *Paracoccidioides brasiliensis* in mycelium and yeast cells. *J. Biol. Chem.* 280 (26), 24706–24714.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Ferreira, M.S., Freitas, L.H., Lacaz, C.da.S., del Negro, G.M., de Melo, N.T., Garcia, N.M., et al., 1990. Isolation and characterization of a *Paracoccidioides brasiliensis* strain from a dog food probably contaminated with soil in Uberlândia, Brazil. *J. Med. Vet. Mycol.* 28 (3), 253–256.
- Franco, M., 1987. Host-parasite relationships in paracoccidioidomycosis. *J. Med. Vet. Mycol.* 25, 5–18.
- Garcia, N.M., Del Negro, G.M., Heins-Vaccari, E.M., de Melo, N.T., de Assis, C.M., Lacaz, C.da.S., 1993. *Paracoccidioides brasiliensis*, a new sample isolated from feces of a penguin (*Pygoscelis adeliae*). *Rev. Inst. Med. Trop. Sao Paulo* 35 (3), 227–235.
- Gimble, F.S., Thorner, J., 1992. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature* 357, 301–306.
- Gonzales, M., Dugan, J., Shafer, R., 2002. Synonymous–non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). *Bioinformatics* 18, 886–887.
- Gorgaten, J.P., Hilario, E., 2006. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* 6, 1–5.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Guindon, S., Lethiec, F., Duroux, P., Gascuel, O., 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33, Web Server issue 10.1093/nar/gki352.
- Hebel-Barbosa, F., Morais, F.V., Montenegro, M.F., Kuramae, E.E., Taylor, J.W., Montes, B., McEwen, J.G., Puccia, R., Bagagli, E., 2003. Sequence comparison of the internal transcribed spacer regions and gp43 in *paracoccidioides gbrasiliensis* from patients and armadillos *Dasypus novemcinctus*. *J. Clin. Microbiol.* 41, 5735–5737.
- 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.
- Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.
- 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 sel-fish gene in yeast: molecular evolution versus site-directed mutagenesis. *Mol. Biol. Evol.* 22, 1535–1538.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5, 150–163.
- Lacaz, C.S., Porto, E., Martins, J.E.C., 1984. *Paracoccidioidomycose*. In: Lacaz, C.S., Porto, E., Martins, J.E.C. (Eds.), *Micologia Médica*. Sarvie, Sao Paulo, pp. 189–216.

- Liu, X.Q., 2000. Protein-splicing intein: genetic mobility, origin, and evolution. *Annu. Rev. Genet.* 34, 61–76.
- Liu, X.Q., Yang, J., 2004. Prp8 intein in fungal pathogens: target for potential antifungal drugs. *FEBS Lett.* 572, 46–50.
- Macoris, S.A.G., Sugizaki, M.F., Peraçoli, M.T.S., Bosco, S.M.G., Hebeler-Barbosa, F., Simões, L.B., et al., 2006. Virulence attenuation and phenotypic variation of *Paracoccidioides brasiliensis* isolates obtained from armadillos and patients. *Mem. Inst. Oswaldo Cruz* 10, 331–334.
- Matute, D.R., McEween, J.G., Montes, B.A., San-Blas, G., Bagagli, E., Rauscher, J.T., et al., 2006a. Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Mol. Biol. Evol.* 23, 65–73.
- Matute, D.R., Sepulveda, E.V., Quesada, L.M., Goldman, G.H., Taylor, J.W., Restrepo, A., et al., 2006b. Microsatellite Analysis of three phylogenetic species of *Paracoccidioides brasiliensis*. *J. Clin. Microbiol.* 44, 2153–2157.
- Matute, D.R., Torres, I.P., Salgado-Salazar, C., Restrepo, A., McEwen, J.G., 2007. Background selection at the chitin synthase II (chs2) locus in *Paracoccidioides brasiliensis* species complex. *Fungal Genet. Biol.* 44 (5), 357–367.
- McCullough, M.J., Disalvo, A.F., Clemons, K.V., Park, P., Stevens, D.A., 2000. Molecular epidemiology of *Blastomyces dermatitidis*. *Clin. Infect. Dis.* 30, 328–335.
- Mendes, R.P., 1994. The gamut of clinical manifestations. In: Franco, M., Lacaz, C.S., Restrepo-Moreno, A., Del Negro, G. (Eds.), *Paracoccidioidomycosis*. CRC Press, Boca Raton, pp. 233–257.
- Morais, F.V., Barros, T.F., Fukada, M.K., Cisalpino, P.S., Puccia, R., 2000. Polymorphism in the gene coding for the immunodominant antigen gp43 from the pathogenic fungus *Paracoccidioides brasiliensis*. *J. Clin. Microbiol.* 38 (11), 3960–3966.
- Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Okuda, Y., Sasaki, D., Nogami, S., Kaneko, Y., Ohya, Y., Anraku, Y., 2003. Occurrence, horizontal transfer and degeneration of VDE intein family in Saccharomycete yeasts. *Yeast* 20, 563–573.
- Perler, F.B., 2000. Inbase, the Intein database. *Nucleic Acids Res.* 22, 344–345.
- Petrokovski, S., 2001. Intein spread and extinction in evolution. *Trends Genet.* 17, 465–472.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Posey, K., Koufopanou, V., Burt, A., Gimble, F., 2004. Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. *Nucleic Acids Res.* 32, 3947–3956.
- Poulter, R.T.M., Goodwin, T.J.D., Butler, M.L., 2007. The nuclear encoded inteins of fungi. *Fungal Genet. Biol.* 44, 153–179.
- Restrepo-Moreno, A., Schneidau Jr., D., 1967. Nature of the skin-reactive principle in culture filtrates prepared from *Paracoccidioides brasiliensis*. *Bacteriology* 93 (6), 1741–1748.
- Ricci, G., Mota, F.T., Wakamatsu, A., Serafim, R.C., Borra, R.C., Franco, M., 2004. Canine *Paracoccidioidomycosis*. *Med. Mycol.* 42, 379–383.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Swofford, D.L., 2002. PAUP-Phylogenetic Analysis Using Parsimony (and other Methods). Version 4b10. Sinauer, Sunderland.
- Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S., et al., 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.* 31 (1), 21–32.
- Teixeira, M.M., De Carvalho, M.J.A., Dantas, A.S., Felipe, M.S.S., 2005. The hsp70 gene presents sequence differences among isolates of *Paracoccidioides brasiliensis*. *Rev. Inst. Med. Trop. Sao Paulo* 47 (14), 48.
- Theodoro, R.C., Bosco, S.M.G., Araújo, J.P.Jr., Candeias, J.M., Macoris S.A.G., Trinca, L.A., Bagagli, E., 2008. Dimorphism, thermal tolerance, virulence and heat shock protein 70 transcription in different isolates of *Paracoccidioides brasiliensis*. On line version, (10.1007/s11046-008-9091-6).
- Xia, X., Xie, Z., 2001. DAMBE: data analysis in molecular biology and evolution. *J. Hered.* 92, 371–373.
- Xu, M.Q., 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.

Artigo 2

Inteins in pathogenic fungi: a phylogenetic tool and perspectives for therapeutic applications

Raquel Cordeiro Theodoro, Eduardo Bagagli⁺

Departamento de Microbiologia e Imunologia, Instituto de Biociências, UNESP-Botucatu, Distrito de Rubião Junior s/n, 18618-000 Botucatu, São Paulo, Brasil

Inteins or "internal proteins" are coding sequences that are transcribed and translated with flanking sequences (exteins). After translation, the inteins are excised by an autocatalytic process and the host protein assumes its normal conformation and develops its expected function. These parasitic genetic elements have been found in important, conserved proteins in all three domains of life. Most of the eukaryotic inteins are present in the fungi kingdom and the PRP8 intein is one of the most widespread inteins, occurring in important pathogens such as Cryptococcus neoformans (varieties grubii and neoformans), Cryptococcus gattii, Histoplasma capsulatum and Paracoccidioides brasiliensis. The knowledge of conserved and non-conserved domains in inteins have opened up new opportunities for the study of population variability in pathogenic fungi, including their phylogenetic relationships and recognition or diagnoses of species. Furthermore, inteins in pathogenic fungi should also be considered a promising therapeutic drug target, since once the autocatalytic splicing is inhibited, the host protein, which is typically vital, will not be able to perform its normal function and the fungal cell will not survive or reproduce.

Key words: Intein - PRP8 species - pathogenic fungi - therapeutic target

The modern synthesis of natural selection with the Mendelian inheritance theory explains how natural selection can interfere with gene frequencies in different populations. But how does natural selection operate? This has been the theme of several discussions among important evolutionists. Some believe that it operates over a group or the individual and others assume that natural selection acts at the molecular level, operating over genes. The latter hypothesis has been consistently defended by Richard Dawkins (1989), in his book entitled "The Selfish Gene", where he proposed that all living organisms are machines for gene survival and replication, such that natural selection will obviously favour those genes that increase the fitness of their hosts. Parasitic genetic elements, such as inteins, are excellent examples of how natural selection occurs at the molecular level. Inteins are considered parasitic genes because they are found in coding genes, with which they are transcribed and translated and, apparently, do not contribute to the organism's fitness. The inteins undergo a post-translational excision by an autocatalytic process and the host protein assumes its normal conformation, by a peptide ligation, and can develop its normal function (Fig. 1) (Cooper & Stevens 1995).

The first described intein was found in the vacuolar ATPase gene from *Sacharomyces cerevisiae*. The central region of this gene presented high similarity to some

endonucleases (Hirata et al. 1990). Interestingly, the ATPase protein showed a molecular weight of 70kDa, which was less than expected and corresponded to the weight without the central region. In 1990, Kane et al. demonstrated that the insertion (corresponding to an endonuclease) was present in the mRNA and that all the mRNA was translated. Therefore, the 70kDa protein was a result of post-translational processing.

Inteins are coding sequences whose translated products can range from 134-608 amino acids. They excise themselves from the host protein and join the flanking protein sequences (exteins) with a peptide ligation, thus generating a functional protein (Gogarten et al. 2002). There are two intein types, mini-inteins and full-length inteins, both of which present a splicing domain containing four blocks of more or less well-conserved amino acid sequences. Blocks A and B constitute the N-terminal splicing domain and blocks F and G constitute the C-terminal splicing domain. In addition, full-length inteins present an endonuclease domain that can perform a homing function [homing endonuclease gene (HEG)] with the potential to make the intein a mobile genetic element, resulting in occupation of empty alleles and duplication of the parasitic genetic element (Liu 2000). The transfer of an intein to a cognate and empty allele considerably increases the frequency of HEG⁺ alleles in the population through super-Mendelian inheritance (Burt & Koufopanou 2004). Therefore, in a few generations, the HEG intein may be fixed in the population. There are different endonuclease types that are encoded by the HEG domain, but endonucleases from the LAGLIDADG family are the most frequent (Perler 2000). The HEG is divided into blocks C, D, E and H, which can be found in some introns that also have a "homing" capacity (Belfort et al. 1995).

⁺ Corresponding author: bagagli@ibb.unesp.br

Received 21 July 2008

Accepted 13 March 2009

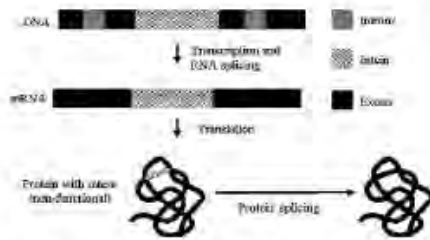


Fig. 1: scheme of intein splicing from the host protein (extein) after its translation.

Inteins have been detected in all three domains of life: eukaryotes, bacteria and archaea, according to the intein database InBase (<http://www.neb.com/neb/inteins.html>). They are usually found at conserved sites of housekeeping proteins that have vital functions in the cell, such as metabolic enzymes, DNA and RNA polymerases, proteases and others (Liu 2000). For this reason, any mutation that disrupts the functionality of the splicing domain can be fatal for the cell. In this manner, natural selection operates at the molecular level: although the inteins do not appear to promote any improvement in the fitness of the organism, their localization at important and thus conserved sites can be the reason for their survival in genomes, since they recognize specific and conserved genetic niches in the genome for the homing process.

Horizontal transfer of the intein coding sequence and a small portion of the flanking sequence readily accounts for the discontinuous intein dispersion in corresponding insertion points, since inteins have an established homing mechanism (Dassa & Pietrokovski 2008). This horizontal transfer, as well as independent loss of inteins from different lineages, contributes to their large and sporadic phylogenetic distribution.

When inteins occupy exactly the same site, they are referred to as “allelic inteins”, even though they are from different species (Butler et al. 2006). Today, more than 130 inteins are known to be distributed among approximately 34 host proteins with distinct functions (Perler 2000). Some inteins are present in distantly related organisms. For example, the DnaB intein is present in a photosynthetic cyanobacterium and the heterotrophic marine thermophile *Rhodothermus marinus* (Liu & Hu 1997). An intein can also parasitise functionally different proteins that display structural similarity, as in the case of the intein TopA, which is present in topoisomerase I and reverse gyrase (Chute et al. 1998). Inteins are inherited vertically when alleles with and without inteins are assembled during sexual reproduction, enabling the homing process. This was documented through the invasion of an intein-containing allele into the *vma* allele without the intein in *S. cerevisiae*. After meiosis, more than 50% of the *vma-1* alleles had the intein (Gimble & Thorner 1992).

The distribution and population dynamics of the VMA intein seem to follow a model of invasion, fixation, degeneration, loss and reinvasion. Once the intein

is established in the population, there is no selection for endonuclease functionality, such that this domain can become degenerate through mutations, deletions and genetic drift. These empty sites will be reoccupied again according to the cyclic model (Burt & Koufopanou 2004) (Fig. 2). However, a recent analysis of HEG inteins suggested that the homing endonuclease domain can persist over long evolutionary times due higher recombination frequency, complex population structures preventing the fixation of HEG, a balance between the spreading of HEG and its possible decrease in relation to the host’s fitness, or a function of HEG that increases the fitness of its host (Gogarten & Hilario 2006).

It is also known that some inteins could have evolved from a parasitic to a functional role (Fig. 2) through positive selection. One example is the endonuclease HO, which is responsible for mating-type switching, between the sexual types α and α , in *S. cerevisiae* (Liu 2000). This endonuclease is very similar to the VMA1 intein from *S. cerevisiae* and, although inactive, it also has conserved regions related to splicing activity. This is a strong indication that HO has evolved from a VMA intein (Gogarten et al. 2002, Haber & Wolfe 2005).

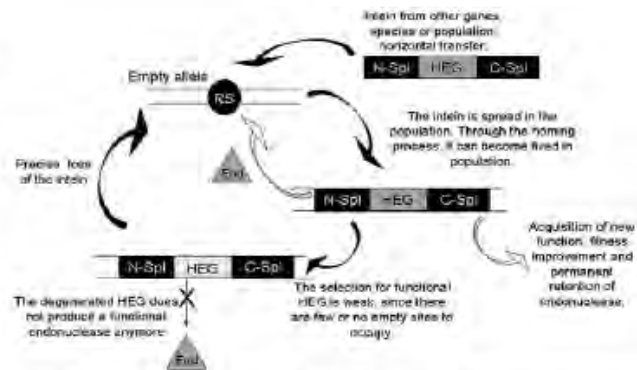


Fig. 2: homing cycle of the VMA intein (modified from Gogarten & Hilario 2006). C-Spl: splicing domain C-terminal; End: endonuclease encoded by the HEG domain; HEG: homing endonuclease gene; N-Spl: splicing domain N-terminal; RS: endonuclease recognition site.

Inteins in fungi - As already mentioned, inteins are present in the three life domains: eukarya, bacteria and archaea. According to Inbase (Perler 2000), about 60 species from the eukarya domain present inteins in some important genes. Among these species are some amoebae and algae, but the majority of the eukaryotes with inteins are fungi (about 50 species), from distinct phyla (Chytridiomycota, Zygomycota, Basidiomycota and Ascomycota). The majority of inteins in fungi are in the host genes Vacuolar ATPase (H⁺transporting ATP synthase) subunit A (VMA intein) and *prp8* (PRP8 intein). The VMA intein is widespread among yeast fungi, such as *Saccharomyces*, *Zygosaccharomyces* and some *Candida* (all Ascomycota) species, while the PRP8

intein is present in all fungi phyla (Perler 2000), being more abundant among Ascomycetes, some species of the orders Eurotiales, Onygenales and Helotiales, a few Basidiomycetes from the order Tremellales and the chytrid fungus, *Batrachochytrium dendrobatidis* (Inbase). Some important human pathogenic species, such as *Histoplasma capsulatum*, *Paracoccidioides brasiliensis* (Ascomycota, order Onygenales), *Cryptococcus neoformans*, and *Cryptococcus gattii* (Basidiomycota, order Tremellales) are known to have the PRP8 intein (Butler et al. 2006).

PRP8 intein

This intein is located in the *prp8* gene, which codes for the PRP8 protein, the largest and most highly conserved nuclear protein (61% similarity between yeast and human), which occupies a central position in the catalytic core of the spliceosome. It is a component of U5 snRNPs and is implicated in several crucial molecular rearrangements during pre-mRNA splicing (for the removal of introns), a central step in gene expression (Grainger & Beggs 2005).

The PRP8 intein was initially described in a large number of *C. neoformans* (varieties *grubii* and *neoformans*) and *C. gattii* isolates as mini-inteins (without the HEG domain) (Butler et al. 2001, Butler & Poulter 2005). Besides these *Cryptococcus* species, the PRP8 intein can also be found as a full-length intein in *Cryptococcus laurentii*. Since this species is moderately distant from *C. neoformans*, Butler & Poulter (2005) suggested that the mini-intein found in *C. neoformans* and *C. gattii* was derived from horizontal transfer, with *C. laurentii* or another species being the donor.

Butler et al. (2006) searched for PRP8 inteins in a database of genomes and discovered it in many species of Ascomycetes. The authors also compared the PRP8 intein with the VMA intein, with relation to their splicing and HEG domains and HEG functionality and, interestingly, noted that the PRP8 intein can be found as a mini-intein or full-length intein, depending on the fungal species, whereas all known VMA inteins are full-length. In contrast to the VMA intein, the PRP8 intein seems to have a functional HEG domain in most of the host species. Analysis of dS/dN (the ratio between synonymous and nonsynonymous substitutions) also revealed that the HEG domain from PRP8 is saturated with synonymous substitutions when compared to the HEG domain from VMA, suggesting that they are constrained by selection and have been diverging for a longer time than the VMA intein. This is also corroborated by the sporadic occurrence of the PRP8 intein, which is missing in some species closely related to fungi that contain the PRP8 intein. This is the case of *Coccidioides immitis* and *Coccidioides posadasii*, two pathogenic "inteinless" species that are closely related to *Uncinocarpus reesii*, which contains a PRP8 mini-intein. The PRP8 intein appears to have an ancient origin in Ascomycota, and the "inteinless" species indicate that the loss of an intein (by deletion or fixation of an empty allele) could be quite frequent.

Besides the *Cryptococcus* species that belonged to the Basidiomycota phylum, inteins in the *prp8* gene were

also described in many Eueascomycetes, such as *Aspergillus nidulans*, *Aspergillus fumigatus*, *P. brasiliensis*, *H. capsulatum* and *Botrytis cinera*. In this group, however, there is no evidence of horizontal transfer, since the phylogenetic analysis of these inteins is similar to those using other nuclear sequences, for example, ribosomal sequences, which suggests an ancient origin and vertical transmission among Eueascomycetes (Butler et al. 2006). Since only a few *Cryptococcus* species (Basidiomycota, order Tremellales) present the PRP8 intein, of which *C. laurentii* is the only one with a full-length intein (Butler & Poulter 2005) and no other basidiomycetes have been shown to have inteins, it seems reasonable to explain this distribution of the PRP8 intein as the result of horizontal transfer between ascomycetes (donors) and basidiomycetes (receptors). This transfer, according to Poulter et al. (2007), could have occurred during the parasitic phase, in co-infection cases, or during the saprobic phase of these pathogens, when different fungal cells are phagocytosed by microorganisms from soil. In fact, the interaction between pathogenic fungi and phagocytic amoebae from soil has been the theme of several studies that proved the similarity between fungus-macrophage and fungus-amoeba interactions. This proposes that virulence features could have emerged in a saprobic environment and today present a dual use (in both saprobic and host environments) (Casadevall et al. 2003, Steenbergen & Casadevall 2003).

Species concept and use of PRP8 intein for phylogenetic studies in pathogenic fungi

The species concept is one of the central themes of evolution, and although a "species" is considered a concrete group, in contrast to artificial ones (such as genera, order, class, phylum and kingdom), its definition and recognition are considered a challenge, even in the genomic and proteomic era.

Under an evolutionary approach, members of the same species are expected to form a genetic and ecological unit, sharing the same alleles and exploring the same ecological niche. This definition seems to be very simple and direct, but how can we recognize it in nature? The morphological recognition of species for example is based on common phenotypes presented in members of the same species. However, it is typical to group genetically different organisms into one species, mainly when we deal with unicellular organisms, whose phenotypes are restricted to chemical interactions with substrates or a few macro and microscopic features. The biological recognition of species takes into account the interbreeding capacity among members from the same species. However, this recognition is not applicable for asexual populations, and even in sexual ones, it may lead to a conflict between potential and real gene flow (Taylor et al. 2006). There is also an ecological recognition that defines a species as a group of individuals that occupy the same niche, a concept that is closely connected to the biological one, as natural selection favours individuals that interbreed only with others with the same ecological adaptation. It is known that the adaptive value of hybrids could be, in many cases, lower than that of the parental species, such that parental species will present better

competitive adaptations for occupying their niches (Ridley 2006). This concept is also very difficult to apply to microorganisms, whose niche is defined by biochemical interactions with substrates (environmental and host) and is compounded by n dimensions (Hutchinson 1959, Bagagli et al. 2008).

Recently, cryptic speciation was revealed in some important fungal pathogens by phylogenetic recognition of species, which detects genetic divergence among populations through multi-locus sequence typing, by concordance of gene genealogies (Taylor et al. 2000). This method also provides information about the reproductive mode of the different divergent species, even without the isolation of sexual structures. Cryptic species have been detected by gene genealogies in important pathogenic fungi, such as *C. immitis* (separated into 2 species: *C. immitis* and *C. posadasii*) (Koufopanou et al. 2001), *H. capsulatum* (separated into at least 7 species) (Kasuga et al. 2003) and *P. brasiliensis*, which is now separated into at least four previously unrecognized species: S1 (species 1 from Brazil, Argentina, Paraguay, Peru and Venezuela), PS2 (phylogenetic species 2 from Brazil and Venezuela), PS3 (phylogenetic species 3 with 21 isolates from Colombia) (Matute et al. 2006) and Pb01-like (from the central region of Brazil) (Carrero et al. 2008, Teixeira 2008).

Knowledge of the variability among fungal isolates that cause important diseases, including systemic mycoses, is very important not only for population studies but also, and mainly, to correlate different clinical manifestations and distinguish among genotypes of isolates to improve diagnoses and treatment. Once the species are genetically separated, they are supposed to accumulate some morphological differences, which may result in the occupation of new ecological niches or different strategies for survival in the saprobe or host environment. However, if the recognition of cryptic species is difficult and laborious (since the use of many nuclear coding sequences is required), the search for morphological, physiological, ecological and clinical markers correlated with the genetic clustering is even more challenging. Usually, after the discovery of cryptic species, some phenotypes that were originally considered merely intra-specific polymorphisms become associated with isolated genetic groups. For example, *C. posadasii* presents slower growth in high salt concentration media than *C. immitis*, however, this feature is not used for diagnosis (Fisher et al. 2002). In some cases, when the host selective pressure on the pathogen population is sufficient to be considered an allopatric speciation factor, the host can be considered a marker for species recognition. For example, the genus *Pneumocystis*, characterized by non-cultivable and highly diversified fungal pathogens that dwell in the lungs of mammals, comprises numerous host-specific species, such as *Pneumocystis jirovecii* in humans, *Pneumocystis carinii* and *Pneumocystis wakefieldiae* in rats, *Pneumocystis murina* in mice, and *Pneumocystis oryctolagi* sp. nov. in old world rabbits (Dei-Cas et al. 2006).

The search for morphological and/or molecular markers that efficiently differentiate the cryptic species in pathogenic human fungi is essential for practical medical purposes. But what constitutes a good marker? Most morphological characteristics are extremely variable because they usually result from the expression of innumerable genes. A good morphological marker should clearly distinguish between various species and also be constant and repeatable among members of the same species. The same is required for molecular markers but, in this case, polymorphisms can be objectively distinguished as it is based on nucleotide substitutions. Generally, coding nuclear genes with sufficient polymorphism(s) to separate the species or genetic groups are the most useful markers.

Inteins also seem to constitute a promising source of phylogenetic information, especially the full-length inteins, because they are expected to have more sequence variation in the endonuclease domain than in the splicing domain due to a more relaxed selection. This is particularly true if the HEG is no longer active (Gogarten & Hilario 2006). In the case of PRP8 inteins, even though their HEG domain appears to be constrained by selection (most PRP8 inteins may have an active endonuclease), they are very diverse, probably due to an ancient origin and a long evolutionary history (Butler et al. 2006) and are useful sequences for phylogenetic inferences in intein-carrying fungal species.

Since the *prp8* gene is highly conserved among eukaryotes, primer design for its amplification by PCR is possible for phylogenetic studies of fungal species and varieties. Using this approach, Butler and Poulter (2005) distinguished between strains of two varieties of *C. neoformans* (*neoformans* and *grubii*) and from strains of *C. gattii* (recently designated as a new species) by the neighbour joining method. These encapsulated yeasts are capable of causing serious infections in both immunocompromised and immunocompetent persons. The varieties *grubii* (serotype A), *neoformans* (serotype D), and *gattii* (serotypes B and C) were initially designated according to antigenic differences in the capsular polysaccharide. This was also confirmed by molecular data analyses (Franzot et al. 1999, Meyer et al. 1999, Ellis et al. 2000). The *gattii* variety has recently been accorded species status as *C. gattii* (Boekhout et al. 2001, Kwon-Chung et al. 2002). Although the use of the PRP8 intein for phylogenetic analysis at a higher taxonomic level is not viable due to its erratic distribution, it proved to be a suitable taxonomic marker at the species or sub-species level for epidemiological studies of *Cryptococcus*.

We applied the same approach to sequence and phylogenetically analyze the PRP8 intein in different isolates of the *P. brasiliensis* species complex (S1, PS2, PS3 and Pb01-like species) (Theodoro et al. 2008). This thermomorphic fungus is the etiologic agent of paracoccidioidomycosis, the most important systemic mycosis in Latin America (Franco 1987). It has been frequently recovered from clinical samples and tissues of the nine-banded armadillo, *Dasypus novemcinctus* (Bagagli et al. 2003). It was also recently detected in and isolated from dogs (Ricci et al. 2004, Farias et al. 2005). However, mo-

lecular polymorphism analysis suggested that the same "ecopathogenotypes" can infect both humans and other animals (Hebeler-Barbosa et al. 2003).

In our assay, the PRP8 intein from 22 isolates of *P. brasiliensis* belonging to the four previously recognized species was sequenced. Phylogenetic analysis by maximum-parsimony, maximum likelihood and Bayesian analysis clearly separated the isolates from the four species and revealed a significant difference between Pb01-like and the remaining species (Theodoro et al. 2008). This corroborated previous phylogenetic analyses of several nuclear encoding regions, such as *GP43* (the gene coding for the immunodominant antigen of *P. brasiliensis*), *CH4*, *actin*, *ODC*, *URA3*, *CHS2*, *FKS1*, *HSP70*, *Hydrophobin*, *Kex*, *Catalase A*, *Catalase P*, *Formamidase* and *Glyoxalase* (Carrero et al. 2008). The Pb01 genotype has been isolated from the central region of Brazil and, interestingly, there is no clear geographic barrier that could explain this divergence, since this species is practically sympatric with respect to the S1 and PS2 groups. Morphological assays could clarify some important features about the exploration of the host and environmental ecological niches of this pathogen. Additionally, other mycological characteristics, such as virulence and antifungal resistance, are also necessary to determine to address whether the different genetic groups present different clinical aspects.

All the evaluated *P. brasiliensis* isolates presented a full-length intein in the *prp8* gene. The HEG domain of the PRP8 intein from *P. brasiliensis* appears to be inactive because the second aspartic acid residue, which is indispensable for functionality, was replaced by a serine in the Pb01 isolate (from the Pb01-like genetic group) and a glycine in the rest of the isolates. However, the ratio between synonymous and nonsynonymous substitutions (dS/dN) produced values greater than 1, indicating that this region is under conservative selection and probably was active until recently. Although the degree of polymorphisms in the PRP8 intein from *P. brasiliensis* was not as high as in the commonly used nuclear-coding genes, this sequence contains sufficient phylogenetic information to separate the cryptic species of *P. brasiliensis* and therefore constitutes a reliable molecular marker for this pathogen.

PRP8 as a therapeutic target

The discovery of inteins has offered new biotechnological tools for protein purification on account of the protein splicing nature of these parasitic genetic elements. The traditional methods for protein purification, which require histidine tags used in affinity chromatography, have low yield and can compromise the stability, solubility and activity of the desired protein (Gogarten et al. 2002). By mutating the N or C terminus of an intein, a desired gene can be ligated and expressed in-frame with the intein and excised with it. The protein can be purified by releasing it from the mutant intein through a change in reducing conditions (Chong et al. 1997, 1998). Inteins have also been used in protein splicing, autoproteolysis and biosynthesis of reconstructed polypeptides with reactive ends necessary for intein-mediated protein ligation (Evans et al. 1998).

Besides these biotechnological applications, the distribution, location and functionality of inteins make them promising therapeutic targets against important pathogens. First, intein splicing is required for the production of mature and functional protein (extein). Since inteins are commonly inserted into vital proteins, the inhibition of its splicing would lead to microorganism death. Second, since no intein has been found in animals or humans, this therapeutic target seems to be very safe for the pathogens hosts, and third, screening systems for the detection of mutations or inhibitors of the splicing domain can be tested in *Escherichia coli* (Derbyshire et al. 1997).

Some inteins occur not in vital proteins, but in some proteins required in specific conditions. For example, the VMA intein is in the vacuolar H⁺-ATPase, which is essential for growth in the presence of CaCl₂. This can be useful for the selection of splicing inhibitors that inhibit yeast growth in medium with CaCl₂ (Nogami et al. 1997). Selection of systems for the inhibition of protein splicing can be done artificially by inserting inteins into proteins that provide some selective advantage.

One of the first systems used to examine the intein splicing function in a non-native host context was based on the β-galactosidase complementation assay for the RecA intein of *Mycobacterium tuberculosis*. RecA splicing yielded an active LacZα fragment, which could be easily detected by the blue/white assay (Davis et al. 1992). Using this approach, Daugelat and Jacobs (1999) inserted a RecA intein from *M. tuberculosis* with a C-terminal cysteine residue next to the enzyme responsible for kanamycin resistance. This same intein was also used in an in vitro screening system for protein splicing inhibitors based on green fluorescent protein (GFP), as an indicator when expressed in *E. coli* (Gangopadhyay et al. 2003).

The first potential genetic selection for intein splicing in yeast was the transfer of the Sce intein (*S. cerevisiae* VMA1 intein) into the *vat2* gene, which encodes the 60kDa subunit of the vacuolar H⁺-ATPase in yeast (Cooper et al. 1993). Knockout strains for this gene grow at pH 5.0, but not at pH 7.5 (Yamashiro et al. 1990). Therefore, growth at pH 7.5 using an intein-Vat2 fusion would select for active splicing.

Similar studies on PRP8 fungal inteins were carried out to evaluate the splicing activity of this intein in a non-native host protein in *E. coli*. The PRP8 intein from *C. neoformans* has been studied by mutagenesis to identify the minimal sequence requirements, as well as the importance of a number of conserved residues for its splicing in a foreign protein context (Pearl et al. 2007). The same mutagenic approach was used for the PRP8 mini-intein of *Penicillium chrysogenum*, the smallest known nuclear-encoded active splicing protein, by introducing deletions at different sites. The authors removed a total of 14 amino acid residues (8 and 6 at 2 different sites) and the resulting 143aa intein was functional when expressed in *E. coli* (Elleuche et al. 2008). Elleuche et al. (2006) also expressed the PRP8 mini-intein of *P. chrysogenum*, *Penicillium expansum* and *Penicillium*

vulpinum in *E. coli* in a model host protein and a GFP model system. This same intein, from *P. chrysogenum*, was also tested in a *trans*-splicing assay by splitting it artificially. Protein *trans*-splicing involves the joining of two halves of an intein sequence through their high affinity for each other followed by ligation of the two extein sequences. Therefore, these split inteins self-associate and catalyze protein-splicing activity in *trans* (Tavassoli et al. 2008). Elleuche & Pöggeler (2007) tested three split sites in the *P. chrysogenum* PRP8 intein. Two of these sites resulted in protein *trans*-splicing: one that corresponded to the insertion site of the endonuclease domain in allelic large PRP8 inteins and the other was detected at a new position, located N-terminal of the endonuclease insertion site.

The splicing of PRP8 inteins from *A. fumigatus*, *A. nidulans*, *H. capsulatum* and *C. neoformans* was also evaluated by inserting them into the expression plasmid pMST between the *Xho*I and *Age*I sites, replacing the *Ssp* DnaB intein coding sequence of the plasmid. This insertion site was flanked by a maltose binding protein at its N-terminus and a thioredoxin at its C-terminus. Western blots were carried out using anti-thioredoxin antibody. The precursor protein, spliced protein and excised intein were identified by their predicted sizes. Intein splicing was evaluated at 25°C and 36°C for the four fungal species and in general splicing efficiency was higher at 25°C. This relationship was more significant for *H. capsulatum*, whose splicing efficiency at 37°C was only 25%, while at 25°C it was 100% (no band for the precursor protein was observed by western blot) (Liu & Yang 2004).

It is interesting to note that, as a dimorphic fungus, *H. capsulatum* must convert from the mycelia (M) to yeast (Y) form during infection or under high temperature (37°C). If the splicing efficiency of the intein in the *prp8* gene at 37°C is as low as in the non-native gene host, the M-Y transition could be disturbed because many different genes must be transcribed and correctly translated for the fungus to assume the yeast morphology and physiology. In this case, the PRP8 intein could decrease the fitness of the pathogen and, as the PRP8 intein from *H. capsulatum* is a full-length intein that codes for a functional HEG (Poulter et al. 2007), one may ask why this HEG appears to be maintained by selection. According Gogarten and Hilario (2006), the balance between decreased fitness of the host and active spread of the parasitic element could lead to long persistence time of the HEG, especially when the selective disadvantage to the carrier and the mobility of the intein were not the same in different subpopulations. Since *H. capsulatum* is a cosmopolite pathogen, with many subpopulations, including cryptic species, this hypothesis for the maintenance of the PRP8 intein in this pathogen seems very reasonable.

The ability of the PRP8 intein to self-splice from a non-native protein from *E. coli* was merely the first step for future studies and the identification of splicing inhibitors may constitute potential antifungal drugs. Since PRP8 inteins from different pathogens have the same corresponding insertion site in the PRP8 protein and, differently from the HEG domain, their splicing domain

is highly conserved (Liu & Yang 2004), splicing inhibitor drugs may have the advantage of simultaneous action against most fungal diseases.

REFERENCES

- Bagagli E, Franco M, Bosco SMG, Hebelbarbosa F, Trinca LA, Montenegro MR 2003. High frequency of *Paracoccidioides brasiliensis* infection in armadillo (*Dasypus novemcinctus*): an ecological study. *Med Mycol* 41: 217-223.
- Bagagli E, Theodoro RC, Bosco SMG, McEwen J 2008. *Paracoccidioides brasiliensis*: phylogenetic and ecological aspects. *Mycopathologia* 165: 197-207.
- Belfort M, Reaban ME, Coetzee T, Dalgaard JZ 1995. Prokaryotic introns and inteins: a panoply of form and function. *J Bacteriol* 177: 3897-3903.
- Boekhout T, Theelen B, Diaz M, Fell JW, Hop WC, Abeln EC, Dromer F, Meyer W 2001. Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. *Microbiol* 147: 891-907.
- Burt A, Koufopanou V 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr Opin Genet Dev* 14: 609-615.
- Butler ML, Goodwin TJD, Poulter RTM 2001. A nuclear-encoded intein in the fungal pathogen *Cryptococcus neoformans*. *Yeast* 18: 1365-1370.
- Butler ML, Gray J, Goodwin TJD, Poulter RTM 2006. The distribution and evolutionary history of the PRP8 intein. *BMC Evol Biol* 6: 1-26.
- Butler ML, Poulter RTM 2005. The PRP8 inteins in *Cryptococcus* are a source of phylogenetic and epidemiological information. *Fungal Genet Biol* 42: 452-463.
- Carrero LL, Niño-Vega G, Teixeira MM, Carvalho MJA, Soares CMA, Pereira M, Jesuino RSA, McEwen JG, Mendoza L, Taylor J W, Felipe MS, San-Blas G 2008. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. *Fungal Genet Biol* 45: 605-612.
- Casadevall A, Steenbergen JN, Nosanchuk JD 2003. "Ready-made" virulence and "dual use" virulence factors in pathogenic environmental fungi: the *Cryptococcus neoformans* paradigm. *Curr Opin Microbiol* 6: 332-337.
- Chong S, Mersha FB, Comb DG, Scott ME, Landry D 1998. Single-column purification of free recombinant proteins using a self-cleavable affinity target derived from a protein splicing element. *Gene* 192: 271-281.
- Chong S, Montello GE, Zhang A, Cantor EJ, Liao W 1997. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nuc Acids Res* 26: 5109-5115.
- Chute IC, Hu Z, Liu XQ 1998. A topA intein in *Pyrococcus furiosus* and its relatedness to the r-gyr intein of *Methanococcus jannaschii*. *Gene* 210: 85-92.
- Cooper AA, Chen YJ, Lindorfer MA, Stevens TH 1993. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J* 12: 2575-2583.
- Cooper AA, Stevens TH 1995. Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends Biochem Sci* 20: 351-356.
- Dassa B, Pietrokovski S 2008. Origin and evolution of inteins and other Hint domains. In M Belfort, V Derbyshire, BL Stoddard, DW Wood, *Homing endonucleases and inteins*, Springer-Verlag Berlin, Heidelberg, p. 211-229.

- Daugelat S, Jacobs WR 1999. The *Mycobacterium tuberculosis* recA intein can be used in an ORFTRAP to select for open reading frames. *Prot Sci* 8: 644-653.
- Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedwick SG 1992. Protein splicing in the maturation of *M. tuberculosis* recA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* 71: 201-210.
- Dawkins R 1989. *The selfish gene*, Oxford Univ. Press, Oxford, 540 pp.
- Dei-Cas E, Chabe M, Moukhliis R, Durand-Jolly I, Aliouat el M, Stringer JR, Cushion M, Noel C, de Hoog GS, Guillot J, Viscogliosi E 2006. *Pneumocystis oryctolagi* sp. nov., an uncultured fungus causing pneumonia in rabbits at weaning: review of current knowledge and description of a new taxon on genotypic, phylogenetic and phenotypic bases. *FEMS Microbiol Rev* 30: 853-871.
- Derbyshire V, Wood DW, Wu W, Dansereau JT, Dagaard JZ, Belfort M 1997. 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.
- Elleuche S, Döring K, Pöggeler S 2008. Minimization of a eukaryotic mini-intein. *Biochem Biophys Res Commun* 366: 239-243.
- Elleuche S, Nolting N, Pöggeler S 2006. Protein splicing of PRP8 mini-inteins from species of the genus *Penicillium*. *Appl Microbiol Biotechnol* 72: 959-967.
- Elleuche S, Pöggeler S 2007. Trans-splicing of an artificially split fungal mini-intein. *Biochem Biophys Res Commun* 355: 830-834.
- Ellis D, Marriott D, Hajjeh RA, Warnock D, Meyer W, Barton R 2000. Epidemiology, surveillance of fungal infections. *Med Mycol* 38: 173-182.
- Evans TC, Benner J, Xu MQ 1998. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 7: 2256-2264.
- Farias MR, Werner J, Muro MD, Marques SA, Marques MEA, Franco MF, Ribeiro MG, Custodio CC, Condas LAZ, Bosco SMG, Bagagli E 2005. Canine paracoccidioidomycosis: case report of generalized lymphadenitis. *Rev Inst Med Trop São Paulo* 47: 64.
- Fisher MC, Koenig GL, White TJ, Taylor JW 2002. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. *Mycologia* 94: 73-84.
- Franco M 1987. Host-parasite relationships in paracoccidioidomycosis. *J Med Vet Mycol* 25: 5-18.
- Franzot SP, Salkin IF, Casadevall A 1999. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. *J Clin Microbiol* 37: 833-840.
- Gangopadhyay JP, Jiang S, Paulus H 2003. An *in vitro* screening system for protein splicing inhibitors based on green fluorescent protein as an indicator. *Anal Chem* 75: 2456-2462.
- Gimble FS, Thorner J 1992. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature* 357: 301-306.
- Gogarten JP, Hilario E 2006. Inteins, introns and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol* 6: 1-5.
- Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E 2002. Inteins: structure, function and evolution. *Ann Rev Microbiol* 56: 263-287.
- Grainger RJ, Beggs JD 2005. Prp8 protein: at the heart of the spliceosome. *RNA* 11: 533-557.
- Haber JE, Wolfe KH 2005. Function and evolution of HO and VDE endonucleases in fungi. *Nucleic Acids Mol Biol* 16: 161-175.
- Hebeler-Barbosa F, Morais FV, Montenegro MR, Kuramae EE, Montes B, McEwen JG, Puccia R, Bagagli E 2003. Sequence comparison of the internal transcribed spacer regions and gp 43 in *Paracoccidioides brasiliensis* for patients and armadillos *Dasyurus novemcinctus*. *J Clin Microbiol* 41: 5735-5737.
- Hirata R, Ohsumi 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.
- Hutchinson GE, MacArthur RH 1959. A theoretical ecological model of size distribution among species of animals. *Am Nat* 93: 117-125.
- Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH 1990. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 250: 651-657.
- Kasuga T, White TJ, Koenig G, McEwen J, Restrepo A, Castañeda E, Lacaz CS, Heins-Vaccari, EM, Freitas RS, Zancopé-Oliveira RM, Qin Z, Negroni R, Carter DA, Mikami Y, Tamura M, Taylor ML, Miller GF, Poonwan N, Taylor JW 2003. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol Ecol* 12: 3383-3401.
- Koufopanou V, Burt A, Szaro T, Taylor JW 2001. Gene genealogies, cryptic species and molecular evolution in the human pathogen *Coccidioides immitis* and relatives (Ascomycota, Onygenales). *Mol Biol Evol* 18: 1246-1258.
- Kwon-Chung KJ, Boekhout T, Fell JW, Diaz M 2002. Proposal to conserve the name *Cryptococcus gattii* against *C. Hondurianus* and *C. Bacillisporus* (Basidiomycota, Hymenomycetes, Tremellomycetiadae). *Taxon* 51: 804-806.
- Liu XQ 2000. Protein-splicing intein: genetic mobility, origin and evolution. *Ann Rev Genet* 34: 61-76.
- Liu XQ, Hu Z 1997. A DnaB intein in *Rhodothermus marinus*: indication of recent intein homing across remotely related organisms. *Proc Natl Acad Sci USA* 94: 7851-7856.
- Liu XQ, Yang J 2004. Prp8 intein in fungal pathogens: target for potential antifungal drugs. *FEBS Letters* 572: 46-50.
- Matute DR, McEwen JG, Montes BA, San-Blas G, Bagagli E, Rauscher JT, Restrepo A, Morais F, Nino-Veja G, Taylor JW 2006. Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Mol Biol Evol* 23: 65-73.
- Meyer W, Marszewska K, Amirmostofian M, Igreja RP, Hardtke C, Methling K, Viviani MA, Chindamporn A, Sukroongreung S, John MA, Ellis DH, Sorrell TC 1999. 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.
- Nogami SY, Satow Y, Anraku Y 1997. Probing novel elements for protein splicing in yeast Vma1 protozyme: a study of replacement mutagenesis and intragenic suppression. *Genetics* 147: 73-85.
- Pearl EJ, Tyndall JDA, Poulter RTM, Wilbanks SM 2007. Sequence requirements for splicing by the Cne PRP8 intein. *FEBS letters* 581: 3000-3004.
- Perler FB 2000. InBase, the intein database. *Nucleic Acids Res* 22: 344-345.
- Poulter RTM, Goodwin TJD, Butler ML 2007. The nuclear encoded inteins of fungi. *Fungal Genet Biol* 44: 153-179.

- Ricci G, Mota FT, Wakamatsu A, Serafim RC, Borra RC, Franco M 2004. Canine paracoccidioidomycosis. *Med Mycol* 42: 379-383.
- Ridley M 2006. Conceitos de espécie e variação intra-específica. In M Ridley, *Evolução*, Artmed, Porto Alegre, p. 481-484.
- Steenbergen JN, Casadevall A 2003. The origin and maintenance of virulence for the pathogenic fungus *Cryptococcus neoformans*. *Microb and Infect* 5: 667-675.
- Steenbergen JN, Nosanchuk JD, Malliaris SD, Casadevall A 2004. Interaction of *Blastomyces dermatitidis*, *Sporothrix schenckii* and *Histoplasma capsulatum* with *Acanthamoeba castellanii*. *Infect Immun* 72: 3478-3488.
- Tavassoli A, Naumann TA, Benkovic SJ 2008. Production of cyclic proteins and peptides. In M Belfort, V Derbyshire, BL Stoddard, DW Wood, *Homing endonucleases and inteins*, Springer-Verlag Berlin, Heidelberg, p. 293-302.
- Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol* 31: 21-32.
- Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D 2006. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philos Trans R Soc Lond B Biol Sci* 361: 1947-1963.
- Teixeira MM 2008. *Tipagem molecular e evolução do gênero Paracoccidioides*. Master Thesis, Faculdade de Medicina da Universidade de Brasília, Brasília, 154 pp.
- Theodoro RC, Bagagli E, Oliveira C 2008. Phylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complex. *Fung Genet Biol* 45: 1284-1291.
- Yamashiro CT, Kane PM, Wolczyk DF, Preston RA, Stevens TH 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.

Artigo 3

(Fungal Genetics and Biology In press, doi:10.1016/j.fgb.2010.07.010)

PRP8 *intein* in Ajellomycetaceae-family pathogens: sequence analysis, splicing evaluation and homing endonuclease activity

Raquel Cordeiro Theodoro^{1,2}, Gerrit Volkmann², Xiang-Qin Liu², Eduardo Bagagli^{1*}

¹ Department of Microbiology and Immunology, Institute of Biosciences, UNESP, Botucatu, SP, Brazil

² Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada

* 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

Inteins are intervening sequences that are transcribed and translated with flanking host protein sequences and then self-excised by protein splicing. Bi-functional *inteins* also contain a homing endonuclease responsible for their genetic mobility. The PRP8 *intein*, the most widespread among fungi, occurs in important pathogens such as *Histoplasma capsulatum* and *Paracoccidioides brasiliensis*, from the Ajellomycetaceae family. Herein, we describe the bi-functional PRP8 *intein* in two other Ajellomycetacean pathogens, *Blastomyces dermatitidis* and *Emmonsia parva*. Sequence analysis and experimental evidence suggest that the homing endonuclease from PbrPRP8 is inactive. The splicing activity of the PRP8 *intein* from the *B. dermatitidis*, *E. parva* and *P. brasiliensis* species complex was demonstrated in a non-native protein context in *Escherichia coli*. Since the PRP8 *intein* is located in a functionally essential nuclear protein, it can be considered a promising therapeutic target for anti-fungal drugs, because inhibition of *intein* splicing should inhibit proliferation of *intein*-containing pathogens.

Key-words: Ajellomycetaceae, *intein*, PRP8, splicing, homing endonuclease

1. Introduction

Inteins, also known as internal proteins, typically range from 134 to 608 amino acids in size. After being self-excised from the host protein, the flanking protein sequences (exteins) are joined by a peptide bond, constituting a functional protein (Gogarten et al. 2002). Three types of *inteins* exist in nature: 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 an insertion of the *intein* coding sequence into *inteinless* alleles and possibly non-homologous locations (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 exons in a protein *trans*-splicing reaction (Wu et al. 1998a). *Inteins* have been detected in all three domains of life: Eukaryotes, Bacteria and Archaea (Perler 2000). They are usually found at conserved sites of essential housekeeping proteins such as metabolic enzymes, DNA and RNA polymerases, and proteases (Liu 2000).

According to Inbase, the *intein* data base (Perler 2000), about 60 species from the Eukarya domain contain an *intein* in at least one important gene. On this list, there are some amoebae and algae, but the majority of eukaryotic *inteins* were found in fungi (about 50 species). Most of these fungal species contain an *intein* in the subunit A of the vacuolar H⁺-translocating ATPase (VMA *intein*) or the spliceosomal Prp8 protein (PRP8 *intein*). The VMA *intein* is widespread among yeast fungi such as *Saccharomyces*, *Zygosacchomyces* and some *Candida* (all Ascomycota species), while the PRP8 *intein* can be found in all fungal phyla (Zygomycetes, Chytridiomycetes, Ascomycetes and Basidiomycetes) according to Inbase (Perler 2000), being more abundant among Euascomycetes (Orders Eurotiales, Onygenales and Helotiales) and in a few Basidiomycetes (Order Tremellales). Among these fungal species, there are important human pathogens: *Aspergillus nidulans* and *Aspergillus fumigatus* (Ascomycota, Order Eurotiales) (Liu and Yang 2004), *Histoplasma capsulatum* and *Paracoccidioides brasiliensis* (Ascomycota, Order Onygenales, Ajellomycetaceae Family) (Butler et al., 2006), as well as *Cryptococcus neoformans* and *Cryptococcus gattii* (Basidiomycota, Order Tremellales) (Butler et al. 2001; 2005). The PRP8 *intein* occurs in the *PRP8* gene, which encodes for one of the largest and most highly conserved nuclear proteins, Prp8, a component of U5 snRNPs implicated in the editing of pre-messenger RNA (intron splicing), a central step in gene expression (Grainger and Beggs 2005).

The PRP8 *inteins* have a functional homing endonuclease domain in most of the host species, in contrast to the VMA *inteins*, where endonuclease was proven to be functional only in three out of the thirteen yeast species that have it (Posey et al., 2004). The dS/dN ratios (the ratio between synonymous and non-synonymous substitutions) revealed that the PRP8 *intein* sequences have been constrained by selection suggesting that the *inteins* are active or were

active during most of the time that they were undergoing divergence. Comparison of the PRP8 *inteins* of some species indicated that the sequences are saturated by synonymous mutations suggesting they have been diverging for a longer time than the VMA *inteins* of yeast. Indeed the VMA homing endonuclease domains are extensively divergent, which has been attributed to their loss of function (Butler et al., 2006).

Taking into account that the PRP8 *intein* has an ancient origin and presents a vertical transfer within the Euscomycetes (the *intein* and the PRP8 phylogenies are the same), it seems reasonable to attribute its sporadic occurrence to successive losses during the evolutionary time. This is the case in *Coccidioides immitis* and *Coccidioides posadasii*, two pathogenic “*inteinless*” species from the Onygenales order that are closely related to *Uncinocarpus reesii*, which presents a PRP8 mini-*intein*. On the other hand, the presence of PRP8 *intein* in *Cryptococcus*, a Basidiomycete genus, suggested a horizontal transfer, once its close similarity to the *inteins* from Euscomycetes does not reflect the length of time since the divergence of their host organisms (Butler et al., 2006).

The PRP8 *intein* has been used as a molecular marker for phylogenetic analysis of fungal pathogens. It was shown to be an effective tool to distinguish strains of two varieties of *Cryptococcus neoformans* (*neoformans* and *grubii*) from each other and from strains of *C. gattii* (Butler and Poulter 2005). Phylogenetic analysis of this *intein* in the *P. brasiliensis* species complex (Matute et al. 2006a; Teixeira et al., 2009) (S1, PS2, PS3 and *P. lutzii* species) could also distinguish isolates from the different species (Theodoro et al. 2008). This thermo-dimorphic Ajellomyceatacean fungus is the etiologic agent of Paracoccidioidomycosis, the most important systemic mycosis in Latin America (Franco 1987). *P. brasiliensis* has been frequently recovered from clinical samples and from tissues of the nine-banded armadillo, *Dasypus novemcinctus* (Bagagli et al. 2003).

Herein, we describe the finding of PRP8 *intein* in *Blastomyces dermatitidis* and *Emmonsia parva*, two other fungal pathogens from the Ajellomycetaceae family which also causes systemic mycosis. Phylogenetic analysis of the *intein* sequences corroborate previous phylogenetic studies of the Ajellomycetaceae family that focused on the LSU and ITS1 regions

(Untereiner et al., 2004). This family is compounded by *Ajellomyces grisea* (the only non-pathogenic species from this family), *A. crescens*, *H. capsulatum* and *B. dermatitidis*, and the anamorphic *P. brasiliensis*, *Emmonsia spp* and *E. parva*. All of them are vertebrate-associated and share similar substrates and physiological characteristics, such as association with feces and soil, and the capacity to grow at 36°C. The species *P. brasiliensis*, *H. capsulatum*, and *B. dermatitidis* are thermo-dimorphic, growing as mycelia at 25°C and as yeast cells at 36°C (Untereiner et al., 2004). *E. parva* is also considered dimorphic, producing hyphae and conidiophores at 25°C, but instead of budding yeast cells, this pathogen presents adiaspores at 37°C under both *in vivo* and *in vitro* conditions. These spores are swollen blastic conidia which, once in host lung tissue, do not spread to other organs.

We evaluated the splicing function of the PRP8 *inteins* from *B. dermatitidis*, *E. parva* and four cryptic species of *P. brasiliensis* in a previously described non-native recombinant protein context in *E. coli* cells (Liu and Yang 2004). All PRP8 *inteins* proved to be active in this extein model, suggesting their usefulness for screening drugs that inhibit the *intein* excision in pathogens from the Ajellomycetaceae family. In contrast, we provide evidence that the *P. brasiliensis* PRP8 *intein* lost its endonuclease activity, which reveals important insights into the evolution of this *intein* in this particular pathogen.

2. Material and Methods

2.1. Fungal samples and DNA isolation

Five isolates of *P. brasiliensis*, one of *E. parva* and three of *B. dermatitidis* were used in this study (Table 1). DNA from *P. brasiliensis* isolates was obtained according to the method described in McCullough et al (2000), with initial cell disruption using glass beads (425-600 microns, 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 10 ng/μL. DNA samples from *B. dermatitidis* were supplied by Dr. Thomas D. Sullivan

(University of Wisconsin Medical School, Madison, Wisconsin, USA) and DNA from *E. parva* was supplied by Dr. Roseli Freitas and Adriana Vicentini Moreira (Instituto de Medicina Tropical, São Paulo, SP, Brazil).

2.2. PCR, cloning, sequencing and phylogenetic analysis of PRP8 inteins

The PRP8 *intein* was amplified by PCR using the primer set PbXhofwd and PbAgeRev for *P. brasiliensis*, and Bd/EpXhoFwd and BdAgeRev for *B. dermatitidis* (Table 2). The forward and reverse primers were designed with *Xho*I and *Age*I restriction sites, respectively, at their 5'-ends. The primers are complementary to 15 nt of the PRP8 exteins, and the design was based on PRP8 sequences from genome projects available at the Broad Institute website (<http://www.broad.mit.edu/annotation/genome>). To detect the PRP8 *intein* in *E. parva*, two degenerated primers, DEG-1fwd and DEG-2rev, were designed (Table 2), since there is no database concerning the PRP8 gene of this fungal species. All the amplicons (~1.8Kb) were inserted into the pJET1.2 cloning vector (Fermentas), and subjected to automatic DNA sequencing. After sequencing, the specific primers Bd/EpXhoFwd (the same forward primer used for *B. dermatitidis*) and EpAgeRev, with *Xho*I and *Age*I restriction sites, were designed for the amplification of the PRP8 *intein* from *E. parva* (Table 2). All sequences, including the PRP8 *intein* sequence from *H. capsulatum* (G816AR, from Broad Institute genome database), were aligned with ClustalW, and phylogenetic analyses based on Neighbor Joining was performed using Mega software 4.0v (Kumar et al., 2004), applying Kimura-2-parameters as the distance model. Bootstrap re-sampling (Felsenstein, 1985) was performed to assess support for individual nodes using 1000 replicates with random additions. The synonymous and non-synonymous substitution rates (dS/dN) were calculated at the syn-Scan web site (Gonzales et al., 2002), according to the Nei-Gojobori method (Nei & Gojobori., 1986) for the splicing and endonuclease domains. The translated sequences were compared to the VMA *intein* from *S. cerevisiae* (GenBank accession number **Q874G3**) in order to evaluate the presence or absence of the two aspartate residues (Asp-218 and Asp-326) that are involved in the formation of

double-strand DNA breaks by the homing endonuclease (Posey et al., 2004; Koufopanou and Burt, 2005).

2.3. Protein-splicing Analysis in *E. coli*

The PRP8 *intein* sequences were released from the pJET1.2-based plasmids via digestion with *Xho*I and *Age*I, and ligated into the similarly digested expression plasmid pMST (Wu et al. 1998b), thereby replacing the *Ssp* DnaB *intein* between the maltose binding protein (MBP) and thioredoxin (T) of the model extein. *E. coli* DH5 α -competent cells (Invitrogen) were transformed with the pMST-based plasmids containing the PRP8 *intein* from all fungi listed in Table 1. After transformation, the cells containing the expression plasmid were grown in liquid Luria-Bertani (LB) medium at 37°C until an OD₆₀₀ of 0.5, at which point IPTG was added to a final concentration of 0.8 mM to induce protein expression, and the cultures were maintained for 3 h at 37°C and overnight at room temperature (~22°C). The cultures were harvested and the cells were re-suspended in SDS-DTT-containing gel loading buffer (0.1 M Tris-HCl (pH 6.8), 11.1% glycerol (v/v), 2% SDS (w/v), 0.1% bromophenol blue (w/v) and 34 mM DTT), boiled for 5 min and chilled on ice for 1 min, followed by a brief centrifugation step. Samples were separated by SDS-PAGE and transferred to PVDF membrane by Western blotting. The membranes were first probed with an anti-thioredoxin antibody, followed by an anti-rabbit HRP-conjugated antibody (both from Sigma). Chemiluminescence was performed using the enhanced chemiluminescence detection kit (GEHealthcare), and signal intensities were estimated using a gel documentation system (Gel Doc 2000 with Quantity One software, Bio-Rad).

2.4. Endonuclease experiments

For expression and purification of PI-*Pbr*I proteins, the coding region of the wild-type protein (with mutations of the first and last *intein* residue to Ala and omission of a stop codon) was amplified from plasmid pMPb04T (pMST derivative plasmid with the PRP8 *intein* of Pb04

isolate) using primers Plex.fwd and Plex.rev; and the PCR product was subcloned into pJET1.2, yielding pJPI(WT). The Gly451 residue was mutated to Asp by inverse PCR on pJPI(WT) using primers PIG451Dex.fwd and PIG451Dex.rev, resulting in plasmid pJPI(G451D). Restriction sites for *Bam*HI and *Not*I were introduced into plasmid pET52b(+)/MR (Narmandakh and Bearne, 2009) by inverse PCR using primers NOT52 and BAM52 to yield pET52m. All these primers are listed in Table 2. The PI-*Pbr*I/WT and G451D coding regions were then released from plasmids pJPI(WT) and pJPI(G451D), respectively, using *Bam*HI and *Not*I, and ligated into similarly digested pET52m, resulting in the plasmid pETPI(WT), containing the wild-type of the PRP8 *intein* from *P. brasiliensis*, and pETPI(G451D), containing the aspartate residue instead of the original glycine at position 451 of the PRP8 *intein* from *P. brasiliensis*. Each plasmid was then introduced into BL21-Codon Plus (DE3)-RIL cells (Stratagene). The proteins were expressed in 120 mL cultures with 0.8 mM IPTG at room temperature for 18 h. Cells were harvested by centrifugation (4,500×g, 15 min, 4°C), resuspended in 9 mL Wash buffer (10 mM Tris-HCl (pH 8.1), 300 mM NaCl, 20 mM imidazole), and lysed by passage through a French pressure cell (14,000 PSI, performed twice). The soluble protein fraction was obtained by centrifugation of the crude cell lysate (10,000×g, 25 min, 4°C), and incubated with 1 mL Ni-NTA agarose (Qiagen) for 1 h at 4°C. The agarose was poured into a disposable column, washed with 20 mL Wash buffer, and proteins were eluted with 2 mL Elution buffer (10 mM Tris-HCl (pH 8.1), 300 mM NaCl, 250 mM imidazole). Proteins were then dialyzed against 10 mM sodium phosphate buffer (pH 7.0) for 18 h at 4°C, and concentrated by centrifugation (4,000×g, 15 min, 4°C) using Amicon Centrifugal devices (MWCO: 10,000 Da, Millipore). Protein concentrations were determined by the DC Protein Assay (Bio-Rad). The final protein preparations had concentrations of ~40 mg/mL, and were >90% pure as judged by Coomassie-stained SDS-PAGE gels.

To construct a plasmid substrate for endonuclease assays, the putative 51 bp *P. brasiliensis* homing endonuclease site (HES) was created by modifying the *inteinless* *S. cerevisiae* genomic *PRP8* locus to the sequence of an *inteinless* *P. brasiliensis* *PRP8* locus. This was achieved by performing two separate PCR reactions on genomic DNA from *S.*

cerevisiae with primer pairs P1-fwd/P1-rev and P2-fwd/P2-rev. The two products were annealed and amplified using primers P1-fwd and P2-rev, and the fusion PCR product was cloned into pJET1.2, yielding plasmid pJHES1. The HES region was then released from pJHES1 by digestion with *Bgl*II, and the fragment was cloned into a modified version of the pDrive cloning vector (Qiagen), which carried a single site for *Bgl*II at the multiple-cloning site, resulting in plasmid pDHES1. Plasmid pDHES2 was obtained by ligating the 814 bp-*Hind*III/*Not*I fragment released from pJHES1 into similarly digested pDHES1.

Endonuclease assays contained ~10 µg PI-*Pbr*I protein and ~2 µg plasmid pDHES2. Incubation was performed for 4-6 h at 37°C in the following buffers: PI-*Scel* buffer (50 mM KCl, 25 mM Tris-HCl, 2.5 mM 2-mercaptoethanol, pH 7.5) with and without the addition of MgCl₂ or MnCl₂ to 2.5 mM, buffer 1 (10 mM Bis Tris Propane-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.0), buffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9), buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9), buffer 4 (50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9), buffer 5 (100 mM KCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 9.2), buffer 6 (100 mM KCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.0), buffer 7 (100 mM KCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 8.0), or buffer 8 (100 mM KCl, 10 mM Tris-HCl, 1 mM DTT, 2.5 mM MnCl₂, pH 8.0) with and without the addition of 0.1 mg/mL bovine serum albumin (BSA). Reactions were stopped by the addition of 6× loading buffer (Fermentas), and DNA was separated by electrophoresis on 1% Tris-Acetate EDTA (TAE) agarose gels containing ethidium bromide.

3. Results

3.1. PRP8 intein in fungal pathogens from *Ajellomycetaceae* Family

All the fungal species evaluated herein have a full-length PRP8 *intein*, containing the splicing domain and the endonuclease domain belonging to the LAGLIDADG family as identified

by the conserved blocks C and E. The new sequences from *B. dermatitidis* and *E. parva* isolates can be accessed in GenBank by the numbers: **GQ980266** – **GQ980269**. The PRP8 *intein* amino-acid sequences from *P. brasiliensis* species complex (S1, PS2, PS3 and *P. lutzii*), *B. dermatitidis* and *E. parva* isolates were aligned together with the PRP8 *intein* amino-acid sequence from *H. capsulatum* (**G816AR**, from the Broad Institute genome database) whereas the splicing and endonuclease domains were identified by comparison with the VMA *intein*. The species S1, PS2 and PS3 from the *P. brasiliensis* complex presented 573 amino acids (167 corresponding to the splicing domain and 406 to the endonuclease domain). The species *P. lutzii* has a deletion of 3 amino acids in the N-terminus splicing domain, totaling 570 amino acids, as documented previously (Theodoro et al. 2008). The respective PRP8 *inteins* from *B. dermatitidis* and *E. parva* were successfully detected and amplified by PCR. Both *inteins* are 526 aa long (148 and 146 amino acids for the splicing domain, and 378 and 380 for the endonuclease domain in *B. dermatitidis* and *E. parva*, respectively) (Figure 1). The splicing domain is more conserved among these fungal species when compared to the endonuclease domain (Figure 1). This is also corroborated by the ratio between synonymous and non-synonymous substitutions. The splicing domain is saturated with synonymous substitutions, having a mean dS/dN ratio of 8.92, while for the endonuclease domain it is only 2.77 (Tables 3 and 4). The first aspartate residue (block C, marked with an arrow in figure 1) is present in all isolates, except in *E. parva*, in which it is replaced by a glutamate residue. The second aspartate (block E, marked with an arrow in figure 1) is substituted for serine in *P. lutzii*, and for glycine in *E. parva* and *P. brasiliensis* isolates from S1, PS2 and PS3. The Neighbor Joining tree for the PRP8 *intein* from the fungal pathogens of Ajellomycetaceae family corroborated, with high bootstrap values, previous analyses (with other nuclear coding genes) concerning the relation among the different species from the family (Untereiner et al., 2004) and also among the cryptic species of *P. brasiliensis* (Teixeira et al., 2009) (Figure 2).

3.2. PRP8 intein splicing in *E. coli*

The PRP8 *intein* sequence from each of the different fungal pathogens was inserted between maltose binding protein (MBP) and thioredoxin (T) sequences in the pMST plasmid. The expected spliced protein (MBP+T) has a molecular weight of 56 kDa, whereas the unspliced precursor protein is much larger (~115 kDa). Analysis of *E. coli* cells expressing the precursor proteins by Western blotting using an antibody directed against thioredoxin revealed a complete protein splicing reaction at 37°C for species S1 (isolate Pb265), PS2 (isolates BT84 and Pb04) and PS3 (isolate 60855) of *P. brasiliensis*, and isolates SLH4081 and MYA-2586 of *B. dermatitidis* (Figure 3), as judged by the presence of a strong protein signal at ~57 kDa and the absence of a signal >100 kDa. In contrast, faint signals for the precursor protein were detectable in the *B. dermatitidis* isolate 26199 and *P. lutzii* from the *P. brasiliensis* species complex, corresponding to splicing efficiencies of approximately 92% and 96%, respectively. All isolates performed a complete splicing reaction at room temperature (Figure 3).

The PRP8 *intein* from the *E. parva* isolate represents a notable exception to all tested isolates because in addition to the product of protein splicing, considerable amounts of a protein species at ~12 kDa were generated both at room temperature and 37°C. This protein likely corresponds to the thioredoxin part of the precursor protein generated by a cleavage reaction at the C-terminal end of the *intein* (C-cleavage). This non-productive cleavage reaction accounts for 26% of precursor protein conversion at room temperature and 68% at 37°C (Figure 3), with protein splicing corresponding to 74% and 24%, respectively. At room temperature, the *E. parva* isolate further showed two protein signals at ~20 kDa, both of which do not conform to the other possible *intein* cleavage reaction (N-terminal cleavage would generate an *intein*-thioredoxin protein of ~73 kDa), and are thus likely degradation products from the precursor or spliced protein.

3.3. Endonuclease activity of the *Pbr Pb04 PRP8 intein (PI-PbrI)*

As mentioned above, our recent analysis of the amino-acid sequences of the bi-functional *inteins* present in the *P. brasiliensis* species complex revealed the presence of only one acidic residue known to be essential for endonuclease activity (Asp159 in motif C), while the second Asp residue appeared to be mutated to either Gly or Ser, depending on the species isolated (Theodoro et al., 2008). We therefore wished to determine whether these mutations abrogated endonuclease activity. We focused our studies on the *intein* of the Pb04 isolate (T10B1-like), which contains a Gly in the second LAGLIDADG endonuclease motif (Gly451 in motif E). As per nomenclature for *intein*-encoded endonucleases (Perler et al. 1994) we designated this bi-functional *intein* PI-*PbrI*.

The protein was produced in recombinant form in *E. coli* cells with a C-terminal His₁₀-tag (PI-*PbrI*/WT), and was purified by metal ion affinity chromatography to near homogeneity (Figure 4A). We also produced a protein, in which the Gly451 residue was restored to Asp (PI-*PbrI*/G451D) in order to assess the endonuclease activity of this “conventional” *intein* endonuclease. The PI-*PbrI*/G451D protein behaved similarly to the wild-type (WT) protein in the purification process (Figure 4A), indicating that the G451D mutation did not negatively affect the overall protein structure.

The endonuclease activity of both proteins was studied using the plasmid pDHES2, which contains two identical homing sites of 51 bp, corresponding to the sequence of an *intein*-less genomic locus in *P. brasiliensis* (Figure 4B). We chose the length of the homing site according to the observation that other homing endonucleases have recognition sites of 12-40 bp (Belfort and Roberts 1997; Guhan and Muniyappa 2003). Using plasmid DNA provides the additional opportunity to detect single-strand breaks (nicks) as a result of enzymatic activity, and can thus reveal more information than linear DNA in endonuclease assays (Guhan and Muniyappa 2002). Introduction of double-strand breaks within the HES regions of pDHES2 should lead to two fragments with respective sizes of 4361 bp and 835 bp. Digestion of pDHES2

with the restriction endonuclease *Bst*BI served as a control, because the HES region contains a site for *Bst*BI (Figure 4B), and should thus lead to DNA fragments of similar size.

We initially tested the endonuclease activity of PI-*Pbr*/WT and PI-*Pbr*/G451D using buffer conditions that allowed efficient DNA cleavage by PI-*Scel*, the bi-functional *Sce* VMA *intein* (Gimble and Thorner 1992); however, we did not observe the expected DNA fragments when compared to digestion of pDHES2 with *Bst*BI (Figure 4C, compare lanes 4 and 7 to lane 3). Addition of divalent cations (Mg^{2+} and Mn^{2+}) to the reactions also did not indicate activity of either the WT or G451D enzyme (Figure 4C, lanes 5-6 and 8-9). We then tested both enzymes for cleavage activity in a variety of different buffers (see Materials and Methods) with and without the addition of bovine serum albumin (BSA), which is known to enhance the stability of numerous restriction endonucleases, but none of the conditions led to the two DNA fragments expected from double-strand cleavage (Figure 4D).

The DNA patterns consistently showed three bands, of which the slowest migrating band likely corresponds to multimeric supercoiled plasmid DNA (Guhan and Muniyappa 2002). The two remaining bands, which migrated more rapidly, could be relaxed plasmid DNA due to single-strand breaks (nicks) and monomeric supercoiled plasmid DNA, of which the latter is expected to present the faster migration (Guhan and Muniyappa 2002). Although incubation of pDHES2 with the restriction endonucleases *Nde*I and *Nhe*I, which both lack cleavage sites in the plasmid, also resulted in this three-banded pattern, we do note that reactions containing PI-*Pbr*/G451D showed an increase in the band intensity of the putative nicked DNA species and a concomitant decrease of the putative monomeric supercoiled DNA when compared to the intensities seen in reactions with PI-*Pbr*/WT (Figure 4D). This difference in band intensity appeared to be BSA-independent.

4. Discussion

New PRP8 inteins in the Ajellomycetaceae Family

Herein we have described two new PRP8 *inteins*, one in *B. dermatitidis* and one in *E. parva*, and compared them with the PRP8 *intein* from *H. capsulatum* and *P. brasiliensis*. All these fungi are human pathogens from Ajellomycetaceae Family (Onygenales Order) and are known to cause systemic mycoses. The geographic distribution of *B. dermatitidis* and *P. brasiliensis* is restricted to North America and Latin America, respectively, while *H. capsulatum* and *E. parva* are distributed more widely. According to phylogenetic analysis of LSU and ITS1 regions, Untereiner et al (2004) proposed this new Family, and a more recent study also includes the uncultivable pathogenic fungus *Lacazia loboi* (Vilela et al., 2009) in this group. This fungus is known to cause a subcutaneous mycosis in dolphins and humans, especially those from the Amazon basin, Brazil (Taborda et al., 1999). We have tried to amplify the *L. loboi* PRP8 *intein* from DNA samples extracted from mice artificially infected with *L. loboi*, using different combinations of degenerated primers; however, the obtained PCR products contained either other genomic regions of the pathogen or mouse genomic regions (data not shown). The *L. loboi* genome project initiative, now with few sequences available in the Broad Institute database, will certainly be helpful for detection of *inteins*, mainly in the PRP8 gene, since all the other Ajellomycetaceae members present it. Moreover, the knowledge of the complete genome of this pathogen will clarify the peculiar lifestyle aspects of *L. loboi* in comparison to those of the other Ajellomycetacean fungi.

Although many *inteins* are thought to be horizontally inherited, *inteins* in the PRP8 gene described in Euascomycetes – including *Aspergillus nidulans*, *A. fumigatus*, *Paracoccidioides brasiliensis*, *Histoplasma capsulatum* and *B. cinera* – seem to be vertically transmitted due to the similarity observed in phylogenetic analyses using these *inteins* and other nuclear sequences such as the ribosomal genes (Butler et al. 2006). The phylogenetic analysis of the PRP8 *intein* in *B. dermatitidis*, *H. capsulatum*, *E. parva* and *P. brasiliensis* corroborated the previous phylogenetic analysis of the LSU and ITS (Untereiner et al., 2004), confirming the use of this *intein* for evolutionary studies at the species level (Theodoro et al. 2008).

Degeneration of PRP8 intein homing endonucleases

The dS/dN analysis of all the bi-functional *inteins* reveals a stronger constrained selection for the protein splicing domain when compared to the endonuclease domain, which is expected since a functional splicing domain is essential for the expression of the mature Prp8 protein, being therefore indispensable for fungal cell life. The dS/dN values higher than 1 for the homing endonuclease domain also suggest a constrained selection and perhaps a functional feature. However, the substitutions in the second essential aspartate in *P. brasiliensis* (species S1, PS2, PS3) and *P. lutzii*, and in both essential aspartates in *E. parva* indicate that this endonuclease might have lost its activity in these fungal lineages. Indeed, we have provided initial evidence that the homing endonuclease found in the *PRP8* gene of the *P. brasiliensis* Pb04-isolate (PI-*Pbr1*) is most likely inactive towards DNA double-strand cleavage at an *inteinless PRP8* locus due to the presence of only one essential Asp residue in the enzyme's active site. According to the homing cycle model of invasion, fixation, degeneration, loss and reinvasion proposed for the VMA *intein* (Burt & Koufopanou 2004), once all the sites are occupied by the *intein* in a population there is no selection for endonuclease activity, so that this domain becomes degenerated by mutations and deletions. This scenario may account for the mutation of the second essential Asp residue in the homing endonuclease from these Ajellomyceatacean fungi.

Introduction of the second Asp in motif E of PI-*Pbr1* did not restore cleavage ability but only led to a potential increase in nicking activity (Figure 4D). Although we limited the putative *P. brasiliensis* homing endonuclease site to 51 bp, the sequence was imbedded into a 455 bp region of the *S. cerevisiae* *inteinless PRP8* gene; these sequences have a similarity of 68.8%, and given the fact that homing endonucleases are less specific towards their recognition sequences than restriction endonucleases (Belfort and Roberts 1997; Posey et al., 2004), one might have expected DNA cleavage activity of the PI-*Pbr1* protein beyond the exact 51 bp *P. brasiliensis* sequence. Moreover, the ability of *intein* homing endonucleases to perform DNA double-strand cleavage has been found to correlate to the isoelectric point (pI) of the protein, with inactive endonucleases having a slightly acidic pI (< 5.8) (Okuda et al. 2003). This

observation appears consistent with our finding that the PI-*Pbrl* endonuclease is inactive because the protein has a pI of 5.5.

Whether the PI-*Pbrl* enzyme is truly inactive towards DNA cleavage and *intein* homing as indicated by our results will ultimately require the study of the PRP8 *intein* in its natural host. Since the *in vitro* conditions used in this study are inherently artificial, we cannot exclude the possibility that the protein is indeed active under as of yet unidentified conditions. It is further possible that the PI-*Pbrl* protein requires cellular co-factors for activity. Most recently, the homing activity of the *Botrytis cinerea* PRP8 *intein* was studied in different *B. cinerea* strains (Bokor et al., 2010), and similar studies could be envisioned with *P. brasiliensis*. However, this represents a significant technical challenge: since *P. brasiliensis* is not polymorphic for the presence of the PRP8 *intein* (in contrast to *B. cinerea*), an *intein*-less strain would have to be engineered for mating with an *intein*-containing strain. Unfortunately, *P. brasiliensis* cannot be genetically modified easily and conditions for controlled mating are still far from optimal. We hope that with further study of this fungus it will soon be possible to overcome these technical challenges, and the activity of the PI-*Pbrl* protein can be evaluated in an unequivocal manner.

Protein splicing of new PRP8 inteins and its implications

Besides being an important phylogenetic tool, the PRP8 *intein* can also be considered a therapeutic target for pathogenic fungi because it is located in an essential and most highly conserved eukaryotic gene, namely, *prp8*, which encodes for the Prp8 protein involved in mRNA splicing. Furthermore, the absence of this *intein* in the vertebrate hosts makes it a safe target for drugs that may inhibit its splicing function (Liu and Yang 2004).

Our results showed that the splicing activity of the PRP8 *intein* in the non-native extein context used is complete at room temperature and is almost 100% efficient at 37°C for *B. dermatitidis* and for the S1, PS2 and PS3 species from the *P. brasiliensis* complex. Interestingly, *P. lutzii* was the only species from the *P. brasiliensis* complex that showed some accumulation of unspliced precursor protein at 37°C. This species presents a deletion of three amino acids in the N-terminal splicing domain, and many substitutions along the *intein* sequence (splicing and

endonuclease domains) when compared to other *P. brasiliensis* species, as already documented (Theodoro et al. 2008). This isolate also differs substantially from the others in relation to nuclear encoded genes and it is the most divergent in the *P. brasiliensis* species complex, having been separated from the other three species for at least 32 million years. (Carrero et al., 2008; Teixeira et al., 2009). This long divergence period may have resulted in the emergence of differences in morphological features including yeast cell size and conidium shape (Teixeira et al., 2009), and also molecular characteristics, such as the sequence and splicing activity of the PRP8 *intein*, due to reproductive isolation and genetic drift.

Among all the fungal isolates used in the present work, the PRP8 *intein* of *E. parva* presented the lowest splicing efficiency at 37°C, and also showed a substantial amount of C-terminal cleavage. It is not clear why this kind of cleavage happens, but it is not likely due to the foreign protein (extein) context because five native extein amino acids were present on both sides of the *intein*, and because the same foreign protein (extein) context did not lead to the cleavage with the other PRP8 *inteins* evaluated. The observed cleavage was most likely a property of the *E. parva* PRP8 *intein*. It is not clear whether the cleavage also happens in the fungal cell, but if it does, the cleavage is an aborted (or incomplete) splicing, and therefore, it would reduce the production of mature/functional PRP8 protein.

A similar low splicing efficiency was previously documented for the PRP8 *intein* of *H. capsulatum* using the same non-native extein approach (Liu and Yang 2004). It is interesting to note that, as dimorphic fungi, all the species herein evaluated must convert from the mycelia (M) to yeast (Y) form during infection, or under high temperature conditions (36-37°C). If the splicing efficiency of the *intein* at 37°C is similarly low in the native Prp8 protein, the M-Y transition could be disturbed because many different genes must be transcribed and correctly translated for the fungus to assume the yeast morphology and physiology. Therefore, in this case, the PRP8 *intein* might decrease the fitness of the pathogen (Theodoro and Bagagli, 2009). Taking into account that the PRP8 *intein* from *H. capsulatum* is a bi-functional *intein*, probably with a functional homing endonuclease domain (Poulter et al. 2007), it is intriguing that this homing endonuclease domain appears to be maintained by selection (high dS/dN values and presence

of both catalytic aspartate residues). In this case the balance between decreased fitness of the host and active spread of the parasitic element could lead to long persistence of the homing endonuclease, especially when the selective disadvantage to the carrier and the mobility of the *intein* are not the same in different subpopulations (Gogarten and Hilario 2006). This hypothesis seems plausible to explain the maintenance of the PRP8 *intein* in *H. capsulatum*, also given its nearly worldwide distribution due to propagation by flying hosts (bats, birds), which resulted in many subpopulations, including cryptic species.

The same could be valid for *E. parva*, which has a wider distribution when compared to *P. brasiliensis* and *B. dermatitidis*, and occurs in endemic zones in southwestern US, Australia and Eastern Europe (Hoog et al., 2000). However, its homing endonuclease domain, despite high dS/dN values (>1), seems to be degenerated and inactive due to substitutions in both essential aspartates. Moreover, in *E. parva*, the low splicing efficiency may contribute to the non-dispersive and asymptomatic aspect of adiaspiromycosis, since the *E. parva* adiaspore yeast phase remains in the lungs and does not spread to other internal organs. Curiously, it is also interesting to note that both *E. parva* and *H. capsulatum* require rich media such as BHI (Brain Heart Infusion) medium for growth at 37°C. Could the low splicing efficiency of the PRP8 *intein* at 37°C observed for *E. parva* and *H. capsulatum* contribute to the slow growth and/or high nutrient demanding at 37°C? A more in-depth investigation into PRP8 *intein* splicing and mycological features of more isolates from *E. parva* and *H. capsulatum* could provide an answer to this intriguing biological question.

PRP8 inteins as potential antifungal drug targets

The ability of *inteins* to function in non-native contexts has been widely studied (Davis et al. 1992; Cooper et al. 1993; Chong et al. 1996; Xu and Perler 1996) and allowed a more detailed understanding about the necessary residues for the occurrence of splicing. Using this approach, Pear and colleagues (2007) analyzed the sequence requirements for splicing by the Cne PRP8 *intein* by alanine scanning mutagenesis. Their results confirm the importance of the splicing junction residues and the threonine and histidine conserved residues from the block B.

In the last twenty years the robustness and flexibility of *inteins* suggested their use in many applications (Perler and Adam 2000). Many selection systems for the cis-splicing activity have been reported using *E. coli* as a model, including employment of the *Mtu* RecA *intein* using β -galactosidase (Davis et al. 1992), thymidilate synthase (Derbyshire et al. 1997), aminoglycoside-3'-phosphotransferase (Daugelat and Jacobs 1999), and green fluorescent protein (GFP) (Gangopadhyay et al. 2003) as reporter proteins, and also for Sce VMA in the *Saccharomyces cerevisiae* model utilizing yeast Vat2p (Cooper et al. 1993) and Gal4/Gal80 (Zeidler et al. 2004) as reporter proteins. All these selection systems use bacterial/fungal growth to address the splicing activity and, although they are good options for fast splicing activity selection, a possible disadvantage of growth-based screens is that substances that attenuate bacterial growth would interfere with the assay. Neither of the cell-based assays for the inhibition of protein splicing has yet been applied successfully to the screening of large compound libraries.

A reported screening system for *intein* splicing *in vitro* has been developed through the insertion of the *Mtu* RecA into a modified GFP (Gangopadhyay et al. 2003). In this case, the accumulated precursor protein in *E. coli* forms insoluble inclusion bodies, which are dissolved in 8M urea, purified and refolded as a stable unspliced precursor by adding EDTA. This system permits a high-throughput screening of compounds that might inhibit the *intein* splicing reaction. Using this approach, more than 85,000 compounds were screened for splicing inhibition of the *Mtu* RecA *intein*, and a number of electrophilic compounds were identified that inhibit splicing by binding to the catalytic cysteine residue (Paulus 2003).

The ability of the PRP8 *intein* from all Ajellomycetaceae pathogens to self-splice from a non-native protein in *E. coli* opens new avenues for future studies and searches for splicing inhibitors that may constitute potential antifungal drugs. Since PRP8 *inteins* from different pathogens have the same corresponding insertion site in PRP8 protein and a highly conserved splicing domain (Liu and Yang 2004), splicing inhibitor drugs may possess the advantage of simultaneous action against most fungal systemic diseases.

Acknowledgments

We thank Dr. Roseli Freitas and Adriana Vicentini Moreira (Instituto de Medicina Tropical, São Paulo, SP, Brazil) for supplying *E. parva* DNA, Dr. Leonel Mendoza (Michigan State University, Michigan, USA) for supplying *L. lobo* DNA, and Dr. Tom Sullivan (Dept of Pediatrics and Internal Medicine, University of Wisconsin-Madison, USA) for supplying *B. dermatitidis* DNA samples used in this study. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp Grant Numbers: 06/03597-4 and 07/01306-5), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes Grant number: 464108-6) and the National Science and Engineering Research Council (NSERC) of Canada.

References

- Bagagli, E., Franco, M., Bosco S.M.G., Hebelers-Barbosa, F., Trinca, L.A., and Montenegro, M.R., 2003. High frequency of *Paracoccidioides brasiliensis* infection in armadillos (*Dasypus novemcinctus*): an ecological study. *Med. Mycol.* 41, 217-223.
- Belfort, M., and Roberts, R.J., 1997. Homing endonucleases: keeping the house in order. *Nucleic acids res.* 25, 3379-3388.
- Bokor, A.A.M., Kan, J.A.L.V., Poulter R.T.M., 2010. Sexual mating of *Botrytis cinerea* illustrates PRP8 *intein* HEG activity. *Fungal Genet. Biol.* doi: 10.1016/j.fgb.2010.01.003.
- Butler, M.L., Goodwin, T.J.D., Poulter, R.T.M., 2001. A nuclear-encoded *intein* in the fungal pathogen *Cryptococcus neoformans*. *Yeast* 18, 1365-1370.
- 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.
- 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.

Cooper, A.A., Chen, Y.J., Lindorfer, M.A., Stevens, T.H., 1993. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. EMBO J. 12, 2575-2583.

Daugelat, S., Jacobs, W.R.Jr., 1999. The Mycobacterium tuberculosis recA *intein* can be used in an ORFTRAP to select for open reading frames. Protein Sci. 8, 644-653.

Davis, E.O., Jenner, P.J., Brooks, P.C., Colston, M.J., Sedgwick, S.G., 1992. Protein splicing in the maturation of *M. tuberculosis* recA protein: a mechanism for tolerating a novel class of intervening sequence. Cell 71, 201-210.

Derbyshire, V., Wood, D.W., Wu, W., Dansereau, J.T., Dalgaard, J.Z., Belfort, M., 1997. Genetic definition of a protein-splicing domain: functional mini-*inteins* support structure predictions and a model for *intein* evolution. Proc. Nat. Acad. Sci. USA 94, 11466-11471.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783-791.

Franco, M., 1987. Host-parasite relationships in paracoccidioidomycosis. J. Med. Vet. Mycol. 25, 5-18.

Gangopadhyay, J.P., Jiang, S.Q., Paulus, H., 2003. An in vitro screening system for protein splicing inhibitors based on green fluorescent protein as an indicator. Anal. Chem. 75, 2456-2462.

Gimble, F.S., Thorner, J., 1992. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. Nature 357, 301-306.

Gogarten, J.P., Hilario, E., 2006. *Inteins*, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. BMC Evol. Biol. 6, 1-5.

Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L., Hilario, E., 2002. *Inteins*: structure, function, and evolution. Annu. Rev. Microbiol. 56, 263-287.

Gonzales, M., Dugan, J., Shafer, R., 2002. Synonymous-non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). Bioinformatics 18, 886-887.

Grainger, R.J., Beggs, J.D., 2005. Prp8 protein: at the heart of the spliceosome. *RNA* 11, 533-557.

Guhan, N., Muniyappa, K., 2002. *Mycobacterium tuberculosis* RecA *intein* possesses a novel ATP-dependent site-specific double-stranded DNA endonuclease activity. *J. Biol. Chem.* 277, 16257-16264.

Guhan, N., Muniyappa, K., 2003. Structural and functional characteristics of homing endonucleases. *Crit. Rev. Biochem. Mol.* 38, 199-248.

Hebeler-Barbosa, F., Morais, F.V., Montenegro, M.R., Kuramae, E.E., Montes, B., McEwen, J.G., Puccia, R., Bagagli, E., 2003. Sequence comparison of the internal transcribed spacer regions and gp 43 in *Paracoccidioides brasiliensis* for patients and armadillos *Dasypus novemcinctus*. *J. Clin. Microbiol.* 41, 5735-5737.

Hoog, G.S., Guarro, J., Gené, J., Figueras, M.J., 2000. Atlas of Clinical Fungi. Centraalbureau voor Schimmelcultures, Utrecht.

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.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5, 150-163.

Liu, X.Q., 2000. Protein-splicing *intein*: Genetic mobility, origin, and evolution. *Annu. Rev. Genet.* 34, 61-76.

Liu, X.Q., Yang, J., 2004. Prp8 *intein* in fungal pathogens: target for potential antifungal drugs. *FEBS Lett.* 572, 46-50.

Matute, D.R., McEwen, J.G., Puccia, R., Montes, B.A., San-Blas, G., Bagagli, E., Rauscher, J.T., Restrepo, A., Morais, F., Nino-Vega, G., et al. 2006a. Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Mol. Biol. Evol.* 23, 65-73.

Matute, D.R., Sepulveda, V.E., Quesada, L.M., Goldman, G.H., Taylor, J.W., Restrepo, A., and McEwen, J.G. 2006b. Microsatellite analysis of three phylogenetic species of *Paracoccidioides brasiliensis*. *J. Clin. Microbiol.* 44, 2153-2157.

McCullough, M.J., DiSalvo, A.F., Clemons, K.V., Park, P., and Stevens, D.A., 2000. Molecular epidemiology of *Blastomyces dermatitidis*. Clin. Infect. Dis. 30, 328-335.

Narmandakh, A., and Bearne, S.L., 2009. Purification of recombinant mandelate racemase: improved catalytic activity. Protein Expres. Purif. 69, 39-46.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418-426.

Okuda, Y., Sasaki, D., Nogami, S., Kaneko, Y., Ohya, Y., Anraku, Y., 2003. Occurrence, horizontal transfer and degeneration of VDE *intein* family in Saccharomycete yeasts. Yeast 20, 563-573.

Paulus, H., 2003. *Inteins* as targets for potential antimycobacterial drugs. Front. Biosci. 8, 1157-1165.

Pearl, E.J., Tyndall, J.D.A., Poulter, R.T.M., Wilbanks, S.M., 2007. Sequence requirements for splicing by the Cne PRP8 *intein*. FEBS Letters 581, 3000-3004.

Perler, F.B., 2000. InBase, the *Intein* Database. Nucleic acids res. 28, 344-345.

Perler, F.B., Adam, E., 2000. Protein splicing and its applications. Curr. Opin. Biotech. 11, 377-383.

Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J., Belfort, M., 1994. Protein splicing elements: *inteins* and *exteins*-a definition of terms and recommended nomenclature. Nucleic acids res. 22, 1125-1127.

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.

Taborda, P.R., Taborda, V.A., McGinnis, M.R., 1999. *Lacazia loboi*, the etiologic agent of lobomycosis. J. Clin. Microbiol. 37, 2031-2033.

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. 2009. Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides* genus. *Mol Phylogenet Evol.* 52, 273-283.

Theodoro, R.C., and Bagagli, E., 2009. *Inteins* in pathogenic fungi: a phylogenetic tool and perspectives for therapeutic applications. *Mem. I. Oswaldo Cruz* 104, 497-504.

Theodoro, R.C., Bagagli, E., Oliveira, C., 2008. Phylogenetic analysis of PRP8 *intein* in *Paracoccidioides brasiliensis* species complex. *Fungal Genet Biol* 45: 1284-1291.

Untereiner, W.A., Scott, J.A., Naveau, F.A., Sigler, L., Bachewish, J., Angus, A., 2004. The Ajellomycetaceae, a new family of vertebrate-associate Onygenales. *Mycol.* 96, 812-821.

Vilela, R., Rosa P.S., Belone, A.F., Taylor, J.W., Diorio, S.M., Mendoza, L., 2009. Molecular phylogeny of animal pathogen *Lacazia loboi* inferred from rDNA and DNA coding sequences. *Mycol. Res.* 113, 851-857.

Wu, H., Hu, Z., Liu, X.Q., 1998a. 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.

Wu, H., Xu, M.Q., Liu, X.Q., 1998b. Protein trans-splicing and functional mini-*inteins* of a cyanobacterial dnaB *intein*. *Biochim. Biophys. Acta* 1387, 422-432.

Xu, M.Q., Perler, F.B., 1996. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* 15, 5146-5153.

Zeidler, M.P., Tan, C., Bellaiche, Y., Cherry, S., Hader, S., Gayko, U., Perrimon, N., 2004. Temperature-sensitive control of protein activity by conditionally splicing *inteins*. *Nat. Biotechnol.* 22, 871-876.

Figure Legends

Figure 1. Amino-acid sequences for the splicing and endonuclease domains of *P. brasiliensis* species complex, *B. dermatitidis*, *H. capsulatum* and *E. parva*.

Alignment of PRP8 *intein* amino acids sequences including splicing domain (A) and endonuclease domain (B) from *B. dermatitidis* (Bd), *H. capsulatum* (Hc), *P. brasiliensis* (Pb), *P. lutzii* and *E. parva* (Ep). Conserved residues are marked with “*”, similar residues are marked with “.”, and the arrows highlight the positions of the two essential aspartate residues in the *Sce* VMA *intein*.

Figure 2. Phylogenetic analysis of the PRP8 *intein* in Ajellomycetacea family.

Neighbor Joining Tree based on the DNA sequences of the PRP8 *intein* from all fungal isolates listed in Table 1. Bootstrap proportions are shown above branches.

Figure 3. Splicing activity of the PRP8 *intein* from *P. brasiliensis* species complex, *B. dermatitidis* and *E. parva*, in a non-native protein model.

Western blotting analysis of PRP8 *intein*-splicing activity. A recombinant fusion protein (MIT) was expressed in *E. coli* at room temperature (RT) and at 37°C to observe protein splicing, in which the *intein* (I) was flanked by maltose binding protein (M) and thioredoxin (T). The *inteins* were from fungal isolates from species of the *P. brasiliensis* complex (Pb01, from *P. lutzii* species, Pb265, from S1 species, Bt84 and Pb04, from PS2 and 60855 from PS3), *B. dermatitidis* (isolates SLH14081, MYA-2586 and 26199) and *E. parva*. The precursor protein (MIT) is indicated by the letter P, the spliced protein (MT) by the letter S, and letter C marks a product possibly resulting from C-cleavage (only the thioredoxin protein).

Figure 4. Expression and activity assay of the wild-type and G451D mutant of PI-*PbrI* endonuclease.

A. The wild-type (WT) and G451D mutant version of the PI-*PbrI* endonuclease were expressed in *E. coli* (lane 1: total cell lysate of uninduced cells; lane 2: lysate of induced cells), and the soluble protein fractions (lanes 3) were incubated with Ni-NTA agarose (lanes 4: flow-through fraction of unbound proteins). After washing, proteins bound to the agarose were eluted with 250 mM imidazole (lanes 5). Proteins were resolved on a 12.5% SDS gel, and stained with Coomassie Brilliant Blue. Lane M shows a protein standard with protein

molecular weights given in kDa. **B.** Map of plasmid pDHES2 used for endonuclease assays. Two putative homing endonuclease sites (HES) for PI-*PbrI* are present in the plasmid, which also contain a cleavage site for the restriction endonuclease *BstBI*. The nucleotide and amino acid sequences of HES are shown in detail, with the arrow indicating the position where the *PbrI* PRP8 intein is inserted in nature. **C.** Plasmid pDHES2 was incubated with 10 µg of either PI-*PbrI*/WT (lanes 4-6) or PI-*PbrI*/G451D (lanes 7-9) in PI-*Scel* buffer for 2 h at 37°C in the absence of divalent cations (lanes 4 and 7), and in the presence of 2.5 mM Mg²⁺ (lanes 5 and 8) or Mn²⁺ (lanes 6 and 9). Control reactions include incubation without enzyme (lane 1), with *PstI* (lane 2), which has only one restriction site in pDHES2, and *BstBI*, which has two sites in pDHES2 (lane 3). Lane M is a DNA size standard (GeneRuler DNA Ladder Mix, Fermentas) with sizes of selected bands indicated in bp. **D.** Plasmid pDHES2 was incubated with either the wild-type version of PI-*PbrI* (WT) or the G451D mutant as indicated for 6 h at 37°C. The numbers above the lanes each indicate a specific buffer (see Materials and Methods for buffer compositions). BSA (at 0.1 mg/mL) was also added to some reactions as indicated. Control reactions include incubation with the restriction endonucleases *NdeI* and *NheI*, neither of which has a restriction site in pDHES2. Lanes M corresponds to DNA size standards with sizes given in bp (shown in the WT panel).

Tables:

Table 1. Fungal isolates used in this work and their origin.

Fungal species	Isolate	Locality	Host	References
<i>P. brasiliensis</i> (S1)	Pb265	Botucatu, SP-Brazil	Human	Hebeler-Barbosa et al. (2003)
<i>P. brasiliensis</i> (PS2)	BT84	Botucatu, SP-Brazil	Human	Hebeler-Barbosa et al. (2003)
<i>P. brasiliensis</i> (PS2)	Pb04	Sao Paulo, Brazil	Human	Hebeler-Barbosa et al. (2003)
<i>P. brasiliensis</i> (PS3)	60855	Colombia	Unknown	Unknown
<i>P. lutzii</i> *	Pb01	Brasilia, DF, Brazil	Human	Teixeira et al. (2009)
<i>A. dermatitidis</i>	SLH14081	Wisconsin, USA	Human	Unknown
	(ATCC # MYA-2586)	Wisconsin, USA	Environment	Unknown
	ATCC# 26199	NK	Human	Unknown
<i>E. parva</i>	Ep1	IMT**, SP, Brazil	NK	NK

*Also known as Pb01-like

** Instituto de Medicina Tropical. NK: not known

Table 2: Primers used in this work.

Primer	Primer Sequence (5'-3')
PbXhofwd	cccctcgagTTCTGGGAAAGAGCC
PbAgeRev	cccaccggtTTCTTCGAAACCACT
Bd/EpXhoFwd	cccctcgagTTTTGGGAACGTGCC
BdAgeRev	cccaccggtTTCTTCAAACCACT
Bd/EpXhoFwd	cccctcgagTTTTGGGAACGTGCC
EpAgeRev	cccaccggtTTCTTCAAACCACT
DEG-1fwd**	ACIGAYGTIATHCARGC
DEG-2rev**	CKRTTIGGDATYTGRTT
P1-fwd	gggatatcCTTTATTTCAAAGAGATCGACATACTCTAGCG
P1-rev	catggattctcgaaaccactggctcttccagaacaggcctcccatgaaTTAAAACCAAGTCCCCTTAAATAAAGTGTGTCC
P2-fwd	ttcatggaaggcctgttctgggaaagagccagtggttccgaagaatccatgCAATTTAAGAAGCTAACTCATGCTCAAAGAAGTGG
P2-rev	gggatatcCTATACTTTTCATGTATTTTTTCCACAAATGAGCACGG
Plex.fwd	gggatcccgGCCCTTGCCAAGGGGACTCTGTTACTGC
Plex.rev	gggcgccgcGGCGTGCAGCACAAGAAAATCATGACG
PIG451Dex.fwd	GATGGTTGGTACGTTGATTTCAGAAAACACGTTCCGGC
PIG451Dex.rev	ACTGTTCGATCAGTCCTGCGAGGAGGGCAAGTCG
NOT52	gcggccGCTCTGGTGCCACGCGGTAGTTCCGCTCATCACC
BAM52	ggatccCGGGTCCCTGAAAGAGGACTTCAAGTGCACCCTTTTCG

*Degenerated primers, in which, Y stands for C and T, H for A, C and T, R for A and G, K for G and T, D for A, G and T and I for inosine.

Table 3: Nucleotide substitution pattern for the splicing domain of PRP8 *intein* from *P. brasiliensis* (S1, PS2 and PS3), *P. lutzii*, *E. parva*, *H. capsulatum* and *B. dermatitidis*. **S_d** indicates the number of observed synonymous substitutions; **N_d** is the number of observed non-synonymous substitutions; **S** denotes the number of potential synonymous substitutions; **N** is the number of potential non-synonymous substitutions; **p_S** indicates the proportion of observed synonymous substitutions; **p_N** is the proportion of observed non-synonymous substitutions; **d_S** denotes the Jukes-Cantor correction for multiple hits of p_S; **d_N** is the Jukes-Cantor correction for multiple hits of p_N. *The Jukes-Cantor correction (d_S) is not relevant due to the high similarity between the sequences.

Sequence 1	Sequence 2	S _d	N _d	S	N	p _S	p _N	d _S	d _N	d _S /d _N
BdMYA2586	BdSLH14081	1.00	0.00	106.33	343.67	0.01	0.00	0.01	-0.00	*
BdMYA2586	Bd26199	0.00	0.00	106.33	343.67	0.00	0.00	-0.00	-0.00	*
BdMYA2586	HcG186AR	49.50	24.50	104.50	345.50	0.47	0.07	0.75	0.07	10.05
BdMYA2586	Ep	26.75	10.25	106.33	340.67	0.25	0.03	0.31	0.03	9.98
BdMYA2586	Plutzii	66.00	36.00	107.83	342.17	0.61	0.11	1.27	0.11	11.20
BdMYA2586	Pb04	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
BdMYA2586	Pb265	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
BdMYA2586	Bt84	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
BdMYA2586	Pb60855	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
BdSLH14081	Bd26199	1.00	0.00	106.33	343.67	0.01	0.00	0.01	-0.00	*
BdSLH14081	HcG186AR	48.50	24.50	104.50	345.50	0.46	0.07	0.72	0.07	9.71
BdSLH14081	Ep	27.75	10.25	106.33	340.67	0.26	0.03	0.32	0.03	10.44
BdSLH14081	Plutzii	67.00	36.00	107.83	342.17	0.62	0.11	1.32	0.11	11.66
BdSLH14081	Pb04	61.50	37.50	107.00	343.00	0.57	0.11	1.09	0.12	9.23
BdSLH14081	Pb265	61.50	37.50	107.00	343.00	0.57	0.11	1.09	0.12	9.23
BdSLH14081	Bt84	61.50	37.50	107.00	343.00	0.57	0.11	1.09	0.12	9.23
BdSLH14081	Pb60855	61.50	37.50	107.00	343.00	0.57	0.11	1.09	0.12	9.23
Bd26199	HcG186AR	49.50	24.50	104.50	345.50	0.47	0.07	0.75	0.07	10.05
Bd26199	Ep	26.75	10.25	106.33	340.67	0.25	0.03	0.31	0.03	9.98
Bd26199	Plutzii	66.00	36.00	107.83	342.17	0.61	0.11	1.27	0.11	11.20
Bd26199	Pb04	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
Bd26199	Pb265	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
Bd26199	Bt84	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
Bd26199	Pb60855	60.50	37.50	107.00	343.00	0.57	0.11	1.05	0.12	8.90
HcG186AR	Ep	51.50	29.50	104.50	342.50	0.49	0.09	0.80	0.09	8.77
HcG186AR	Plutzii	68.00	44.00	106.00	344.00	0.64	0.13	1.45	0.14	10.34
HcG186AR	Pb04	67.00	47.00	105.17	344.83	0.64	0.14	1.42	0.15	9.44
HcG186AR	Pb265	66.00	47.00	105.17	344.83	0.63	0.14	1.36	0.15	9.04
HcG186AR	Bt84	67.00	47.00	105.17	344.83	0.64	0.14	1.42	0.15	9.44
HcG186AR	Pb60855	66.00	47.00	105.17	344.83	0.63	0.14	1.36	0.15	9.04
Ep	Plutzii	61.38	42.62	107.83	339.17	0.57	0.13	1.07	0.14	7.76
Ep	Pb04	63.62	40.38	107.00	340.00	0.59	0.12	1.18	0.13	9.13
Ep	Pb265	62.62	40.38	107.00	340.00	0.59	0.12	1.14	0.13	8.79
Ep	Bt84	63.62	40.38	107.00	340.00	0.59	0.12	1.18	0.13	9.13
Ep	Pb60855	62.62	40.38	107.00	340.00	0.59	0.12	1.14	0.13	8.79
Plutzii	Pb04	12.00	8.00	117.33	368.67	0.10	0.02	0.11	0.02	4.99
Plutzii	Pb265	11.00	8.00	117.33	368.67	0.09	0.02	0.10	0.02	4.55
Plutzii	Bt84	12.00	8.00	117.33	368.67	0.10	0.02	0.11	0.02	4.99
Plutzii	Pb60855	11.00	8.00	117.33	368.67	0.09	0.02	0.10	0.02	4.55
Pb04	Pb265	1.00	0.00	118.33	373.67	0.01	0.00	0.01	-0.00	*
Pb04	Bt84	0.00	0.00	118.33	373.67	0.00	0.00	-0.00	-0.00	*
Pb04	Pb60855	1.00	0.00	118.33	373.67	0.01	0.00	0.01	-0.00	*
Pb265	Bt84	1.00	0.00	118.33	373.67	0.01	0.00	0.01	-0.00	*
Pb265	Pb60855	0.00	0.00	118.33	373.67	0.00	0.00	-0.00	-0.00	*
Bt84	Pb60855	1.00	0.00	118.33	373.67	0.01	0.00	0.01	-0.00	*
Mean		42.42	25.77	109.01	349.19	0.40	0.08	0.75	0.08	8.92

Table 4: Nucleotide substitution pattern for the endonuclease domain of PRP8 *intein* from *P. brasiliensis* (S1, PS2, PS3), *P. lutzii*, *E. parva*, *H. capsulatum* and *B. dermatitidis*. **S_d** indicates the number of observed synonymous substitutions; **N_d** is the number of observed non-synonymous substitutions; **S** denotes the number of potential synonymous substitutions; **N** is the number of potential non-synonymous substitutions; **p_S** indicates the proportion of observed synonymous substitutions; **p_N** is the proportion of observed non-synonymous substitutions; **d_S** denotes the Jukes-Cantor correction for multiple hits of p_S; **d_N** is the Jukes-Cantor correction for multiple hits of p_N. *The Jukes-Cantor correction (d_S) is not relevant due to the high similarity between the sequences.

Sequence 1	Sequence 2	S _d	N _d	S	N	p _S	p _N	d _S	d _N	d _S /d _N
BdMYA2586	BdSLH14081	2.00	3.00	252.33	794.67	0.01	0.00	0.01	0.00	2.11
BdMYA2586	Bd26199	2.00	0.00	253.00	794.00	0.01	0.00	0.01	-0.00	*
BdMYA2586	HcG186AR	101.62	150.38	241.00	770.00	0.42	0.20	0.62	0.23	2.74
BdMYA2586	Ep	54.62	88.38	244.50	769.50	0.22	0.11	0.27	0.12	2.13
BdMYA2586	Plutzii	133.38	231.62	248.83	783.17	0.54	0.30	0.94	0.38	2.50
BdMYA2586	Pb04	137.62	227.38	248.00	784.00	0.55	0.29	1.01	0.37	2.75
BdMYA2586	Pb265	138.50	227.50	247.67	784.33	0.56	0.29	1.03	0.37	2.80
BdMYA2586	Bt84	137.62	227.38	248.00	784.00	0.55	0.29	1.01	0.37	2.75
BdMYA2586	Pb60855	137.50	227.50	248.00	784.00	0.55	0.29	1.01	0.37	2.75
BdSLH14081	Bd26199	2.00	3.00	252.33	794.67	0.01	0.00	0.01	0.00	2.11
BdSLH14081	HcG186AR	99.62	149.38	240.33	770.67	0.41	0.19	0.60	0.22	2.69
BdSLH14081	Ep	53.62	89.38	243.83	770.17	0.22	0.12	0.26	0.13	2.06
BdSLH14081	Plutzii	131.38	231.62	248.17	783.83	0.53	0.30	0.92	0.38	2.44
BdSLH14081	Pb04	135.62	227.38	247.33	784.67	0.55	0.29	0.99	0.37	2.69
BdSLH14081	Pb265	136.50	227.50	247.00	785.00	0.55	0.29	1.00	0.37	2.73
BdSLH14081	Bt84	135.62	227.38	247.33	784.67	0.55	0.29	0.99	0.37	2.69
BdSLH14081	Pb60855	135.50	227.50	247.33	784.67	0.55	0.29	0.98	0.37	2.68
Bd26199	HcG186AR	101.62	150.38	241.00	770.00	0.42	0.20	0.62	0.23	2.74
Bd26199	Ep	54.62	88.38	244.50	769.50	0.22	0.11	0.27	0.12	2.13
Bd26199	Plutzii	131.38	231.62	248.83	783.17	0.53	0.30	0.91	0.38	2.43
Bd26199	Pb04	135.62	227.38	248.00	784.00	0.55	0.29	0.98	0.37	2.67
Bd26199	Pb265	136.50	227.50	247.67	784.33	0.55	0.29	1.00	0.37	2.71
Bd26199	Bt84	135.62	227.38	248.00	784.00	0.55	0.29	0.98	0.37	2.67
Bd26199	Pb60855	135.50	227.50	248.00	784.00	0.55	0.29	0.98	0.37	2.66
HcG186AR	Ep	110.38	186.62	251.67	795.33	0.44	0.23	0.66	0.28	2.34
HcG186AR	Plutzii	133.62	240.38	253.00	794.00	0.53	0.30	0.91	0.39	2.36
HcG186AR	Pb04	132.75	237.25	251.83	795.17	0.53	0.30	0.91	0.38	2.39
HcG186AR	Pb265	133.50	236.50	251.50	795.50	0.53	0.30	0.92	0.38	2.44
HcG186AR	Bt84	132.75	237.25	251.83	795.17	0.53	0.30	0.91	0.38	2.39
HcG186AR	Pb60855	132.50	236.50	251.83	795.17	0.53	0.30	0.91	0.38	2.39
Ep	Plutzii	135.62	250.38	253.50	787.50	0.54	0.32	0.94	0.41	2.27
Ep	Pb04	137.75	250.25	252.33	788.67	0.55	0.32	0.98	0.41	2.37
Ep	Pb265	138.38	249.62	252.00	789.00	0.55	0.32	0.99	0.41	2.40
Ep	Bt84	137.75	250.25	252.33	788.67	0.55	0.32	0.98	0.41	2.37
Ep	Pb60855	137.38	249.62	252.33	788.67	0.54	0.32	0.97	0.41	2.36
Plutzii	Pb04	18.00	26.00	263.83	834.17	0.07	0.03	0.07	0.03	2.25
Plutzii	Pb265	18.00	24.00	263.50	834.50	0.07	0.03	0.07	0.03	2.44
Plutzii	Bt84	18.00	26.00	263.83	834.17	0.07	0.03	0.07	0.03	2.25
Plutzii	Pb60855	17.00	24.00	263.83	834.17	0.06	0.03	0.07	0.03	2.30
Pb04	Pb265	4.00	2.00	262.00	836.00	0.02	0.00	0.02	0.00	6.44
Pb04	Bt84	0.00	0.00	262.33	835.67	0.00	0.00	-0.00	-0.00	*
Pb04	Pb60855	3.00	2.00	262.33	835.67	0.01	0.00	0.01	0.00	4.81
Pb265	Bt84	4.00	2.00	262.00	836.00	0.02	0.00	0.02	0.00	6.44
Pb265	Pb60855	1.00	0.00	262.00	836.00	0.00	0.00	0.00	-0.00	*
Bt84	Pb60855	3.00	2.00	262.33	835.67	0.01	0.00	0.01	0.00	4.81
Mean		90.09	152.87	251.85	796.22	0.36	0.19	0.62	0.24	2.77

Figure 2

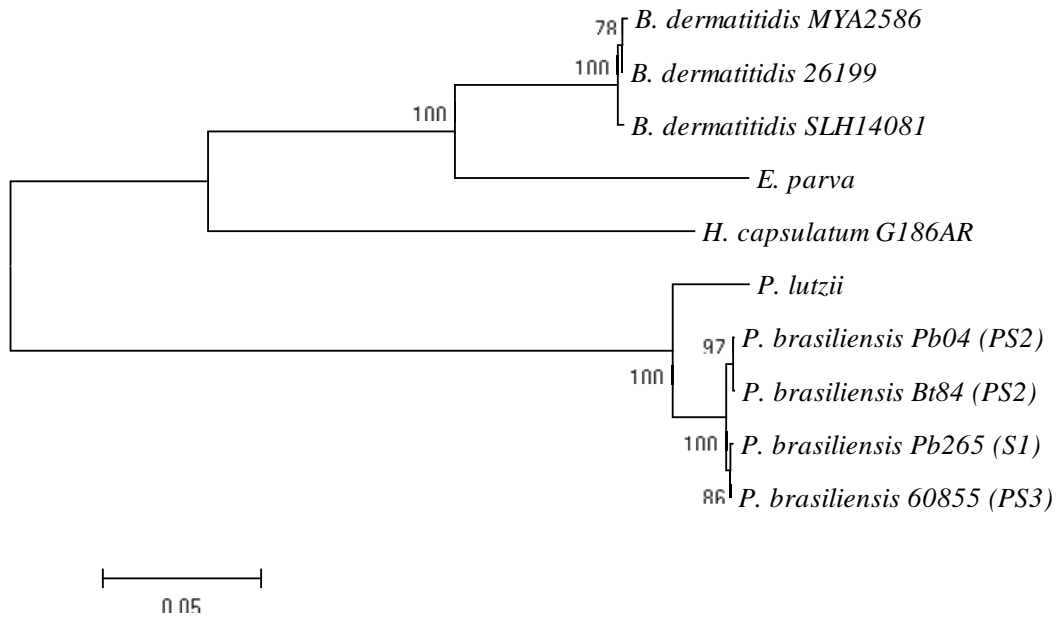


Figure 3

