

JULIANA BENEVENUTO

**EVOLUÇÃO MOLECULAR DO GENE METIONINA SINTASE EM
POPULAÇÕES DISJUNTAS DE *Carapichea ipecacuanha* (RUBIACEAE)**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

**VIÇOSA
MINAS GERAIS – BRASIL
2013**

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

B465e
2013

Benevenuto, Juliana, 1987-

Evolução molecular do gene metionina sintase em população
disjuntas de *Carapichea ipecacuanha* (Rubiaceae) /
Juliana Benevenuto. – Viçosa, MG, 2013.
vii, 33f. : il. ; (algumas color.) ; 29cm.

Orientador: Luiz Orlando de Oliveira.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 31-33.

1. Ipecacuanha - Populações. 2. Diversidade de plantas.
3. Enzimas. 4. Seleção natural. 5. Ipecacuanha - Evolução.
I. Universidade Federal de Viçosa. Departamento de
Bioquímica. Programa de Pós-Graduação em Genética e
Melhoramento. II. Título.

CDD 22. ed. 583.93

JULIANA BENEVENUTO

**EVOLUÇÃO MOLECULAR DO GENE METIONINA SINTASE EM
POPULAÇÕES DISJUNTAS DE *Carapichea ipecacuanha* (RUBIACEAE)**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADA:1º de fevereiro de 2013.

Valéria Monteze Guimarães
(Coorientadora)

Christina Cleo Vinson

Luiz Orlando de Oliveira
(Orientador)

*Aos meus pais, José Geraldo e Marta,
e aos meus irmãos, Ivo e Ivan,
Dedico!*

AGRADECIMENTOS

À Universidade Federal de Viçosa e à todos professores e funcionários que contribuíram para minha formação acadêmica e profissional desde o ensino médio no COLUNI, no curso de graduação em Ciências Biológicas e no mestrado em Genética e Melhoramento.

Ao PPGGM pela oportunidade de desenvolver o mestrado neste programa e ao grupo GenMelhor pelo trabalho em equipe para a ampliação do conhecimento, socialização e melhoria do programa.

Ao CNPq pela concessão de bolsa e, também à CAPES e FAPEMIG, pelo apoio financeiro para o desenvolvimento deste projeto.

Ao professor Luiz Orlando de Oliveira pela orientação, confiança e pelo exemplo de profissionalismo. À você, professor, meu respeito e gratidão por todos ensinamentos durante esses 2 anos.

Aos meus coorientadores, pesquisadora Ana Paula Gomes Soares e professora Valéria Monteze Guimarães, pela amizade e contribuições para o desenvolvimento deste trabalho.

Aos membros de minha banca examinadora, professora Valéria Monteze Guimarães e pesquisadora Christina Cleo Vinson por terem aceitado o convite.

À todos os membros, visitantes e agregados do laboratório de “Biologia Molecular e Filogeografia” (Alberto, Ana Paula, Alejandra, Alex, Amanda, Camila, Cassiana, Christina, Eduardo, Érica, Evan, Flávio, Gislaine, Hugo, Larissa, Leandro, Lígia, Lis, prof. Luiz Orlando, Magali, Natália, Noga, Rafaela, Thaís, Thamyres, prof. Valéria e Vanderson). Com certeza a compreensão com novatos, o auxílio técnico, o convívio agradável, as amizades construídas, as boas risadas, os “bolos & cafezinhos” e as inúmeras confraternizações foram de grande importância para o desenvolvimento deste trabalho. Muitíssimo obrigada! Vou sentir muito a falta de vocês, o melhor lab de todos!

Às amizades construídas em Viçosa (David, Marcela, Gisele, Iara, Lucas, Christiane, Cláudia, Bárbara, Cacá, Haroldo, prof. Lino e petianos) pelo apoio e bons momentos compartilhados. Em especial, agradeço ao Luís Felipe pelo companheirismo, cumplicidade e amor e às amigas da minha república (Lekinha e Ferdinanda) pelo convívio fraterno, paciência, ajuda e jantas de todo dia.

Às minhas amigas de Visconde do Rio Branco (Pozinha, Lolô, Poly, Deinha, Lêh,) pela compreensão com meus finais de semana e feriados comprometidos e pela amizade mantida apesar da distância. Em especial, agradeço à minha grande amiga Fabíola (Bibi) por tudo que vivemos juntas, pela presença em todos momentos importantes da minha vida e que, nesta conquista, infelizmente não estará fisicamente aqui comigo.

À minha família, meus irmãos Ivo e Ivan e meus pais José Geraldo e Marta, pelo amor incondicional, por apoiarem minhas decisões e não medirem esforços para realizar meus sonhos. Essa conquista é para vocês!

À Deus pelo dom da vida, por todas as bênçãos derramadas sobre mim e, principalmente, pela FÉ que fez com que eu não desanimasse nos momentos difíceis ao longo desses dois anos.

BIOGRAFIA

JULIANA BENEVENUTO, filha de José Geraldo Ferraz Benevenuto e Luzia Marta da Silva Benevenuto, nasceu em Visconde do Rio Branco, Minas Gerais, no dia 05 de novembro de 1987.

Em 2002 concluiu o Ensino Fundamental no Colégio Municipal Rio Branco em sua cidade natal. E cursou o ensino médio no Colégio de Aplicação – COLUNI da Universidade Federal de Viçosa no período de 2003 a 2005.

Graduou-se como Bacharel em Ciências Biológicas pela Universidade Federal de Viçosa em janeiro de 2011, recebendo homenagem do Centro de Ciências Biológicas e da Saúde por ter se destacado entre seus pares ao longo do curso. Durante a graduação foi bolsista do Programa de Educação Tutorial (PET-BIO) realizando atividades de ensino, pesquisa e extensão. Também foi bolsista de Iniciação Científica desenvolvendo o projeto “Subsídios biológicos e moleculares para a preservação de *Melipona capixaba* (Hymenoptera: Apidae): desenho de primers microssatélites e análise populacional” sob orientação da professora doutora Mara Garcia Tavares. Ainda durante a graduação, complementou sua formação em um intercâmbio de seis meses no Instituto Politécnico de Bragança – Portugal, onde cursou duas disciplinas e atuou como voluntária no Laboratório de Biologia Molecular de Fungos sob orientação do professor doutor Altino Branco Choupina.

Em março de 2011, iniciou o mestrado *stricto sensu* no Programa de Pós-Graduação em Genética e Melhoramento da Universidade Federal de Viçosa sob orientação do professor doutor Luiz Orlando de Oliveira desenvolvendo a dissertação aqui apresentada.

Em fevereiro de 2013 iniciará o doutorado em “Genética e Melhoramento de Plantas” na Universidade de São Paulo - campus ESALQ - sob orientação da professora doutora Cláudia Barros Monteiro Vitorello.

SUMÁRIO

RESUMO	VI
ABSTRACT	VII
I) INTRODUÇÃO	1
II) MATERIAIS E MÉTODOS	4
1) Material biológico	4
2) Extração de DNA genômico total.....	4
3) Isolamento e sequenciamento do gene <i>MetE</i>	5
4) Alinhamento e Edição das Sequências	6
5) Análises de Diversidade e Testes de Neutralidade com Simulações Coalescentes.....	6
6) Estrutura Populacional de <i>C. ipecacuanha</i>	7
III) RESULTADOS	9
1) Caracterização do gene <i>MetE</i> e identificação de erros gerados pela <i>Taq</i> DNA polimerase	9
2) Análises Populacionais	10
3) Análises evolutivas do gene <i>MetE</i>	17
IV) DISCUSSÃO	26
1) O gene <i>MetE</i> em <i>C. ipecacuanha</i>	26
2) Estrutura populacional	26
3) Diversidade e Desvios da Neutralidade.....	27
4) Evolução molecular da região codificante de <i>MetE</i>	28
V) CONCLUSÕES	30
V) REFERÊNCIAS	31

RESUMO

BENEVENUTO, Juliana. M.Sc., Universidade Federal de Viçosa, fevereiro de 2013. **Evolução molecular do gene metionina sintase em populações disjuntas de *Carapichea ipecacuanha* (Rubiaceae)**. Orientador: Luiz Orlando de Oliveira. Coorientadores: Valéria Monteze Guimarães e Ana Paula Gomes Soares.

Metionina sintase independente de cobalamina (MetE) é uma enzima chave na via de síntese “de novo” de metionina em plantas e também é responsável pela regeneração de S-adenosilmetionina, metabólito associado à vias de interação com patógenos, crescimento vegetal e resposta à estresses. Neste estudo, 95 sequências do gene para metionina sintase foram obtidas a partir de 26 indivíduos de *Carapichea ipecacuanha* (espécie da família Rubiaceae, popularmente conhecida com ipeca) com objetivo de entender e modelar os eventos evolutivos que ocorreram sobre o *locus MetE* e sobre os indivíduos de três regiões disjuntas de ocorrência desta espécie (Mata Atlântica, Amazônia e América Central). Sequências de ITS do nrDNA e dos espaçadores *trnT-trnL* do cpDNA também foram utilizadas para uma análise populacional *multiloci*. A região sequenciada de *MetE* abrangeu 3.890pb, sendo 2.298pb correspondentes à região codificante (composta por 11 exons) e 1.592pb à região não codificante (composta por 10 introns e uma parte da região 3’UTR). No total, foram obtidos 75 haplótipos e uma diversidade nucleotídica de 0,00899. As análises F_{ST} e AMOVA de *MetE* indicaram ausência de estruturação em relação às três regiões, diferentemente do observado para os demais *loci* aqui analisados e em estudos prévios em que elevada divergência entre as três regiões foi observada. Por meio rede de haplótipos foi possível observar a distribuição simpátrica destes; no entanto, o excesso de haplótipos de ponta e o pequeno número de passos mutacionais entre haplótipos vizinhos das três regiões não permitem que a diferenciação destes seja significativa. A árvore filogenética Bayesiana enraizada por *Carapichea lucida* (grupo externo) utilizando a região codificante também mostrou a formação de clados por regiões geográficas e forneceu evidências de que sequências da “Mata Atlântica” sejam mais próximas do estado ancestral do gene *MetE* que as sequências das demais regiões. A heterogeneidade dos resultados encontrados para os testes de neutralidade e para os valores de diversidade nucleotídica e haplotípica nos diferentes *loci* sugerem que estes estão evoluindo independentemente e que eventos demográficos não explicam o padrão obtido. Assim, análises para a presença de seleção foram conduzidas para a região codificante do gene *MetE* e também foi realizado alinhamento de proteínas preditas para esta espécie contra proteínas MetE de outras espécies vegetais disponíveis no *GenBank*. Em quatro sítios houve evidência de seleção positiva, mas os resultados sugerem que a seleção negativa (purificadora) é a principal força atuando sobre o *locus*, favorecendo a conservação da estrutura e função proteica em todas as cópias gênicas de *C. ipecacuanha*.

ABSTRACT

BENEVENUTO, Juliana. M.Sc., Universidade Federal de Viçosa, February, 2013. **Molecular evolution of the methionine synthase gene in disjuncts populations of *Carapichea ipecacuanha* (Rubiaceae)**. Adviser: Luiz Orlando de Oliveira. Co-advisers: Valéria Monteze Guimarães and Ana Paula Gomes Soares.

Cobalamin-independent methionine synthase (MetE) is a key enzyme for *de novo* synthesis of methionine in plants and it is also responsible for the regeneration of S-adenosylmethionine, which is a metabolite associated with the process of interaction with pathogens, plant growth and stresses response. In this study, 95 complete sequences of the gene for methionine synthase were obtained from 26 specimens of *Carapichea ipecacuanha* (species of Rubiaceae family, which is popularly known as ipecac) with aims to understand and model evolutionary events that occurred on *MetE* locus and on the specimens from three disjuncts ranges of occurrence of this species (Atlantic, Amazon and Central America ranges). Sequences of ITS and 5.8S from nrDNA and *trnT-trnL* spacers from cpDNA were also used for a *multiloci* population analysis. The sequenced region of *MetE* covered 3,890pb, where 2,298pb correspond to the coding region (composed of 11 exons) and 1,592pb to the noncoding region (comprising 10 introns and a portion of 3'UTR). In the entire sample were obtained 75 haplotypes and a nucleotide diversity of 0.00899. F_{ST} and AMOVA analyzes of *MetE* indicated no structure in three regions, opposite to detected for other *loci* analyzed here and in previous studies in which high divergence among the specimens from three ranges was observed. However, in the network haplotype was possible to observe the sympatric distribution of the haplotypes. Possibly, excess of tip haplotypes and the small number of mutational steps among neighboring haplotypes from the three ranges do not allow that the differentiation among these was significant. The Bayesian phylogenetic tree rooted by *Carapichea lucida* (outgroup) using the coding region sequences showed the formation of clades by geographic regions and provided evidence that sequences from "Atlantic" range is closer to the ancestral state of the gene *MetE* than sequences from others ranges. The heterogeneity of results for neutrality tests and for values of haplotype and nucleotide diversity at different *loci* suggests that these are evolving independently and also suggests that demographic events do not explain the pattern found. Thereby, tests for the presence of selection site-by-site were performed for the coding region of the gene *MetE* and protein alignment against others MetE from diferents plant species available in the Genbank was develop too. Four sites under positive selection were found, but the results show that negative selection (purifying selection) is the main force acting on the locus, favoring the conservation of protein structure and function in all gene copies from *C. ipecacuanha*

I) INTRODUÇÃO

As plantas, diferentemente de outros eucariotos superiores, possuem todo o aparato enzimático para a síntese “de novo” de metionina. Metionina é um aminoácido que, além da composição de proteínas e iniciação do processo de tradução, participa indiretamente de várias outras vias metabólicas, principalmente por ser o precursor imediato de S-adenosilmetionina (AdoMet). AdoMet participa como doador primário de grupos metil em reações de transmetilação e é o intermediário na biossíntese de poliaminas, etileno, vitamina B1, 3-dimetilsulfonopropionato (um osmoprotetor), além de apresentar funções regulatórias na interação com patógenos, controle do enxofre circulante, resposta a estresse e crescimento vegetal (Ravanel *et al.*, 1998; Hesse & Hoefgen, 2003; Narita *et al.*, 2004). A enzima metionina sintase (EC 2.1.1.14) atua nesta via de biossíntese transferindo um grupo metil do N₅-metiltetrahidrofolato poliglutamato para a homocisteína, formando a metionina. A metionina sintase também atua na reposição de grupos metil após reações de transmetilação de AdoMet, sendo esta a principal rota do metabolismo de metionina. Cerca de 20% da metionina produzida é incorporada a proteínas e 80% é convertida em AdoMet (Hesse & Hoefgen, 2003). Portanto, esta enzima não só catalisa a última reação de síntese de metionina “de novo”, como também serve para regenerar o grupo metil do AdoMet (Figura 1). Dois tipos de metionina sintase já foram descritas, uma dependente de co-fator cobalamina para sua atividade (MetH) e outra cobalamina independente (MetE), sendo somente observado em plantas o tipo independente de cobalamina.

Como a enzima metionina sintase (MetE) está indiretamente envolvida em mecanismos de interação com patógenos, o gene que a codifica é especulado como de rápida evolução nas populações. As interações patógeno-hospedeiro podem ser uma poderosa força para a evolução adaptativa, moldando os padrões de variação molecular dos genes envolvidos (Chiang *et al.*, 2002). Assim, este gene pode ser um modelo ideal para ser utilizado em estudos de variações genéticas recentes associadas à seleção natural, história populacional e demografia. Além disso, dados de polimorfismos de DNA são não-ambíguos, irreduzíveis e representam informações genéticas completas (Wright & Gaut, 2005). O uso deste tipo de dados, aliado à teoria do coalescente e da evolução neutra, fornecem *insights* quanto à frequência e às assinaturas moleculares da presença de seleção.

Na região codificante de um gene, as mutações podem ser sinônimas (não alterando o aminoácido codificado) ou não-sinônimas (substituindo um aminoácido por outro). A estimativa da taxa de mutações sinônimas (dS) em relação a não-sinônimas (dN) constitui uma ferramenta amplamente utilizada para compreender a dinâmica da evolução molecular de uma sequência, pois enquanto mutações sinônimas são “invisíveis” à seleção natural e podem ser consideradas próximas da neutralidade, mutações não-sinônimas podem estar sob forte pressão seletiva (Kreitman, 2000). Se uma sequência está evoluindo de forma neutra, mutações sinônimas e não-sinônimas são esperadas em igual proporção ($dN/dS \cong 1$) e a sequência pode ser um pseudogene. Se a proporção de substituição dN/dS for baixa ($dN/dS < 1,0$) é provável que esteja ocorrendo seleção negativa (purificadora), sendo as mutações que alteram o aminoácido codificado eliminadas da população e a estrutura primária da proteína permanece conservada, porém isso não exclui a possibilidade de seleção positiva. E, se a proporção dN/dS for muito maior que 1 ($dN/dS > 1$), há uma forte evidência para presença de seleção positiva atuando a favor da proteína modificada, podendo levar a uma diversificação da função protéica, o que também não implica que a seleção negativa não esteja atuando. No entanto, outras forças evolutivas, tais como eventos demográficos, migração, recombinação e sistemas de acasalamento, podem mascarar os efeitos da seleção natural (Kreitman, 2000), sendo necessária a aplicação de várias análises estatísticas e comparações entre diferentes *loci* para distinguir quais forças estão atuando e sobre quais regiões do genoma.

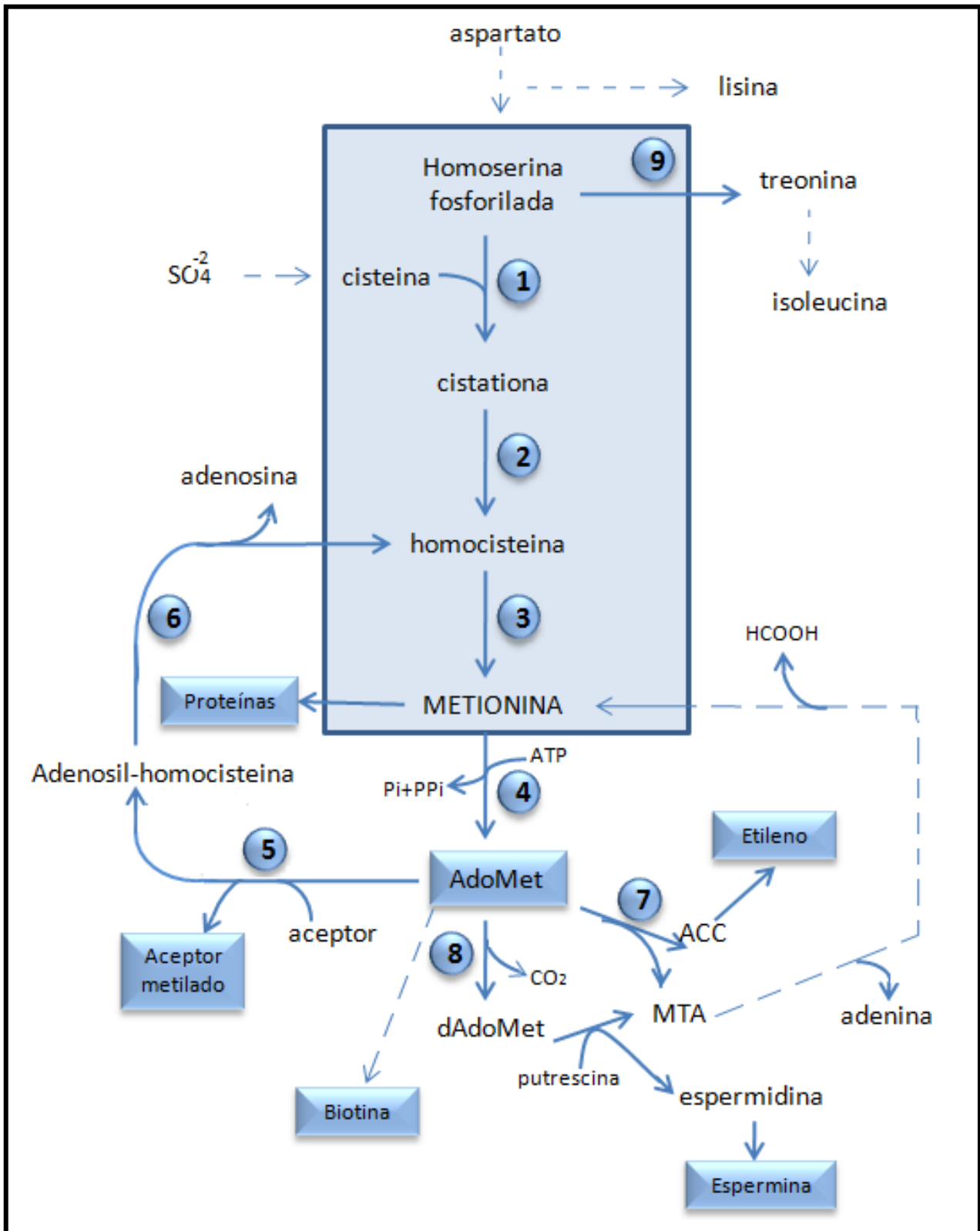


Figura 1: Relação entre as vias biossintéticas de metionina, outros aminoácidos, AdoMet, poliaminas, biotina, espermina e etileno em plantas superiores. Enzimas: 1) Cistationina γ -sintase; 2) Cistationina B-liase; 3) Metionina sintase; 4) AdoMet sintase; 5) Metilase AdoMet-dependente; 6) Adenosil-homocisteína hidrolase; 7) Ácido 1-aminociclopropano-1-carboxílico sintase; 8) AdoMet decarboxilase; 9) Treonina sintase. Adaptado de Ravel *et al.* (1998).

O cDNA e a estrutura proteica primária predita de MetE já foram descritos e disponibilizados no *GenBank* para diversas espécies vegetais, tais como, *Catharanthus roseus* (X83499), *Nicotiana suaveolens* (DQ984138), *Solanum tuberosum* (AF082893), *Carica papaya* (EF512313), *Camellia sinensis* (JF795472), *Glycine max* (NM_001248865), *Phelipanche ramosa* (DQ849630) e *Arabidopsis thaliana* (U97200). Recentemente, Batista (2010) isolou e caracterizou o cDNA de MetE para a espécie *Carapichea ipecacuanha*.

A espécie *C. ipecacuanha*, popularmente conhecida como ipeca ou poaia, é uma planta da família Rubiaceae, mundialmente conhecida pelas propriedades medicinais do extrato de suas raízes (Skorupa & Assis, 1998; WHO, 2007; Brandão, *et al.*, 2008). Populações naturais de ipeca são encontradas em três grandes regiões geográficas bem definidas e disjuntas: 1) América Central em continuidade com o norte da América do Sul; 2) sul da floresta estacional semi-decídua do bioma Amazônia; 3) floresta estacional semi-decídua do bioma Mata Atlântica. Durante o período em que foi descoberta (século XVII) até a década de 80, o Brasil explorou intensivamente o processo extrativista em populações naturais de ipeca (Skorupa & Assis, 1998). Essa intensiva coleta de plantas aliada à progressiva devastação das florestas nas zonas de ocorrência natural, ao uso das áreas produtoras para abertura de novas fronteiras agrícolas e à expansão das cidades e rodovias, resultou em uma gradual redução e fragmentação de suas populações, sendo a espécie considerada “ameaçada de extinção” pelo Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis – IBAMA (Oliveira & Martins, 2002). A reprodução principalmente por via assexuada, a limitada dispersão entre populações, o intenso processo extrativista e a fragmentação dos habitats de *C. ipecacuanha* conduziram à perda de elementos genéticos por processos estocásticos, cujos efeitos foram refletidos no padrão de diversidade genética da espécie. Estudos de filogeografia e de diversidade genética de suas populações revelaram que a variabilidade genética dentro de uma mesma região é baixa e que não há fluxo gênico entre elas devido às limitações inerentes à própria reprodução da espécie e às barreiras geográficas de longa data, formando populações altamente divergentes e estruturadas nas três grandes regiões de ocorrência natural (Rossi *et al.*, 2009; Oliveira *et al.*, 2010; Queiroz *et al.*, 2011). Desta forma, a espécie *C. ipecacuanha* foi escolhida para a condução das análises do gene *MetE*, uma vez que o uso de indivíduos provenientes de populações cuja filogeografia já está bem definida, como a de ipeca, constitui uma excelente estratégia para estudos de evolução gênica molecular intraespecífica.

Com base nestas informações, sequências completas do gene para metionina sintase foram obtidas para entender e modelar os eventos evolutivos que ocorreram sobre o *locus* e sobre as populações disjuntas de *C. ipecacuanha*. As sequências das regiões do DNA ribossomal nuclear (ITS1–5.8S–ITS2 do nrDNA) e de espaçadores intergênicos do DNA cloroplastídico (*trnT-trnL* do cpDNA), obtidas do trabalho de Oliveira *et al.* (2010), também foram utilizadas para fornecer uma análise *multiloci*, o que torna as conclusões mais informativas e confiáveis em estudos de genética populacional. Por conseguinte, o objetivo do presente estudo é responder às questões: 1) Como é a organização estrutural do gene *MetE* no genoma de ipeca?; 2) Como a diversidade está distribuída ao longo da sequência e das populações?; 3) A variabilidade genética de *MetE* está particionada entre as três grandes regiões e produz padrões de estruturação consistentes como em estudos prévios com outras regiões do genoma?; 4) Há eventos demográficos ou de seleção atuando sobre o gene causando um desvio do esperado por evolução neutra?; 5) As cópias gênicas de *MetE* estão seguindo diferentes destinos evolutivos, uma vez que pressões de seleção diferentes entre os habitats e efeitos estocásticos podem estar atuando sobre elas?

II) MATERIAIS E MÉTODOS

1) Material biológico

Com base em estudos prévios, que indicam alta divergência genética entre indivíduos oriundos das três regiões de ocorrência de ipeca e baixa variabilidade genética intra-regional (Rossi *et al.*, 2009; Oliveira *et al.*, 2010; Queiroz *et al.*, 2011), uma grande amostragem dentro da mesma região torna-se desnecessária. Portanto, no presente trabalho, foram avaliados apenas 26 indivíduos de *C. ipecacuanha* amostrados nas três regiões disjuntas de ocorrência da espécie (Figura 1), sendo 13 oriundos da região de Mata Atlântica, 09 da região Amazônica e 04 da região do Panamá na América Central. A espécie *Carapichea lucida* foi utilizada como grupo externo (Tabela 1).

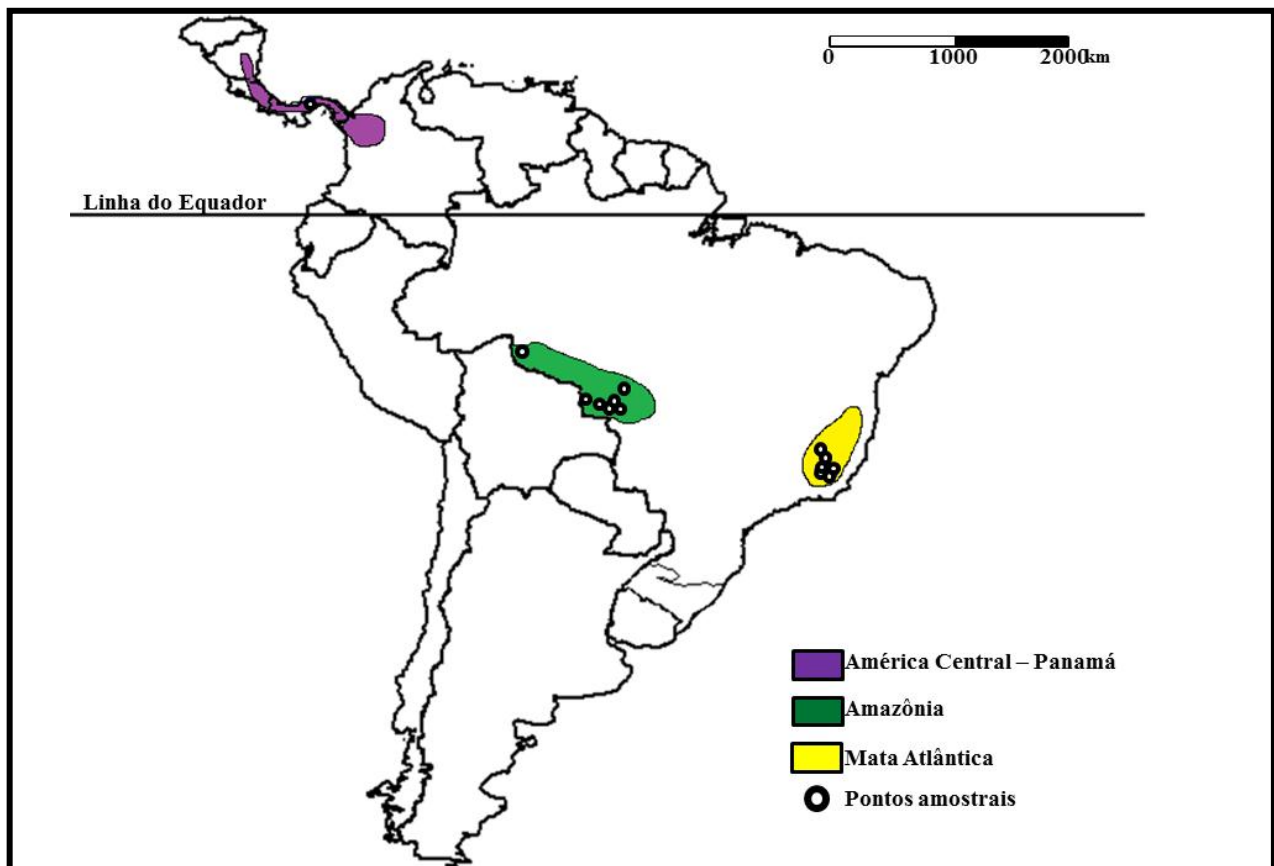


Figura 1: Distribuição geográfica aproximada dos indivíduos amostrados (nomes e coordenadas de cada indivíduo estão disponíveis na Tabela 1) e das três grandes regiões disjuntas de ocorrência de *Carapichea ipecacuanha* representadas pelas diferentes cores.

2) Extração de DNA genômico total

A extração de DNA genômico foi realizada conforme o protocolo descrito em Rossi *et al.* (2009). O DNA concentrado está armazenado em freezer (-20 °C) no Laboratório de Biologia Molecular e Filogeografia da Universidade Federal de Viçosa.

Tabela 1: Localidades e coordenadas geográficas dos indivíduos amostrados para análise do gene *MetE*. Na tabela, **N** corresponde ao tamanho amostral e **n** ao número de clones sequenciados.

Regiões (localidade/ abreviatura)	Coordenadas		n
	Latitude	Longitude	
<i>Região de Mata Atlântica (N=13)</i>			48
Dores de Guanhões (DOG1)	S19°02'17''	W42°57'27''	3
Faria Lemos (FRL2)	S20°48'41''	W42°01'49''	5
Guaraciaba (GUA5)	S20°34'14''	W43°00'28''	2
Guaraciaba (GUA10)*	S20°34'14''	W43°00'28''	1
Ponte Nova (PTN1)	S20°25'26''	W42°53'42''	4
Raposo (RAP5)	S21°06'40''	W42°05'43''	6
Raposo (RAP6)*	S21°06'40''	W42°05'43''	3
Raposo (RAP18)	S21°06'40''	W42°05'43''	5
Trilha da Lagoa do Meio (TLM3)*	S19°38'34''	W42°30'56''	1
Trilha da Lagoa do Meio (TLM6)	S19°38'34''	W42°30'56''	4
Trilha da Lagoa do Meio (TLM11)	S19°38'34''	W42°30'56''	5
Visconde do Rio Branco (VRB5)	S21°00'37''	W42°50'26''	5
Visconde do Rio Branco (VRB9)	S21°00'37''	W42°50'26''	4
<i>Região Amazônica (N=09)</i>			36
Colorado (COL6)	S11°14'49''	W64°24'40''	4
Exu (EXU2)	S15°40'41''	W57°32'03''	4
Figueirópolis d'Oeste (FIG1)	S15°26'37''	W58°47'06''	4
Lambari d'Oeste (LAM4)	S15°19'27''	W58°00'10''	3
Mozar (MOZ1)	S15°04'57''	W57°58'00''	4
Prata (PRA4)	S15°30'26''	W58°01'50''	4
Tangará da Serra (TAN3)	S14°04'38''	W57°03'45''	3
Tangará da Serra (TAN9)	S14°04'38''	W57°03'45''	6
Vila B. S. Trindade (VBS2)	S15°00'29''	W59°57'02''	4
<i>Região América Central (N=04)</i>			11
Panamá (DRA6)*	N09°09'25''	W79°51'16''	2
Panamá (MIL1)*	N09°09'25''	W79°51'16''	2
Panamá (TCS2)*	N09°09'25''	W79°51'16''	3
Panamá (ZET5)*	N09°09'25''	W79°51'16''	4
<i>Grupo Externo (N=01)</i>			1
<i>Carapichea lucida*</i>			1
TOTAL (N=25)			96

* Amostras cujo gene *MetE* foi particionado para amplificação

3) Isolamento e sequenciamento do gene *MetE*

Primers de anelamento específico às regiões internas e às extremidades do gene *MetE* foram desenhados baseados na sequência do cDNA de *C. ipecacunha* isolado por Batista (2010) (ver Tabela 2). Para a amplificação do gene completo de aproximadamente 4kb (que abrange desde o códon de iniciação da tradução – ATG – até a região 3'UTR), foram utilizados os *primers* met_ATG_F e met_polA_R nas concentrações e condições de amplificação recomendadas no protocolo da GoTaq® DNA polimerase (Promega, USA).

Para amostras de difícil amplificação, o gene foi dividido em dois fragmentos. O fragmento inicial do gene foi obtido com os *primers* met_ATG_F e met_exon7_R e possui aproximadamente 2,4Kb. O segundo fragmento, localizado na porção terminal do gene, foi obtido com os *primers* met_exon6_F e met_polA_R e possui aproximadamente 2Kb. Entre estes dois fragmentos, há uma região de sobreposição de aproximadamente 250pb que foi utilizada para uni-los.

Os produtos da PCR foram clonados em vetor p-GEM-Teasy (Promega, USA) e sequenciados pela empresa Macrogen Inc. (Coréia do Sul). Mais de um clone por indivíduo foram sequenciados para detectar a presença de múltiplos haplótipos e controlar erros de fidelidade da *Taq* DNA polimerase. Para o sequenciamento do gene completo foram necessários 05 *primers* por clone (met_exon4_R, met_exon6_F, met_exon7_R e os *primers* universais M13F e M13R).

Após o sequenciamento e alinhamento, foi observada a presença de um *indel* (inserção/deleção) de 120pb no segundo intron do gene em alguns clones. Então, os *primers* met_intron2_F e met_intron2_R foram desenhados para fazer um *screening* nas populações cujos DNAs dos indivíduos já estavam disponíveis em nosso laboratório. O polimorfismo devido ao *indel* foi detectado em gel de agarose 2%.

Tabela 2: *Primers* utilizados neste trabalho e suas respectivas sequências, temperaturas de *melting* e uso. A sigla PCR significa *primers* utilizados na amplificação do gene; SEQ significa uso para sequenciamento e SCR significa uso para *screening* das populações.

<i>Nome</i>	<i>Sequência (5'-3')</i>	<i>Tm (°C)</i>	<i>Uso</i>
met_ATG_F	ATGGCTTCCCACATTGTTGG	56,0	PCR
met_polA_R	CCCTCAAGATGCACACAGAA	60,0	PCR
met_exon4_R	TTGGCCCAGATATTCCTTCC	61,0	SEQ
met_exon6_F	TCTCCCAGGGTAACAAATGA	58,0	PCR, SEQ
met_exon7_R	GATCCAATGGTGGTGGTAGG	60,0	PCR, SEQ
M13F (universal)	GTAAAACGACGGCAGT	57,2	SEQ
M13R (universal)	GCGGATAACAATTCACACAGG	60,8	SEQ
met_intron2_F	AATACAAAGAGGCCAAGGCG	60,4	SCR
met_intron2_R	GGAACACTGTCAACGCCAAG	62,4	SCR

4) Alinhamento e Edição das Sequências

As sequências foram alinhadas utilizando o programa Sequencher 4.8 (Gene Codes Corporation) com edições e refinamentos manuais. As substituições nucleotídicas únicas por sítio foram consideradas erros de fidelidade da *Taq* DNA polimerase e regiões de introns com repetições mononucleotídicas variáveis também foram excluídas das análises.

5) Análises de Diversidade e Testes de Neutralidade com Simulações Coalescentes

Para avaliar o efeito de forças evolutivas internas, genes não ligados ao locus *MetE*, como os espaçadores de DNA ribossomal nuclear (ITS) e espaçadores intergênicos do DNA cloroplastático (*trnT-trnL*) também foram analisados.

O programa DNAsp v5 (Librado & Rozas, 2009) foi utilizado para estimar parâmetros genéticos populacionais e tendências do espectro de frequência alélica. As diversidades haplotípicas (*h*) e nucleotídicas (π) foram calculadas para cada uma das diferentes regiões e para as amostras como um todo. Para o cálculo destes índices, os *indels* não foram considerados. Para visualizar a conservação e estimar a taxa de diversidade nucleotídica (π) ao longo da sequência de

MetE, uma análise em janela deslizante foi realizada com base em uma janela de 50pb que se move através do gene em intervalos de 25pb.

A história demográfica de populações e a presença de seleção natural são eventos não aleatórios que podem modificar o padrão de diversidade nucleotídica e de distribuição de polimorfismos esperado sob o modelo Neutralista de Evolução. Para inferir sobre a ocorrência destes eventos em relação ao gene *MetE*, foram utilizados os testes D de Tajima (1989), F de Fu e Li (1993), H de Fay e Wu (2000) que são testes complementares para se obter uma conclusão mais confiável sobre quais eventos evolutivos estão atuando sobre as cópias gênicas de *MetE* nas populações de ipeca. O teste D de Tajima é mais sensível aos desvios no espectro das frequências de polimorfismos por ser baseado nas diferenças entre o número de sítios segregantes e o número médio de diferenças nucleotídicas estimado a partir da comparação entre pares de sequências. O teste F de Fu e Li é mais sensível a substituições nucleotídicas únicas e baseia-se no número de mutações que aparecem uma vez entre as sequências e no número médio de diferenças nucleotídicas entre pares de sequências. Já o teste H de Fay e Wu é mais sensível a um excesso de variantes derivados em alta frequência e, portanto, requer um grupo externo para a condução das análises. O estado ancestral dos nucleotídeos foi inferido pelo uso de uma única sequência do gene *MetE* obtida de *C. lucida* para ser usado como grupo externo. Todos estes testes (D, F e H) são esperados serem próximo de zero no modelo padrão de neutralidade e foram realizados considerando o número total de mutações e ausência de recombinação, o que pode tornar os testes mais conservadores. Simulações coalescentes com 1.000 repetições foram realizadas para cada um destes testes para modelar o impacto de eventos demográficos sobre a diversidade das sequências. Um intervalo de distribuição (com 95% de confiança) e o p-valor dos parâmetros foram obtidos após as simulações.

6) Estrutura Populacional de *C. ipecacuanha*

A análise de variância molecular (AMOVA) e a estatística F_{ST} de Wright foram realizadas com o intuito de verificar a partição da variação genética entre e dentro de regiões e o grau de diferenciação das populações gerado pelo gene *MetE*. Para isso, o programa Arlequin v 3.1 (Excoffier et al., 2005) foi utilizado.

Com base na Teoria do Coalescente, a rede de haplótipos nos permite reconstruir a relação de ancestralidade entre as sequências e também observar a estruturação geográfica (Posada & Crandall, 2001). A rede de haplótipos foi calculada utilizando o algoritmo *median joining* implementado no programa NETWORK v4.6 (Fluxus Technology). Para esta análise, além dos quatro tipos de desoxirribonucleotídeos que compõe o DNA, os *indels* foram considerados como o quinto caráter de estado e codificados de forma que cada *indel*, independente de seu tamanho (pb), fosse considerado como um único evento de mutação.

Análises filogenéticas do gene e da região codificante de *MetE* foram realizadas utilizando inferência bayesiana, em que métodos de simulação Monte Carlo em Cadeias de Markov (MCMC) estimam a distribuição a posteriori dos parâmetros dos modelos, gerando uma árvore filogenética enraizada pelo gene ortólogo de *C. lucida*. Primeiramente, programa MrModeltest v2.3 (Nylander, 2004) foi utilizado para testar, dentre 24 modelos de substituição nucleotídica, qual melhor se ajusta aos nossos dados. Como modelos distintos foram obtidos para a região codificante e não codificante, a análise do gene completo foi realizada de forma particionada. As duas regiões (codificante e não codificante) foram analisadas simultaneamente, porém cada uma sob o melhor modelo indicado pelo MrModeltest. O programa MrBayes v3.2 (Ronquist *et al.*, 2012), com mínimo de 1.000.000 de gerações, amostradas a cada 100 gerações (sendo as primeiras 250.000 gerações descartadas e definidas como *burn-in*), foi, então, utilizado para fazer as inferências filogenéticas Bayesianas.

7) Evolução molecular de *MetE*

Para explorar a história evolutiva do gene *MetE*, somente o CDS (*Coding DNA Sequence*) foi considerado nas análises subsequentes. As análises para detecção de sítios sob seleção, baseadas nas taxas dN/dS, foram conduzidas no pacote de programas HYPHY implementado no servidor web DataMonkey (Delpont *et al.*, 2010 - disponível em www.datamonkey.org).

Antes de fazer inferências filogenéticas e evolutivas é necessário detectar a presença de eventos de recombinação, pois a evolução em sequências recombinantes não pode ser adequadamente explicada por uma árvore filogenética única e várias filogenias são requeridas para modelar corretamente a evolução de cada fragmento não recombinante (Pond *et al.*, 2006). Assim, se a recombinação não for levada em consideração, os métodos de máxima verossimilhança de substituição de códons para identificar a presença de seleção podem ser erroneamente interpretados devido à presença de muitos sítios falsos positivos. Portanto, uma triagem deste parâmetro foi realizada para os alinhamentos múltiplos do gene completo e CDS de *MetE* utilizando o algoritmo GARD (*Genetic Algorithm Recombination Detection*) para evidenciar incongruências filogenéticas, identificar o número e localização de pontos de quebra e as sequências envolvidas em putativos eventos de recombinação. As análises foram conduzidas utilizando o modelo HKY85 e demais parâmetros por *default*.

Após as análises de recombinação, três algoritmos (SLAC, REL, FUBAR) implementados no programa HYPHY (servidor DataMonkey) foram utilizados para detecção de seleção sítio-por-sítio. O algoritmo SLAC (*Single-Likelihood Ancestor Counting*) é um conjunto de rápidos "métodos de contagem" baseado em máxima verossimilhança que permite a estimação do número de substituições sinônimas e não sinônimas que ocorreram em cada códon ao longo da história evolutiva da amostra. O método realiza a reconstrução de sequências ancestrais e não faz nenhuma suposição sobre a distribuição das taxas sinônimas e não sinônimas dos sítios (Pond & Frost, 2005). A abordagem REL (*Random Effects Likelihood*), uma extensão do método originalmente descrito por Nielsen e Yang em 1998, permite variações em ambas as taxas de substituição (sinônimas e não sinônimas) de acordo com um modelo geral de substituição nucleotídica pré-definido e, em seguida, infere sobre as taxas de variação com que cada sítio individual evolui. A pressão de seleção em sítios individuais é inferida usando abordagens Bayesianas empíricas (Pond & Frost, 2005). FUBAR (*Fast unconstrained Bayesian approximation*) é um método que utiliza uma rápida simulação MCMC que permite uma distribuição flexível sobre os parâmetros de seleção a serem obtidos a partir dos dados, sem restrições paramétricas na forma desta distribuição. Isto permite que as informações sejam compartilhadas entre sítios, gerando maior poder de detecção que SLAC e sem potencial viés como o apresentado pelas distribuições excessivamente restritivas utilizadas por REL. Neste método, a forma de distribuição das taxas sinônimas e não sinônimas pode ser inferida, usando MCMC, sem ter que recalculá-la a função de probabilidade de cada iteração (Murrell, 2012).

Identificados os sítios sob seleção, o modelo de estrutura tridimensional da proteína metionina sintase descrito por Ferrer *et al.* (2004) para *Arabidopsis thaliana* (código Protein Data Bank: 1U1H), o qual contém, em geral, 85% de identidade com a proteína de ipeca, foi utilizado para localizar os sítios e inferir sobre possíveis diversificações funcionais da proteína de acordo com o domínio proteico em que ocorrem. As porções N-terminal (resíduos 1-391) e C-terminal (resíduos 392 – 765) de *MetE* também foram definidas no trabalho de Ferrer *et al.* (2004). Foi feito o alinhamento múltiplo da proteína *MetE* descrita para várias espécies vegetais (disponíveis no *GenBank*) para observar a conservação e/ou perda de aminoácidos associados à função proteica utilizando o programa MEGA v5.1 (Tamura *et al.*, 2011).

III) RESULTADOS

1) Caracterização do gene *MetE* e identificação de erros gerados pela *Taq* DNA polimerase

Após a clonagem e sequenciamento, foram obtidos, no total, 95 clones do gene *MetE* a partir de 26 indivíduos de *C. ipecacuanha*, que continham um fragmento de aproximadamente 4.001pb. Após remoção de regiões mononucleotídicas de tamanho variável entre os clones, obteve-se um *contig* de 3.890pb. O número e tamanho putativos dos exons foi definido por Batista (2010) pela comparação do cDNA de metionina sintase isolado de *C. ipecacuanha* com o seu homólogo em *A. thaliana*. Com a obtenção do gene *MetE* completo neste trabalho, foi possível também definir as regiões de introns, conforme apresentado na Figura 3.

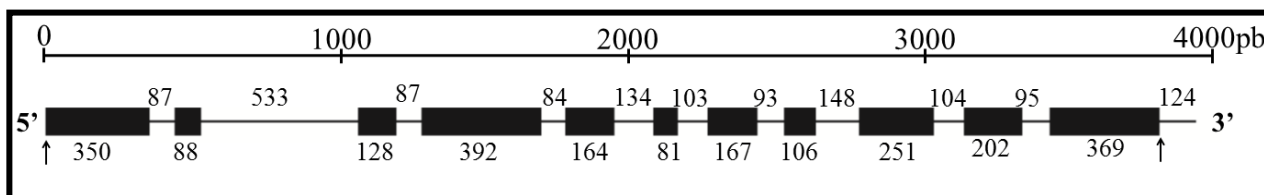


Figura 3: Organização estrutural do gene *MetE* resultante do alinhamento de 95 clones obtidos a partir de 26 indivíduos de *C. ipecacunha*. Os retângulos representam a região codificante (exons) e as linhas, a região não codificante (introns + 3'UTR). Os números posicionados abaixo e acima dos segmentos, correspondem aos seus respectivos tamanhos em pares de base (pb) no *contig*. As setas indicam os códons de iniciação (ATG) e parada (TGA), respectivamente.

O CDS de *MetE* em ipeca é composto por 11 exons, totalizando 2.298pb. Após o alinhamento dos clones para esta região, nenhum *indel* ou mutação sem sentido (que gera um *stop* códon) foram observados ao longo da sequência, gerando, portanto, uma proteína de 765 aminoácidos para todos os clones. O gene é composto por 10 introns, que totalizam 1.468pb. Fez parte dos clones um segmento da região 3'UTR do gene constituída de 124pb, obtendo, então, uma região não codificante total de 1.592pb. Na região não codificante, foram observadas variações de tamanho inter e intraindividuais, sendo notável o *indel* de 120pb no segundo intron. Para detectar a presença deste *indel* em outros indivíduos, foi feito um *screening* em todos DNAs de ipeca disponíveis em nosso laboratório. Somente nos indivíduos da região “Mata Atlântica” foi encontrado este *indel* (Figura 4).

O alinhamento do gene completo revelou, no total, 539 substituições nucleotídicas únicas. A aleatoriedade da ocorrência destas substituições únicas pode ser observada quando comparamos a sua proporção em exons e introns. A região de exons (2.298pb) corresponde a 59,07% do total de pares de base do gene (3.890pb) e o número de substituições únicas observadas nesta região corresponderam a 59,00% do total destas mutações (318 de 539). E para a região não codificante, que corresponde a 40,93% do tamanho do gene, foram encontradas 221 mutações únicas (41,00% do número total destas mutações). Esta aleatoriedade também foi observada quando as mutações únicas encontradas no CDS foram analisadas em relação às três bases do códon, as quais recaíram em cerca de 1/3 sobre cada posição do códon. Assim, todas essas substituições nucleotídicas únicas foram consideradas erros de fidelidade da *Taq* DNA polimerase e removidas das análises subsequentes.

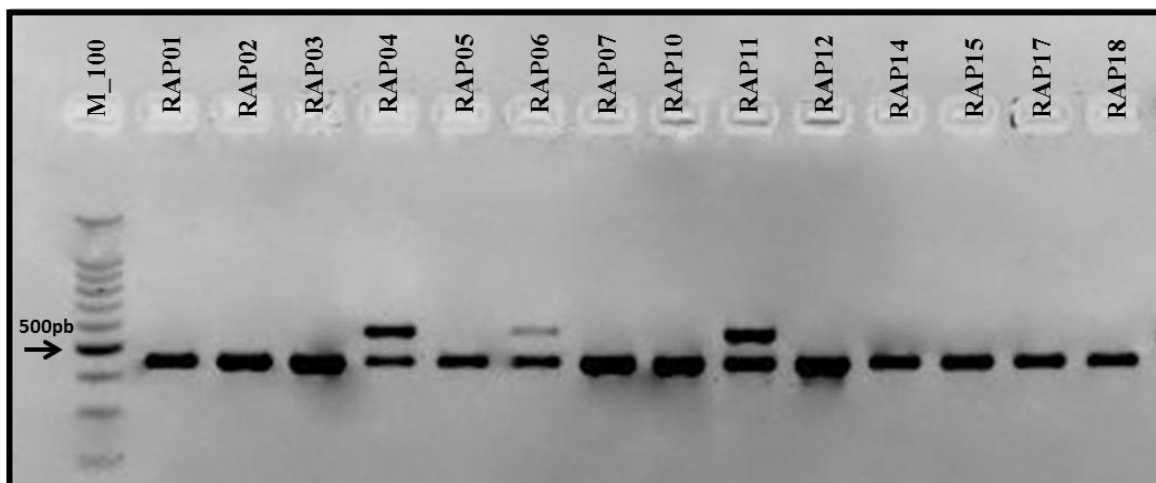


Figura 4: Screening em gel de agarose 2% do *indel* de 120 pb presente no intron2 do gene *MetE* entre indivíduos da população de Raposo (RAP). Indivíduos que apresentam duas bandas são polimórficos para este *indel*. Enquanto indivíduos com uma única banda não apresentam a inserção.

2) Análises Populacionais

Na tabela 3 estão descritas estatísticas das variações nucleotídicas intra-regionais e de todas as amostras, analisadas para diferentes partes do genoma desta espécie. No total, ocorreram 213 sítios polimórficos, sendo 154 substituições do tipo transição (72,30%) e 59 transversões (27,70%).

Tabela 3: Estatísticas descritivas para as variações nucleotídicas e haplotípicas encontradas em todas as amostras e dentro de cada grande região de ocorrência de *C. ipecacuanha* para o gene *MetE*, ITS do nrDNA e espaçadores *trnT-trnL* do cpDNA (sequências obtidas no trabalho de Oliveira *et al.*, 2010).

Região	N	n	S	indel	h	Hd	π_{TOTAL}	D_T	F_{FL}	H_{FW}	F_{ST}
<i>MetE</i> (3890pb)											
Mata Atlântica	13	48	73	5	40	0,991	0,00285	-1,24	-1,75	3,96	-
Amazônia	9	36	54	0	27	0,965	0,00140	-2,17*	-2,17**	0,64	-
América Central	4	11	63	9	8	0,927	0,00760	1,50	1,35	-7,34	-
Total	26	95	213	17	75	0,992	0,00899	-0,66	-0,46	7,33	0.02961
<i>ITS</i> (532pb)											
Mata Atlântica	21	41	173	0	20	0,801	0,02211	-2,17 **	-5,75**	-	-
Amazônia	19	19	0	0	1	0,000	0,00000	-	-	-	-
América Central	6	6	0	1	1	0,000	0,00000	-	-	-	-
Total	46	66	182	1	22	0,837	0,02060	-2,51**	-6,39**	-	0.54882
<i>cpDNA</i> (1044pb)											
Mata Atlântica	73	73	5	5	6	0,617	0,00111	-0,54	0,64	-	-
Amazônia	41	41	0	0	1	0,000	0,00000	-	-	-	-
América Central	5	5	0	0	1	0,000	0,00000	-	-	-	-
Total	119	119	15	13	7	0,726	0,00556	0,91	1,59	-	0.66685

N, número de indivíduos amostrados;

n, número total de sequências amostradas;

S, número de sítios polimórficos;

indel, número de inserções e/ou deleções;

h, número de haplótipos;

Hd, diversidade haplotípica (indels não considerados);

π_{TOTAL} , diversidade nucleotídica total;

D_T , teste D de Tajima sem grupo externo (* $P < 0,05$ e ** $P < 0,02$);

F_{FL} , teste F de Fu e Li sem grupo externo(* $P < 0,05$ e ** $P < 0,02$);

H_{FW} , teste H de Fay e Wu com grupo externo;

F_{ST} , índice de divergência genética entre as regiões.

Quando a sequência completa de *MetE* foi analisada para os 95 clones, 75 haplótipos (*h*) foram encontrados a partir das 213 substituições (*indels* não considerados), o que gerou uma diversidade haplotípica (*Hd*) de 0,992. A região “Mata Atlântica” apresentou o maior índice de diversidade haplotípica (0,991), porém “Amazônia” e “América Central” também apresentaram índices elevados (*Hd* = 0,965 e *Hd* = 0,927, respectivamente). Para os outros loci gênicos analisados (ITS e cpDNA), somente “Mata Atlântica” apresentou diversidade em seus haplótipos.

A diversidade nucleotídica (π_{TOTAL}) de *MetE* para todas as amostras foi de 0,00899. Quando avaliada dentro de cada região, a diversidade nucleotídica apresentou o maior índice na “América Central” (0,00760). E, assim como a diversidade haplotípica, este padrão de diversidade nucleotídica (total e dentro das três regiões) não é observado para os outros loci do genoma avaliados. Para ITS, valores maiores de diversidade foram observados nas amostras (π_{TOTAL} = 0,02060) e somente dentro de “Mata Atlântica” houve variação entre os indivíduos. Para cpDNA, valores baixos de diversidade também foram encontrados (π_{TOTAL} = 0,00556), principalmente porque *indels* foram o principal tipo de mutação nesta região e estes não são computados no cálculo de diversidade nucleotídica. Além disso, variações nos espaçadores de cpDNA também só foram encontradas dentro da região “Mata Atlântica”, assim como para ITS.

Para diferenciar os possíveis eventos evolutivos que atuaram sobre indivíduos de *C. ipecacuanha* das três regiões, foram feitos três testes estatísticos para os três loci analisados. O Teste H de Fay e Wu (H_{FW}) se mostrou o mais rigoroso, não rejeitando a hipótese nula de evolução neutra. Valores negativos significativos dos testes D de Tajima (D_T) e F de Fu e Li (F_{FL}) foram observados para o gene *MetE* somente na “Amazônia”. Para ITS, valores negativos e significativos dos testes D de Tajima (D_T) e F de Fu e Li (F_{FL}) foram observados na “Mata Atlântica” e no “total”. Para os espaçadores cpDNA, nenhum destes testes foi significativo.

Tabela 4: Análise de variância molecular (AMOVA) para dois e três níveis hierárquicos em populações de *C. ipecacuanha* em relação ao gene *MetE*.

<i>Fonte de Variação</i>	<i>Graus de liberdade</i>	<i>Soma de Quadrados</i>	<i>Componentes de Variância</i>	<i>Porcentagem da variância total (%)</i>
<i>Dois níveis hierárquicos</i>				
Entre regiões	2	1,804	0,01486 Va	2,96
Dentro de regiões	92	44,817	0,48714 Vb	97,04
Total	94	46,621	0,50200	100
<i>Três níveis hierárquicos</i>				
Entre regiões	2	1,804	0,01060 Va	2,11
Entre indivíduos dentro das regiões	23	13,884	0,04278 Vb	8,53
Dentro de indivíduos	69	30,993	0,44831 Vc	89,36
Total	94	46,621	0,50169	100

A análise de variância molecular AMOVA a partir de sequências *MetE* considerando dois níveis hierárquicos evidencia que o principal componente da variância total corresponde a variações dentro das diferentes regiões (97,04%) (Tabela 4). Assim, dividindo as regiões em seus indivíduos e considerando os clones dentro de cada indivíduo como um terceiro nível hierárquico e recalculando a AMOVA, observamos que a maior parte da variação anteriormente encontrada era devido às diferenças entre clones dentro de cada indivíduo (89,36% da variação total). O baixo valor da estatística F_{ST} para o locus *MetE* (F_{ST} = 0.02961) corrobora com os resultados da AMOVA, evidenciando que as diferenças de frequências alélicas não foram significativas entre as três regiões. No entanto, elevados valores de F_{ST} para ITS (0.54882) e cpDNA (0.66685) indicam a elevada diferenciação entre as três grandes regiões geográficas de ocorrência da espécie.

Por meio das redes de haplótipos obtidas para o gene (Figura 5) e para o CDS (Figura 6) podemos observar uma relação entre as sequências de *MetE* consistente em nível geográfico, em que haplótipos mais proximamente relacionados foram simpatricamente distribuídos. Além disso, todos os haplótipos foram restritos a uma única região, ou seja, não foram observados haplótipos compartilhados entre as três regiões. Não é possível fazer inferências sobre um ancestral comum, uma vez que os haplótipos não coalescem num único ponto da rede que corresponda a um haplótipo de alta frequência e de distribuição geográfica ampla. Pela rede de haplótipos do gene também é possível observar a presença de seis clones distintos dentro de um mesmo indivíduo (TAN9), sugerindo que mais de uma cópia do gene *MetE* pode estar presente no genoma de ipeca.

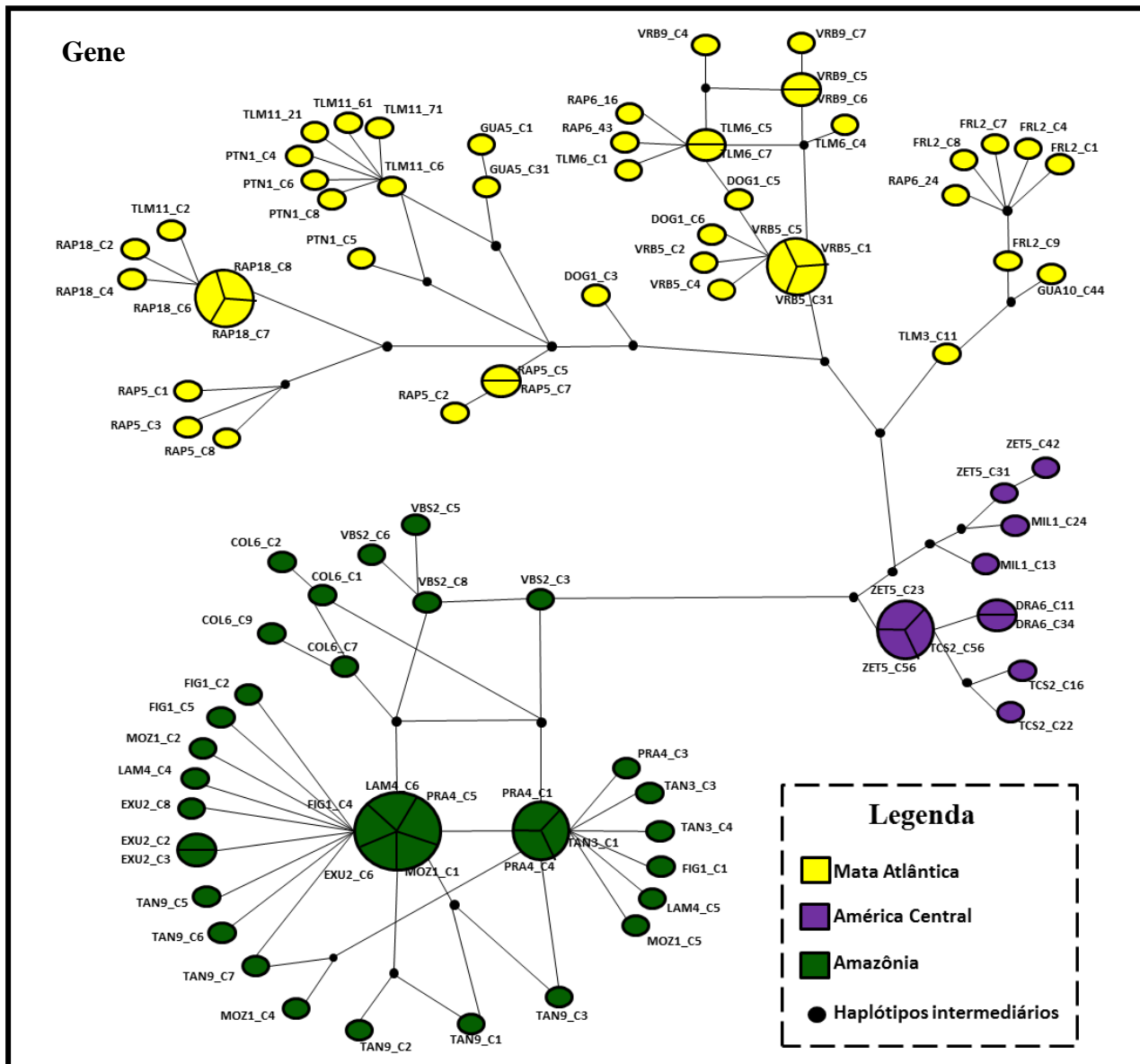


Figura 5: Rede de haplótipos obtida a partir de sequências do gene de metionina sintase em indivíduos de *C. ipecacuanha*, utilizando o algoritmo de *median joining*. Cada círculo representa um haplótipo e os clones dos indivíduos associados são indicados. Os tamanhos dos círculos correspondem às frequências dos haplótipos. Haplótipos intermediários referem-se à haplótipos não amostrados ou extintos na população.

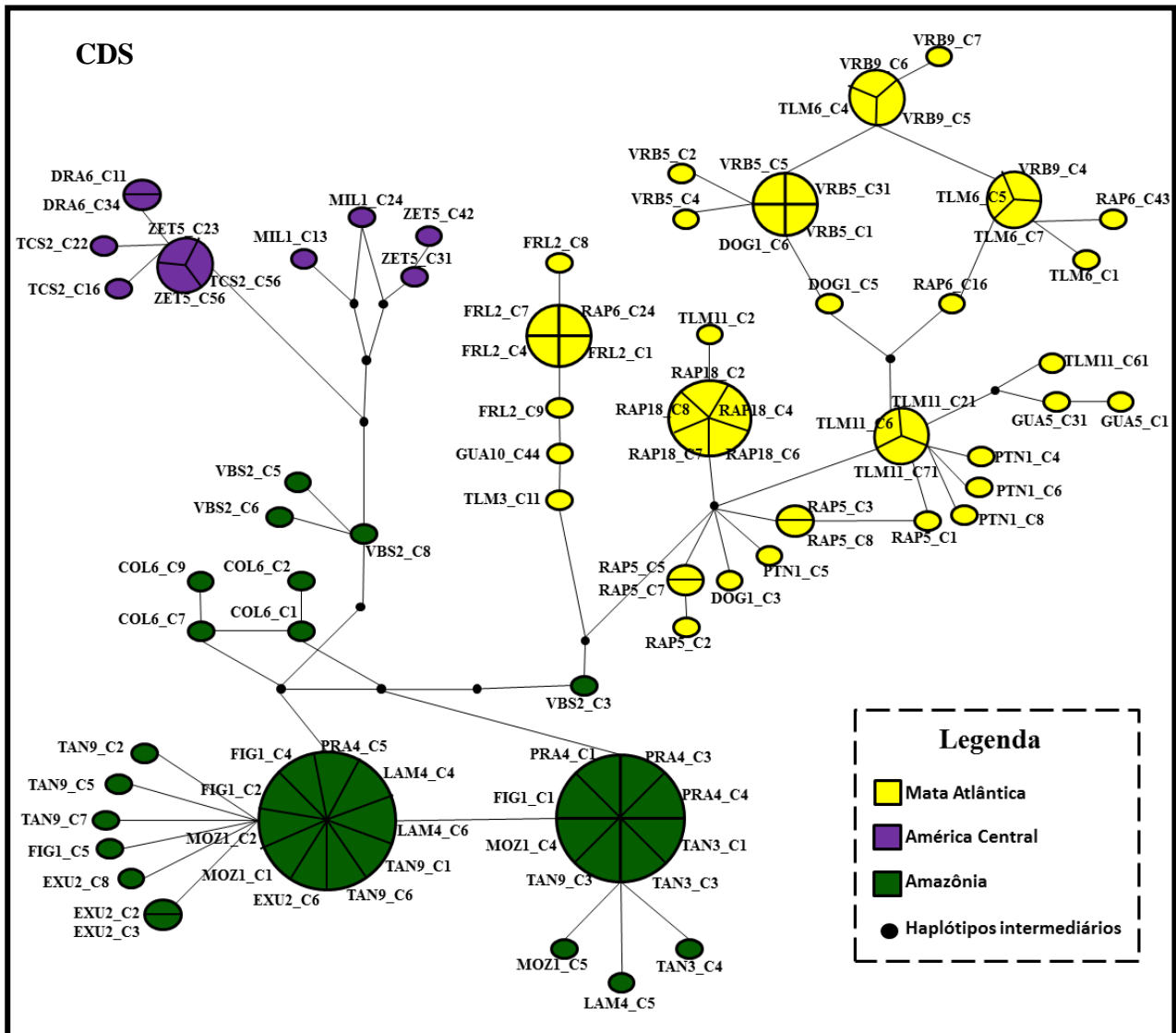


Figura 6: Rede de haplótipos obtida a partir de seqüências CDS do gene metionina sintase em indivíduos de *C. ipecacuanha*, utilizando o algoritmo de *median joining*. Cada círculo representa um haplótipo e os clones dos indivíduos associados são indicados. Os tamanhos dos círculos correspondem às frequências dos haplótipos. Haplótipos intermediários referem-se à haplótipos não amostrados ou extintos na população.

Os testes de ajustamento de modelos de substituição nucleotídica mostraram que para a região de exons, o modelo HKY+I+G obtido por AIC (Akaike Information Criterion) foi o que melhor se ajustou. Para a região não codificante (introns + 3`UTR), o melhor modelo indicado foi HKY+G obtido por hLRT (hierarquical Likelihood Ratio Test).

A árvore filogenética do gene completo reconstruída por inferência Bayesiana e enraizada (Figura 7) também não nos permitiu identificar a história genealógica das sequências, uma vez que os clones *MetE* de *C. ipecacuanha* formam uma politomia basal em relação à espécie *C. lucida*. Também não foi observada a formação de clados por regiões geográficas, pois os clones de GUA5 foram mais associados às outras regiões que à “Mata Atlântica”.

Explorando a árvore filogenética obtida para a região codificante - CDS (Figura 8), é possível observar uma melhor resolução dos clados em relação às três regiões geográficas de ocorrência da espécie, sugerindo que homoplasias podem ter ocorrido nas regiões não codificantes que levaram à ausência de estruturação geográfica dos clados na análise do gene completo. Portanto, nesta árvore, é possível observar que as regiões “Amazônia” e “América Central” são mais próximas entre si em relação à “Mata Atlântica”. Quanto à ancestralidade dos haplótipos, podemos observar que dos ramos que interligam “*C. lucida*” e “Mata Atlântica” não emergiram relações robustas (baixa probabilidade posterior dos nós), o que sugere uma maior proximidade filogenética entre estes haplótipos.

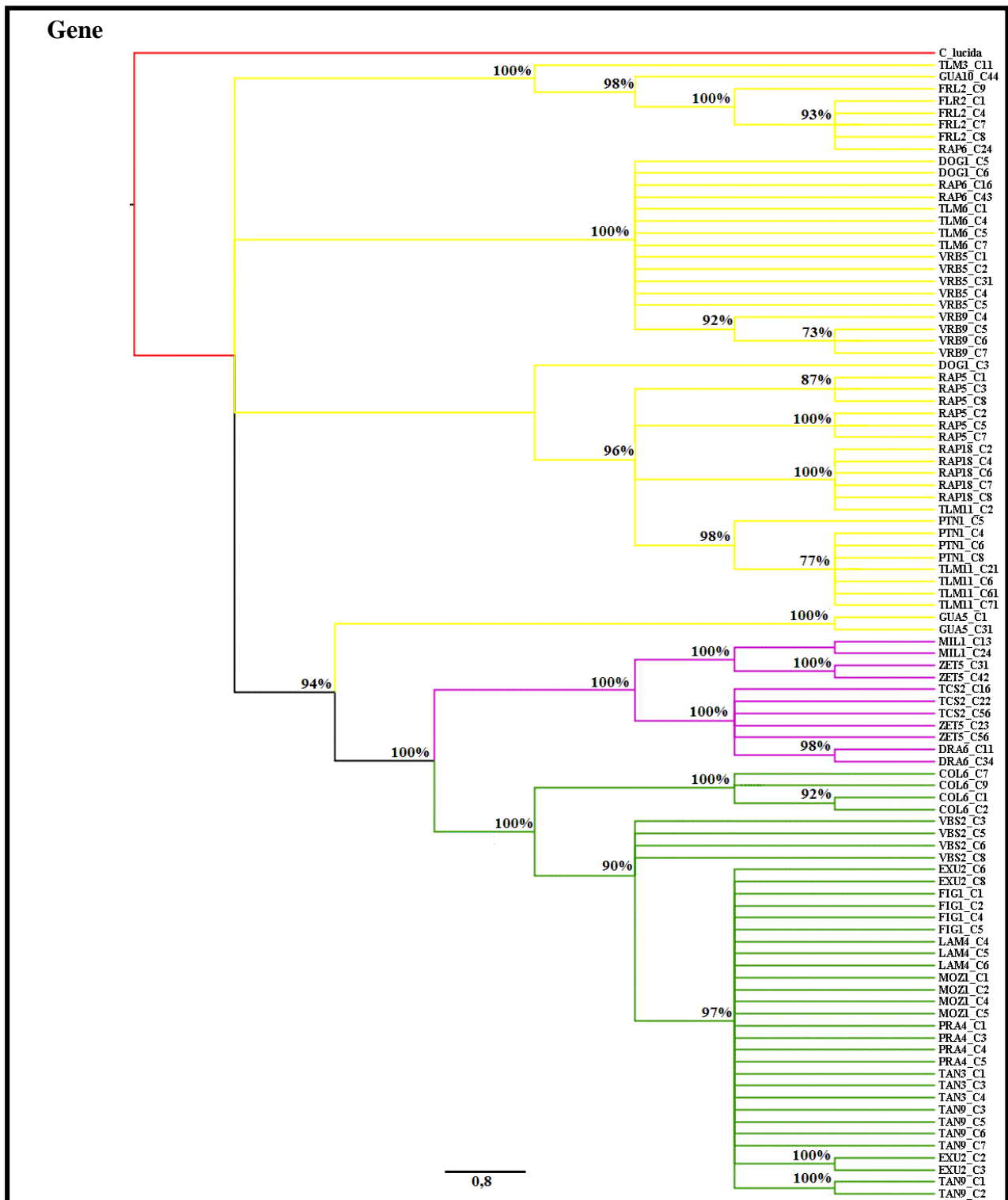


Figura 7: Arvore filogenética bayesiana enraizada obtida a partir da sequência do gene de metionina sintase (*MetE*) a partir de 26 indivíduos de *Carapichea ipecaçuanha*. Para a região de exons foi usado o modelo HKY+I+G e para a região não codificante foi usado o modelo HKY+G. O tamanho de cada ramo é proporcional à barra de escala que representa o número esperado de substituições por sítio. Valores suporte de probabilidade posterior acima de 70% são apresentados para cada ramo que se subdivide. As cores vermelha, amarela, roxo e verde correspondem, respectivamente, a clones de indivíduos pertencentes o grupo externo (*C. lucida*), Mata Atlântica, América Central e Amazônia. Foram realizadas 1.000.000 gerações de MCMC para o gene completo e houve convergência a 0,99%.

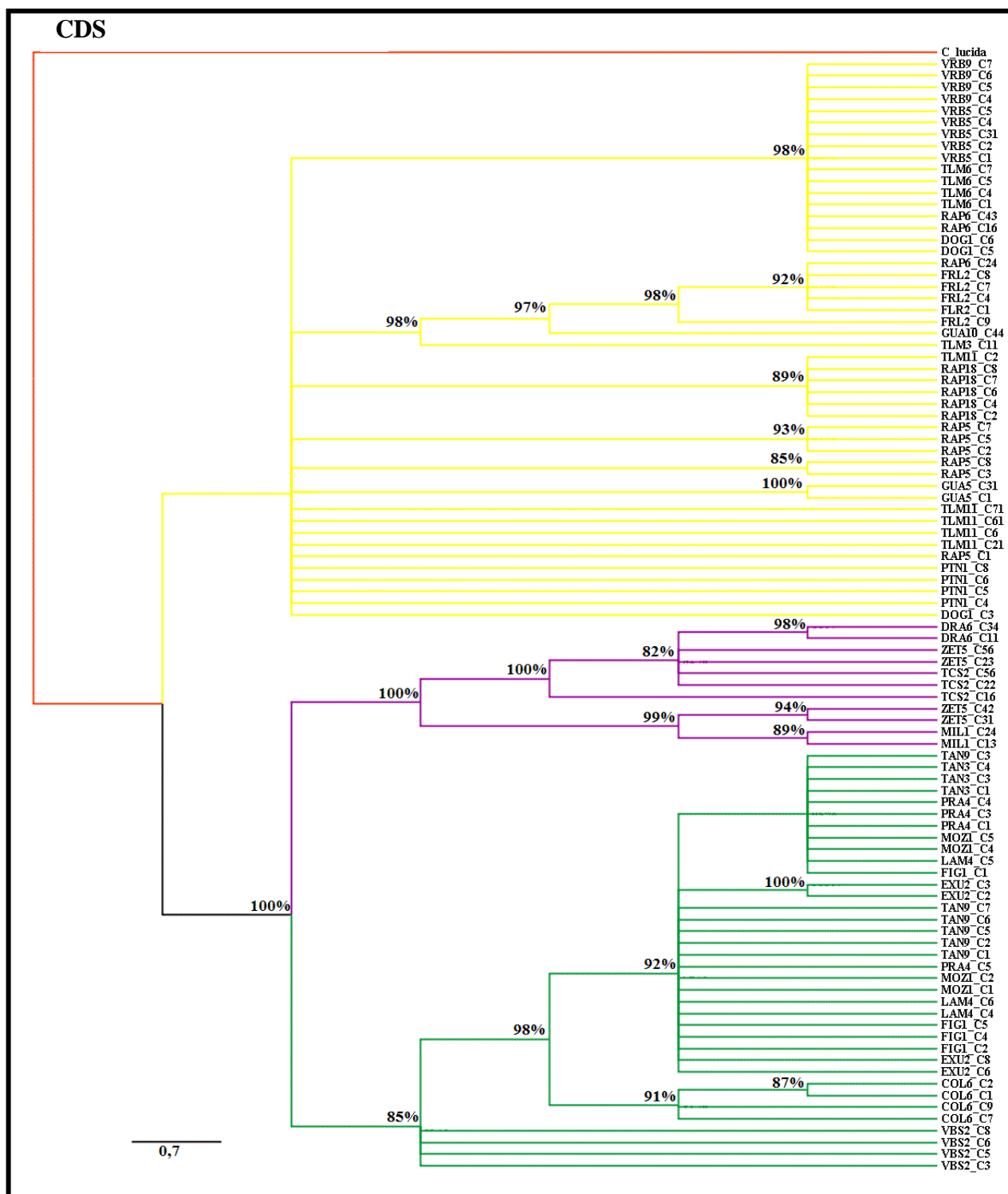


Figura 8: Árvore filogenética bayesiana enraizada obtida a partir das sequências de CDS de metionina sintase (*MetE*) a partir de 26 indivíduos de *Carapichea ipecacuanha*. O tamanho de cada ramo é proporcional à barra de escala que representa o número esperado de substituições por sítio. Valores suporte de probabilidade posterior acima de 80% são apresentados para cada ramo que se subdivide. As cores vermelha, amarela, roxo e verde correspondem, respectivamente, a clones de indivíduos pertencentes o grupo externo (*C. lucida*), Mata Atlântica, América Central e Amazônia. Foram realizadas 1.500.000 gerações para o CDS e houve convergência a 0,99%.

3) Análises evolutivas do gene *MetE*

Uma maior diversidade na região de introns ($\pi_{TOTAL} = 0,01320$) foi observada em relação aos exons ($\pi_{TOTAL} = 0,00641$). A diversidade nucleotídica ao longo de toda sequência pode ser visualizada através da análise em janela deslizante (Figura 9).

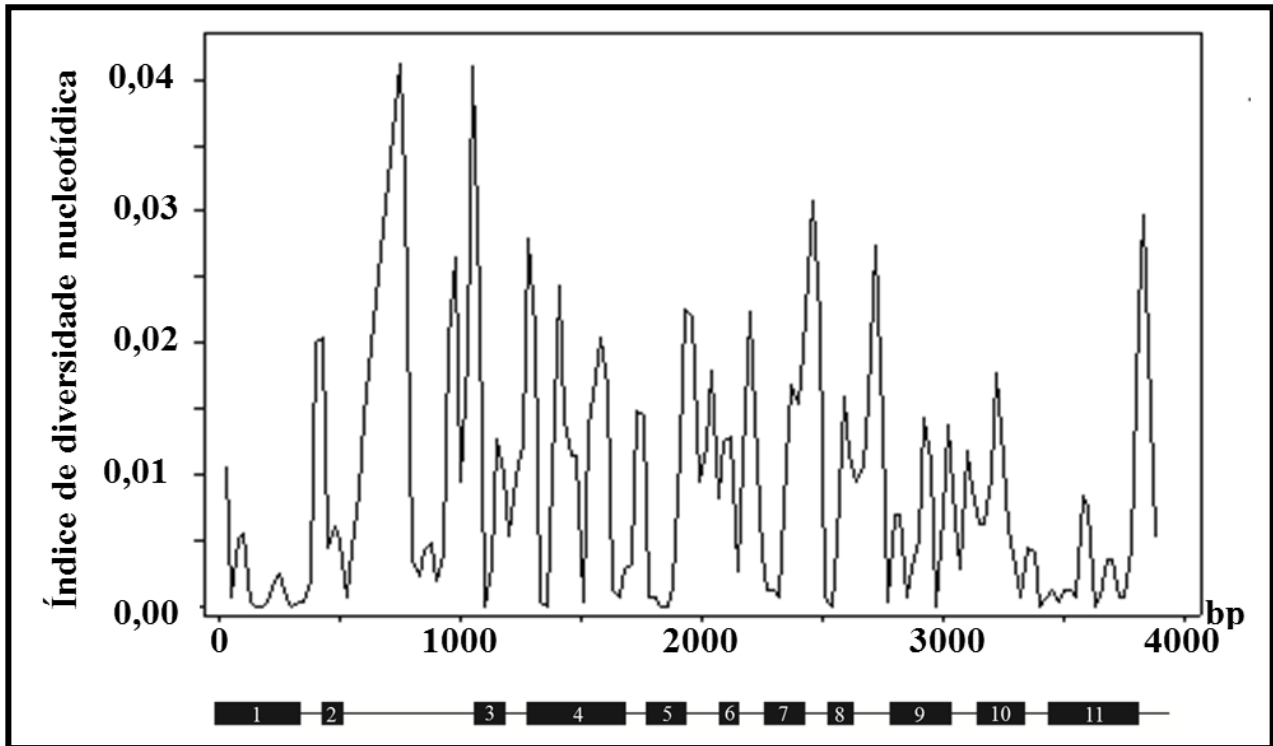


Figura 9: Análise em janela deslizante da diversidade nucleotídica. A análise foi baseada em uma janela de 50pb que se move através do gene em intervalos de 25pb. A estrutura exon/intron do gene é indicada para fornecer uma referência da diversidade nucleotídica em cada segmento.

Nos exons, ocorreram 97 sítios polimórficos, sendo 48 substituições sinônimas e 49 substituições não sinônimas cujas proporções variaram entre os diferentes exons e entre as três regiões (Tabela 5). A diversidade nucleotídica foi maior para sítios sinônimos ($\pi_{SIN}=0,01667$) em relação a sítios não-sinônimos ($\pi_{NSIN}= 0,00324$).

Os testes de neutralidade aplicados para todos os fragmentos do gene foram significativos apenas para o exon11 ($D_T = -1,68$ e $F_{FL} = -2,13$). Ainda que não significativo, o teste H de Fay e Wu, também apresentou valor negativo ($H_{FW} = -1,30$). Para as análises do CDS em relação às regiões de ocorrência, assim como para o gene completo, apenas a região “Amazônia” apresentou desvios significativos da neutralidade.

Tabela 5: Estatísticas descritivas para as diferentes fragmentos do gene *MetE* (exons e introns) e para o CDS nas três regiões de ocorrência de *Carapichea ipecacuanha*.

<i>Fragmentos do gene MetE</i>	<i>bp</i>	<i>S</i>	<i>Sin</i>	<i>NSin</i>	π_{TOTAL}	π_{Sin}	π_{Nsin}	D_T	F_{FL}	H_{FW}
<i>CDS</i>	2298	97	48	49	0,00641	0,01667	0,00324	-0,73	-0,96	0,99
Exon01	350	10	3	7	0,00280	-	-	-1,28	-1,62	0,60
Exon02	88	6	2	4	0,00448	-	-	-1,50	-1,20	0,36
Exon03	128	7	4	3	0,01395	-	-	0,73	1,23	-0,83
Exon04	392	20	12	8	0,01177	-	-	0,53	-0,11	0,14
Exon05	164	4	1	3	0,00384	-	-	-0,38	0,61	0,13
Exon06	81	7	2	5	0,01253	-	-	-0,61	-0,66	0,43
Exon07	167	9	3	6	0,00663	-	-	-0,93	-1,67	0,51
Exon08	106	3	3	0	0,00227	-	-	-1,07	-0,85	0,22
Exon09	251	11	7	4	0,00492	-	-	-1,11	0,12	0,17
Exon10	202	9	6	3	0,00969	-	-	0,29	-0,01	0,55
Exon11	369	11	5	6	0,00210	-	-	-1,68**	-2,13*	-1,30
<i>Introns</i>	1592	116	-	-	0,01320	-	-	-0,59	-0,01	6,34
<i>Regiões</i>										
Mata Atlântica	2298	44	24	20	0,00290	0,00983	0,00076	-1,13	-1,77	0,955
Amazônia	2298	27	8	19	0,00121	0,00253	0,00081	-1,98*	-3,09**	-1,61
América Central	2298	23	12	11	0,00399	0,00901	0,00243	0,76	0,62	-5,87

bp, tamanho da sequência em pares de base;

S, número de sítios polimórficos;

Sin, número total de sítios sinônimos;

NSin, número total de sítios não sinônimos;

π_{TOTAL} , diversidade nucleotídica considerando todos os sítios;

π_{Sin} , diversidade nucleotídica para sítios sinônimos;

π_{Nsin} , diversidade nucleotídica para sítios não sinônimos;

D_T , teste D de Tajima sem grupo externo (* $P < 0,05$ e ** $P < 0,02$);

F_{FL} , teste F de Fu e Li sem grupo externo(* $P < 0,05$ e ** $P < 0,02$);

H_{FW} , teste H de Fay e Wu com grupo externo;

Não foram detectados eventos de recombinação nas análises por GARD para o gene completo e CDS de *MetE*. Portanto, não há evidências para discordâncias filogenéticas e as análises de seleção positiva foram realizadas considerando uma única árvore filogenética para o gene.

Dentre os algoritmos utilizados para inferir sobre eventos de seleção, SLAC mostrou-se o mais conservador, em que apenas 04 sítios sob seleção negativa foram detectados e nenhum sítio foi detectado para seleção positiva. REL mostrou-se o mais sensível para seleção negativa com 48 sítios detectados, sendo este teste mais sujeito a falsos positivos (na Tabela 6 apenas os sítios significativos em consenso com as outras análises são apresentados). E o algoritmo FUBAR foi o único em que eventos de seleção positiva foram detectados (04 sítios). Ainda de acordo com este último, é provável que 50,90% dos sítios estejam experimentando seleção negativa, 6,1% evolução neutra e 43,0% seleção positiva, sendo encontradas fortes evidências de seleção apenas para os sítios apresentados na Tabela 6.

Tabela 6: Sítios com evidência de seleção negativa (azul) e positiva (vermelho) pela inferência de três algoritmos (SLAC, REL, FUBAR). Taxas dN/dS normalizadas estatisticamente significativas são apresentadas para cada sítio. Os valores médios da taxa dN/dS também são apresentados para cada análise.

Posição	SLAC		REL		FUBAR	
	dN/dS	p-valor	dN/dS	Fator Bayesiano	dN/dS	Prob. posterior
7	-123,60	0,0001	-139,85	1,86x10 ⁸⁶	-19,824	0,9999
79	-	-	-4,2637	9743,65	-5,6618	0,9678
179	-	-	-2,4409	10306,10	-6,4298	0,9131
297	-32,29	0,0803	-4,3835	12279,10	-9,3143	0,9848
539	-	-	-4,335	11599,20	-8,6892	0,9838
611	-1,835	0,0799	-4,9282	13627,50	-10,693	0,9885
621	-	-	-4,3308	12118,70	-9,0963	0,9855
708	-2,267	0,0370	-6,4308	1,73 x10 ⁶	-13,618	0,9988
731	-	-	-4,3081	11599,10	-8,6818	0,9829
151	-	-	-	-	4,862	0,9559
336	-	-	-	-	3,828	0,9041
509	-	-	-	-	7,130	0,9690
747	-	-	-	-	3,296	0,9083
dN/dS médio	0,4317		-0,6010		4,7500	

Um p-valor menor que 0,1 foi considerado estatisticamente significativo para presença de seleção nas análises SLAC. Fatores Bayesianos maiores que 50 foram considerados estatisticamente significativos para análises REL. Probabilidades posteriores maiores que 0,9 foram consideradas estatisticamente significativas para as análises FUBAR. Valores não significativos não são apresentados.

O sítio positivamente selecionado na posição 151 corresponde a uma substituição do tipo transversão na segunda base do códon (AGU → ACU) que gera a alteração de uma Serina por uma Treonina. O códon AGU ocorre apenas em indivíduos da “Mata Atlântica” e no indivíduo MIL1 da “América Central”. E o códon ACU ocorre nos demais indivíduos da “América Central” e em todos os indivíduos da “Amazônia”.

Na posição 336, uma transição altera a segunda base do códon (GAU → GGU), causando a substituição de um Aspartato por uma Glicina e ocorre apenas nos clones TLM11_C2 e GUA5_C1 da “Mata Atlântica”.

Na posição 509, ocorre uma transição na primeira base do códon (UUU → CUU) que altera uma Fenilalanina por uma Leucina e aparece apenas nos clones RAP5_C2, TLM11_C61, VRB5_C2 da “Mata Atlântica”.

Na posição 747, ocorre uma transição na segunda base do códon (CUC → CCC), alterando uma Leucina por uma Prolina e aparece nos clones VRB5_C4 da “Mata Atlântica” e TCS2_C22 da “América Central”.

Analisando a posição destas mutações na estrutura tridimensional de metionina sintase (Figura 10), foi observada a ocorrência em regiões de α -hélice (336 e 747) e regiões de *loops* (151 e 509).

As mutações negativamente selecionadas ocorreram na terceira base do códon, causando substituições que não alteraram o aminoácido codificado. Quatro delas (7, 297, 611 e 731) ocorreram em regiões de estrutura secundária não definida (*loops*) e cinco (79, 179, 539, 621 e 708) em regiões α -hélice (Figura 11).

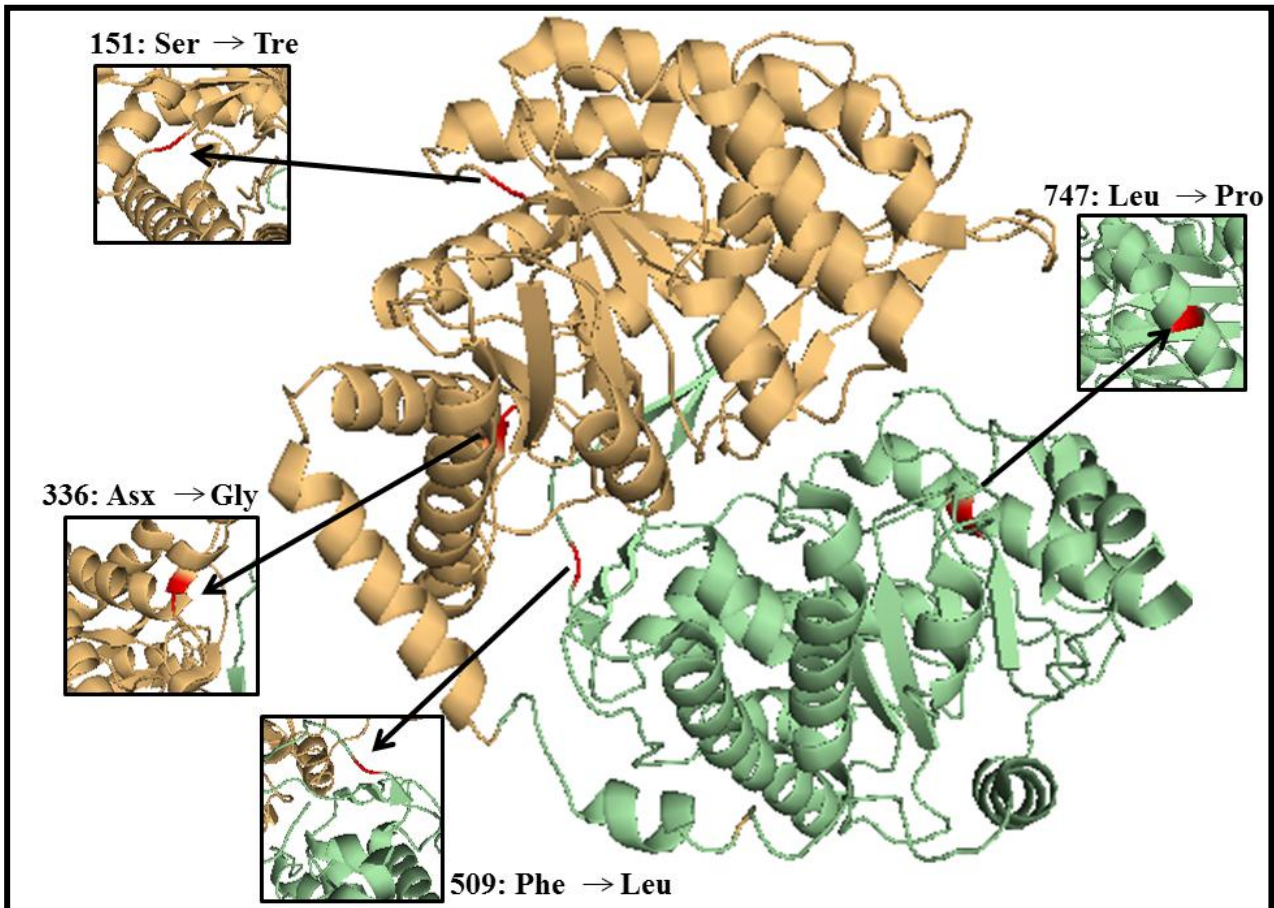


Figura 10: Estrutura tridimensional da enzima MetE de *Arabidopsis thaliana* (código 1U1H no *Protein Data Bank*) representada em diagrama *cartoon*. As cores laranja e verde correspondem às porções N-terminal e C-terminal, respectivamente. Os sítios positivamente selecionados no gene homólogo de *C. ipecacunha* são marcados em vermelho. As setas mostram, em melhor ângulo, o sítio onde ocorreu a mutação. Os números correspondem ao posicionamento da mutação na estrutura primária da proteína (conforme a Tabela 6) e, em seguida, é indicada a substituição aminoacídica que ocorreu no sítio.

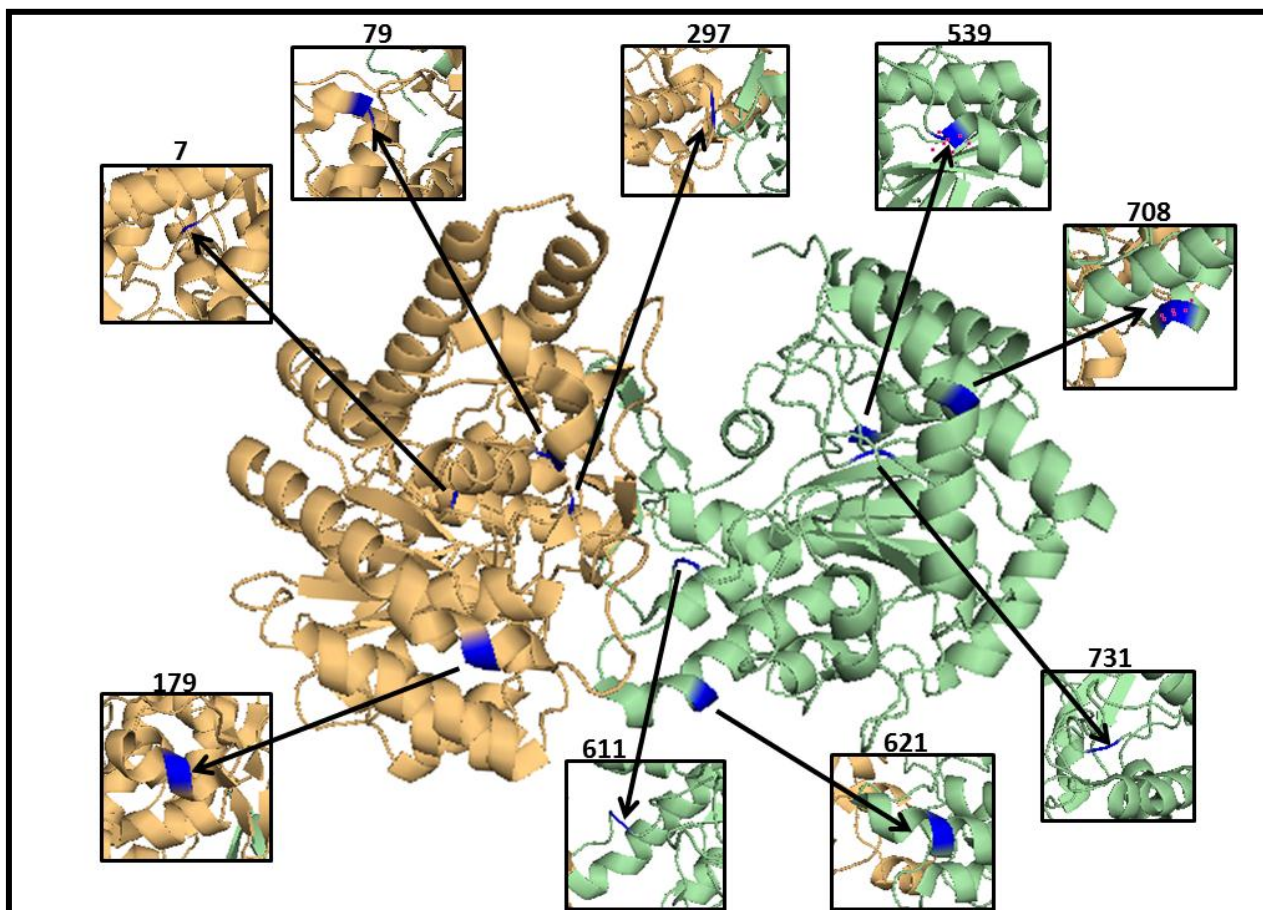
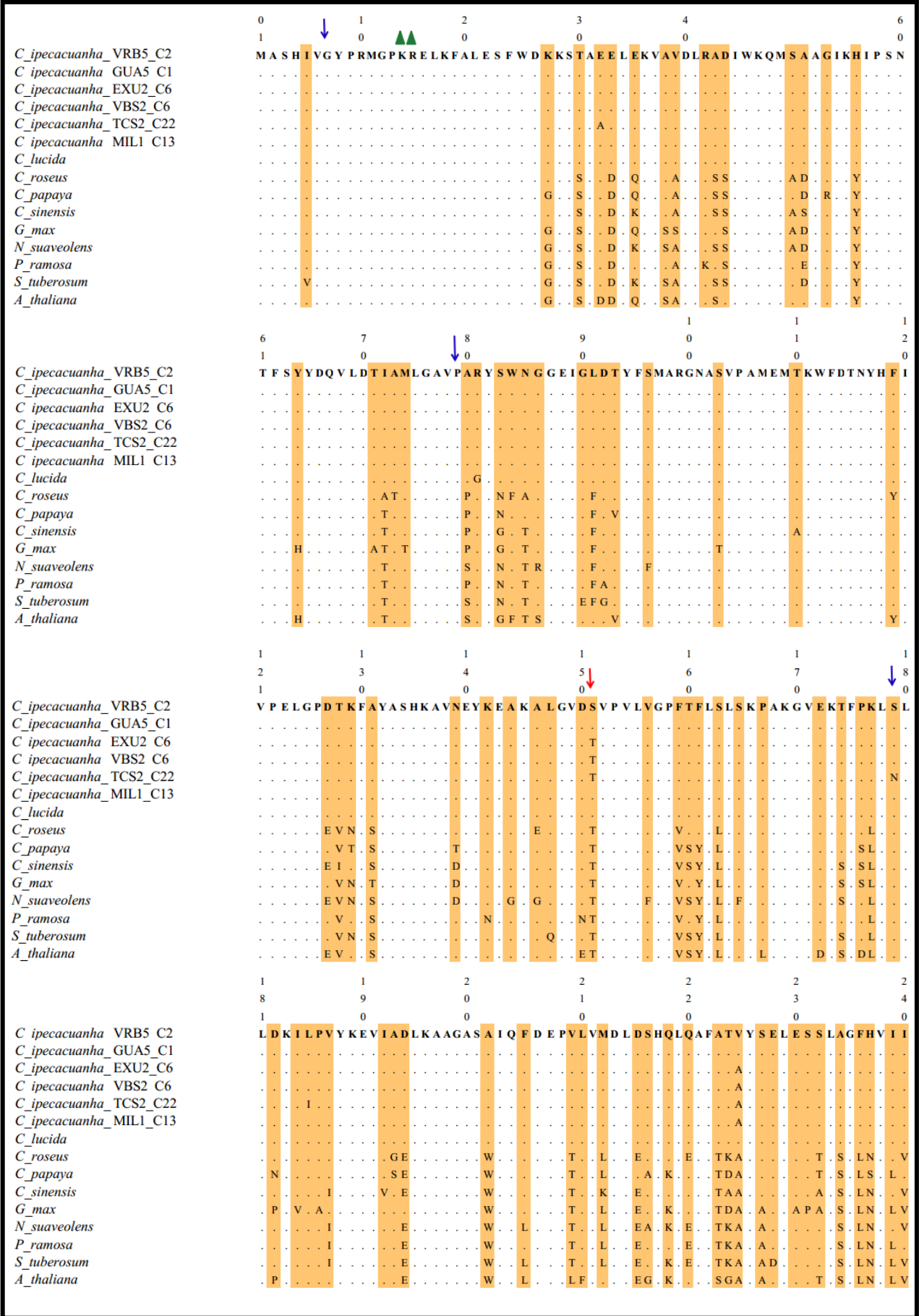


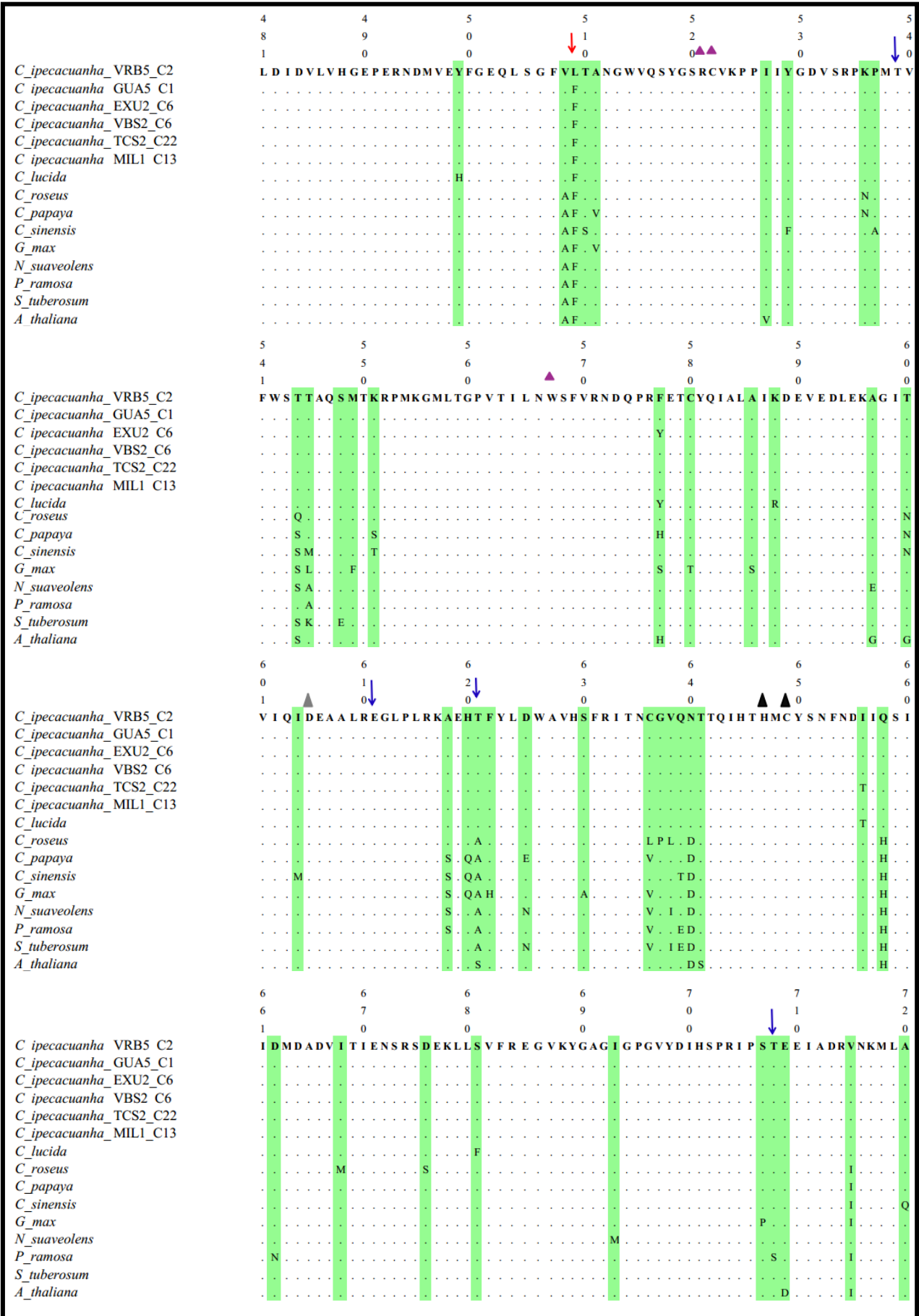
Figura 11: Estrutura tridimensional da enzima MetE de *Arabidopsis thaliana* (código 1U1H no *Protein Data Bank*) representada em diagrama *cartoon*. As cores laranja e verde correspondem às porções N-terminal e C-terminal, respectivamente. Os sítios negativamente selecionados no gene homólogo de *C. ipecacunha* são marcados em azul. As setas mostram, em melhor ângulo, o sítio onde ocorreu a mutação e os números acima correspondem ao posicionamento da mutação na estrutura primária da proteína, conforme a Tabela 6.

No sítio ativo, localizado na porção C-terminal, os aminoácidos envolvidos na coordenação de átomos de zinco (His⁶⁴⁷, Cys⁶⁴⁹ e Cys⁷³³) e na interação com a molécula homocisteína (Asp⁶⁰⁵, Ile⁴³⁷ e Ser⁴³⁹), segundo Ferrer *et al.* (2004), foram conservados em todos os clones de *MetE* em ipeca (dados não mostrados) e entre as proteínas de outras espécies vegetais (Figura 12). O resíduo de Trp⁵⁶⁷ que permite o empilhamento da metade pterina do N⁵-metiltetrahydrofolato e alguns resíduos da porção N-terminal (Lys¹⁴, Arg¹⁵ e Asn³³⁹) que estão relacionados à estabilização da cadeia de poliglutamato negativamente carregada do N⁵-metiltetrahydrofolato, também se mostraram conservados. Somente a mutação positivamente selecionada na posição 509, localiza-se em uma região associada à função da proteína (*loop* catiônico composto pelos resíduos 507-529 que ancora o primeiro resíduo glutamil do N⁵-metil-tetrahydrofolato).

Ainda observando o alinhamento múltiplo de proteínas (Figura 12), é possível notar que a maioria das mutações encontra-se na porção N-terminal da proteína (64,94% de todas as mutações detectadas no alinhamento).



Continua...



Continua...

IV) DISCUSSÃO

1) O gene *MetE* em *C. ipecacuanha*

O cDNA para Metionina Sintase (*MetE*) foi descrito para diversas plantas e comparações destas sequências nucleotídicas e de aminoácidos por elas preditos em relação ao putativo CDS de *MetE* em ipeca mostraram elevada similaridade, principalmente em regiões de motivos associados à função da enzima (Batista, 2010). Porém, poucos trabalhos descrevem análises para a sequência completa (exons e introns) deste gene. Neste trabalho, isolou-se, pela primeira vez, o gene *MetE* completo de *C. ipecacuanha* e análises populacionais e evolutivas foram conduzidas.

Nas análises intraespecíficas, o gene *MetE* de *C. ipecacuanha* apresentou alta similaridade entre os indivíduos amostrados, sendo as divergências na região codificante devido apenas à substituições nucleotídicas. Alterações estruturais na organização exon-intron (por exemplo: ganho ou perda de exons ou introns; exonização e pseudoexonização; inserções e deleções intraexônicas – Xu *et al.*, 2012) não foram observadas nas comparações intraespecíficas, nem interespecíficas.

O gene *MetE* possui uma ou poucas cópias no genoma de plantas (Hesse *et al.*, 2004). Análises de *Southern Blot* realizadas, por exemplo, em *A. thaliana*, *Solanum tuberosum*, *Catharanthus roseus*, *Hordeum vulgare* sugeriram a presença de apenas uma ou poucas cópias gênicas de *MetE* devido à uma ou duas bandas observadas por diferentes enzimas de restrição utilizadas (Gakière *et al.*, 1999; Ravanel *et al.*, 2004; Zeh *et al.*, 2002; Eichel, *et al.*, 1995, Narita *et al.*, 2004). Em ipeca, o sequenciamento do gene *MetE* completo, mostrou a presença de vários clones distintos, sugerindo que até 6 alelos (3 genes) podem ser encontrados dentro de um mesmo indivíduo. No entanto, muitos destes clones se diferenciavam por apenas um passo mutacional, o que pode ter ocorrido por erros de fidelidade da *Taq* DNA polimerase que, porventura, não foram excluídos das análises. Experimentos específicos para a detecção do número de cópias de *MetE* nesta espécie ainda são necessários.

2) Estrutura populacional

Estudos anteriores com marcadores ISSR (Rossi *et al.*, 2009), espaçadores intergênicos cloroplastídicos e espaçadores transcritos internos do DNA ribossomal nuclear (Oliveira *et al.*, 2010) revelaram um alto nível de diferenciação genética entre os indivíduos de *C. ipecacuanha* amostrados nas três grandes regiões de ocorrência da espécie (Mata Atlântica, Amazônia e América Central). No presente trabalho, as análises populacionais de *MetE* mostraram por meio das redes de haplótipos (para o gene completo e CDS) e da árvore filogenética (para CDS) uma relação de ancestralidade comum entre haplótipos de uma mesma região e, conseqüentemente, uma correspondência entre distribuição geográfica dos indivíduos e arquitetura da rede de haplótipos. No entanto, como a maioria destes haplótipos se comporta como *singletons* (haplótipos representados por uma única sequência na amostra), uma grande variação dentro da mesma região é observada e poucos passos mutacionais conectam clones de diferentes regiões. Esta baixa estruturação dos haplótipos de *MetE* em relação às três regiões é evidenciada pelo baixo valor da análise de F_{ST} (0.02961) e pelo resultado da AMOVA, em que a maior parte da variação foi devido às diferenças entre clones dentro de um mesmo indivíduo (89,36%) e não entre as três regiões (2,11%). Pousada e Crandall (2001) explicam que a relação entre genes amostrados a partir de indivíduos da mesma espécie realmente não é hierárquica como na filogenia *strictu sensu*. As relações intraespecíficas entre cópias gênicas ancestrais e descendentes podem ser resultados da sobreposição de *pools* gênicos, de reprodução sexuada, de recombinação, de pequenos números de mutações relativamente recentes, além de efeitos de eventos demográficos e de seleção, resultando em redes de haplótipos com multifurcações, homoplasias e relações reticuladas dentro de populações. Além disso, o uso de genes nucleares em estudos populacionais pode levar à reconstrução de eventos de duplicação ao

invés de eventos populacionais, devido à complexa dinâmica evolutiva de famílias gênicas nucleares (Sang, 2002).

3) Diversidade e Desvios da Neutralidade

Como a maioria dos genes nucleares de eucariotos, o gene *MetE* em ipeca é composto por exons e introns. A diversidade nucleotídica é cerca de 2 vezes maior na região não-codificante, composta pelos introns e parte da região 3'UTR, do que na região codificante, composta pelos exons. Isto ocorre porque, geralmente, exons estão sobre forte seleção purificadora e mutações deletérias são eliminadas, sendo encontradas, relativamente, baixas taxas de substituição nucleotídica; enquanto em introns, mutações podem ser acumuladas por não causar prejuízos para o *fitness* do indivíduo (Sang, 2002).

A diversidade nucleotídica para o locus *MetE* considerando todas as amostras de ipeca foi estimada em 0,00899. Resultados semelhantes foram obtidos para populações de *Dunnia sinensis* (0,00520, Chiang *et al.*, 2002), *Quercus mongólica var. crispula* (0,00646, Quang *et al.*, 2009) e *Fokienia hodginsii* (0,00499, Quang *et al.*, 2012) quando a sequência parcial do gene *MetE* (abrangendo do exon1 ao exon4) foi analisada.

As análises da diversidade nucleotídica intra-regional de *MetE* mostraram a presença de variação dentro das três regiões, enquanto para outros *loci* do genoma analisados (espaçadores *trnT-trnL* do DNA cloroplastídico e ITS do DNA ribossomal nuclear) apenas foi observado variação dentro de “Mata Atlântica”. As análises comparativas dos três *loci* revelaram alta heterogeneidade quanto à diversidade genética e quanto ao desvio da neutralidade.

Em plantas, sequências do genoma de cloroplastos evoluem muito lentamente em relação ao DNA nuclear e apresentam baixos níveis de variação genética intraespecíficos. O cpDNA, na maioria das angiospermas, também é mais sensível à mudanças demográficas, por representarem a metade do tamanho efetivo populacional de marcadores biparentalmente herdados e por não sofrerem recombinação (Sang, 2002; Mitton, 1994). No caso de ipeca, baixos valores de diversidade nucleotídica e haplotípica foram detectados para os espaçadores *trnT-trnL* do cpDNA, o que pode ser consequência das características inerentes ao genoma cloroplastídico, como também dos eventos que levaram esta espécie ao *status* de ameaçada de extinção.

As análises do locus ITS mostraram um maior valor de diversidade nucleotídica que os outros *loci* analisados e 22 haplótipos foram obtidos a partir de 66 sequências, sendo a maioria encontrada na região “Mata Atlântica”. Sequências repetidas em tandem como ITS do nrDNA, devido à evolução em concerto, tendem a se homogeneizar e reduzir o número de sítios polimórficos em nível intraindividual e populacional. Queiroz *et al.* (2011) sugeriram que há presença de evolução em concerto, porém ainda incompleta, e que há presença de pseudogenes altamente divergentes nas cópias de ITS do genoma nuclear de *C. ipecacuanha*, o que pode explicar a variação encontrada.

Apenas uma tendência de desvio da neutralidade (valores de D_T e F_{FL} negativos e significativos) foi observada em relação às regiões de ocorrência de ipeca e aos diferentes *loci*. Valores negativos e significativos estimados em testes de neutralidade indicam que há um excesso de polimorfismos de baixa frequência, maior do que o esperado com base na Teoria Neutralista de Evolução. Neste caso, três inferências são possíveis: 1) houve uma recente expansão do tamanho populacional; ou 2) presença de seleção purificadora recente de uma mutação ligada; 3) erros de fidelidade da *Taq* DNA polimerase não detectados e, portanto, não excluídos das análises.

Os testes de neutralidade não foram significativo para os espaçadores *trnT-trnL* do cpDNA, não rejeitando, portanto, a hipótese de evolução neutra, a qual condiz com regiões do genoma de cópia única, não codificantes e que cuja variação em nível molecular não afeta o *fitness* do indivíduo. Portanto, o destino evolutivo da variação observada no cpDNA é melhor explicado por processos estocásticos.

Em ITS, valores negativos significativos foram obtidos para as amostras como um todo e para “Mata Atlântica”. O fenômeno da evolução em concerto, que ocorre nestas regiões gênicas, exerce uma forte pressão purificadora devido a esta conversão gênica viesada que leva a homogeneização das cópias de ITS. Assim, este processo evolutivo não adaptativo, pode ser o fator responsável pelos desvios da neutralidade encontrados.

Para o gene *MetE*, valores negativos e significativos dos testes de neutralidade foram obtidos somente para a região “Amazônia” quanto ao gene completo e CDS. E quando o CDS foi fragmentado em seus exons, apenas o exon11 apresentou valores significativos, sugerindo a presença de um processo evolutivo distinto atuando neste gene, mais especificamente no exon11 e na região amazônica.

Portanto, os resultados sugerem que os três *loci* evoluíram independentemente e que eventos demográficos não explicam a variação observada, pois sua ocorrência afetaria profundamente todo o genoma. O mais provável é que estes *loci* estejam sob o efeito de diferentes processos evolutivos, característicos de cada região do genoma e influenciados pelos diferentes habitats em que os indivíduos se encontram (adaptação local).

4) Evolução molecular da região codificante de *MetE*

O gene *MetE* codifica a enzima metionina sintase, que por atuar na via de síntese e regeneração de AdoMet, está indiretamente associada ao metabolismo de crescimento vegetal, de interação com patógenos e de resposta a estresses (Ravanel *et al.*, 1998; Chiang, *et al.*, 2002; Narita *et al.*, 2004). A rápida evolução de genes relacionados às estas vias e os desvios da hipótese de evolução neutra encontrados, motivou-nos a testar a hipótese de seleção positiva para os sítios polimórficos presentes na região codificante do gene *MetE* nas populações disjuntas de *C. ipecacuanha*.

Mudanças na função do gene podem ser devido à substituições que alteram o aminoácido e/ou por mudanças estruturais na organização exon-intron (Xu, *et al.*, 2012). Não foram encontradas evidências de alterações estruturais no gene *MetE* em ipeca e somente substituições nucleotídicas foram detectadas na região codificante. A presença de 49 substituições não-sinônimas pode indicar a ocorrência de mutações vantajosas mantidas devido às diferenças de pressões seletivas em cada uma das três regiões de ocorrência da espécie. No entanto, apenas quatro destes sítios foram estatisticamente detectados como positivamente selecionados (151, 336, 50 e 747).

Na posição 151, a alteração de uma Serina por uma Treonina pode ser considerada uma substituição conservativa, pois ambos os aminoácidos possuem cadeias laterais não carregadas e polares e ocorrem em uma região de *loop*.

Na posição 336, há alteração de um Aspartato por uma Glicina. O Aspartato possui em sua cadeia lateral um segundo grupo hidroxila, o que lhe confere uma carga líquida negativa em pH7,0 e, portanto, propriedades hidrofílicas. Sua alteração por uma Glicina, um aminoácido de estrutura simples e que realiza interações fracas, pode causar uma quebra da estrutura secundária onde esta alteração ocorre.

A mutação na posição 509, que altera uma Fenilalanina por uma Leucina, posiciona-se em um *loop* catiônico associado à funcionalidade da enzima. Apesar de ambos aminoácidos terem características hidrofóbicas, a presença de um anel aromático na cadeia lateral da Fenilalanina pode tornar a região mais rígida. No entanto, se esta mutação causar algum prejuízo ao *fitness* do indivíduo, a ocorrência em apenas um dos clones dos indivíduos RAP5 (clone 2), TLM11 (clone 61), VRB5 (clone 2), pode ser compensada pelas demais cópias gênicas de *MetE* que podem assumir a função enzimática que lhe é convencionalmente atribuída.

Na posição 747, ocorre a alteração de uma Leucina por uma Prolina. Ambos são aminoácidos alifáticos não-polares, no entanto, a Prolina possui uma estrutura cíclica e um grupo amino secundário (imino) que é mantido em uma conformação rígida. Portanto, esta alteração pode

levar à redução da flexibilidade estrutural ou mesmo torções na estrutura de α -hélice onde ocorre. Entretanto, ainda são necessários estudos de modelagem protéica para realizar inferências mais confiáveis sobre os efeitos destas alterações na estrutura e atividade da enzima.

Os testes estatísticos em que apenas o critério $dN/dS > 1$ considera a presença de seleção positiva são extremamente estridentes e, provavelmente, levam à perda da maioria dos casos em que a seleção positiva está operando (Kreitman, 2000). O teste para detecção de seleção utilizando a taxa dN/dS foi originalmente desenvolvido para ser aplicado à sequências com grandes divergências genéticas, cujas diferenças entre elas representassem substituições ao longo de linhagens evolutivamente independentes (mutações fixadas entre espécies). Mesmo quando as amostras são oriundas de populações evoluindo independentemente, como os indivíduos das três regiões de ocorrência de ipeca, o poder da estatística dN/dS para detectar seleção positiva é baixo quando a maioria dos sítios em uma proteína evolui sobre seleção purificadora/negativa (Kryazhimskiy & Plotkin, 2008). Além disso, é importante ressaltar que polimorfismos são mutações cuja disposição final (fixação ou perda) ainda não está definida e estão sob efeito da deriva genética e da seleção dentro de uma espécie. Mas, ainda assim, fornecem informações sobre o processo de evolução, porém em diferentes profundidades de tempo que comparações realizadas entre diferenças fixadas em alelos oriundos de diferentes espécies (Kreitman, 2000).

O algoritmo FUBAR indicou que a maioria das substituições nucleotídicas encontradas em *MetE* para ipeca parecem estar evoluindo sob seleção negativa (50,90%), o que sugere que a seleção está operando a favor da conservação da estrutura e função protéica. Portanto, ainda que elevada variabilidade tenha sido encontrada entre clones de um mesmo indivíduo (análise de AMOVA), os sítios associados à função enzimática se mantiveram preservados, o que sugere uma alta adaptabilidade desta enzima nas três regiões de ocorrência da espécie. O alinhamento múltiplo das proteínas das diversas espécies também corrobora com a hipótese de conservação da função enzimática, uma vez que há grandes motivos extremamente conservados e os principais resíduos associados à função da enzima, de acordo com Ferrer *et al.* (2004), também são mantidos em todas as espécies analisadas.

Em muitos organismos, a metade C-terminal da enzima metionina sintase é mais conservada que a metade N-terminal, uma vez que a primeira abriga os sítios de ligação ao co-fator zinco e ao substrato (Pejchal & Ludwig, 2005). E, interessantemente, a maioria das mutações não-sinônimas encontradas no CDS de ipeca encontram-se na porção N-terminal (30 versus 19) e, ainda, 64,94% das mutações observadas no alinhamento interespecífico das proteínas também ocorreram nessa porção da enzima, aumentando as evidências de conservação da função proteica.

V) CONCLUSÕES

A região sequenciada de *MetE* abrangeu 3.890pb, sendo 2.298pb correspondentes à região codificante (composta por 11 exons) e 1.592pb à região não codificante (composta por 10 introns e uma parte da região 3'UTR). Não há evidências de alterações na organização estrutural exon-intron. Na região codificante, apenas substituições nucleotídicas foram detectadas, sendo 48 sinônimas e 49 não sinônimas. E, na região não codificante, além de substituições, também foi observada a presença de *indels*.

Elevada diversidade haplotípica foi observada (0,992), apesar de uma diversidade nucleotídica baixa (0,0899). Maior diversidade foi detectada nos introns em relação ao CDS e, dentre as regiões de ocorrência da espécie, “América Central” foi a que apresentou maior diversidade nucleotídica.

Apesar de observada uma distribuição simpátrica dos clones *MetE* de indivíduos de ipeca através da rede de haplótipos e relações de ancestralidade comum das sequências dentro de cada região serem inferidas pela árvore filogenética bayesiana de CDS, as análises de AMOVA e Fst indicam uma baixa estruturação em relação à formação de três grupos geneticamente divergentes por região de ocorrência, diferentemente do padrão obtido para ITS e cpDNA em que alta divergência genética é observada.

Valores negativos e significativos dos testes de neutralidade (D de Tajima e F de Fu e Li) foram obtidos para *MetE* (região “Amazônia” e exon11) e para ITS (“Mata Atlântica” e amostras como um todo). Valores não significativos foram obtidos para cpDNA. As diferenças quanto aos desvios da hipótese de evolução neutra e quanto às diversidades nucleotídicas e haplotípicas entre os três diferentes *loci* indicam que eventos demográficos não são a causa da variação observada e que os três *loci* estão evoluindo independentemente e sob efeito de distintas forças evolutivas.

Análises sitio-por-sítio da presença de seleção no locus *MetE*, mostraram nove sítios com forte evidência de seleção negativa ($dN/dS < 1$) e quatro sítios com forte evidência de seleção positiva ($dN/dS > 1$). De modo geral, as análises fornecem evidências para a presença de seleção purificadora/negativa atuando sobre as cópias gênicas de *MetE* em ipeca, favorecendo a conservação da estrutura e funcionalidade proteica e sugerindo elevada adaptabilidade desta enzima nas três regiões de ocorrência da espécie.

V) REFERÊNCIAS

BATISTA F.R.C. **Caracterização molecular da região codificadora do gene de metionina sintase em *Carapichea ipecacuanha* (Brot.) L. Andersson.** Monografia apresentada à Universidade Federal de Viçosa – Brasil. 2010.

BRANDÃO, M.G.L. *et al.* **Brazilian medicinal plants described by 19th century European naturalists and in the Official Pharmacopoeia.** Journal of Ethnopharmacology, v.120 , p.141–148. 2008.

CHIANG, Y-C. *et al.* **Nucleotide sequence diversity at the methionine synthase locus in endangered *Dunnia sinensis* (Rubiaceae): An evaluation of the positive selection hypothesis.** Molecular Biology and Evolution, v.19, n.8, p.1367–1375. 2002.

DELPORT, W. *et al.* **Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology.** Bioinformatics, v.26, n.19, p.2455-2457. 2010.

EICHEL, J. *et al.* **Vitamin-B12-independent methionine synthase from a higher plant (*Catharanthus roseus*): Molecular characterization, regulation, heterologous expression, and enzyme properties.** European Journal of Biochemistry, n.230, p.1053-1058. 1995.

EXCOFFIER, L. *et al.* **Arlequin (version 3.0): An integrated software package for population genetics data analysis.** Evolutionary Bioinformatics Online, v.1, p. 47–50. 2005.

FAY, J.C. & WU, C.I. **Hitchhiking under positive Darwinian selection.** Genetics, v.155, n.3, p.1405-1413. 2000.

FERRER, J-L. *et al.* **Crystal structures of cobalamin-independent methionine synthase complexed with zinc, homocysteine, and methyltetrahydrofolate.** The Journal of Biological Chemistry, v.279, n.43, p. 44235–44238. 2004.

FU, Y.X. & LI, W.H. **Statistical tests of neutrality of mutations.** Genetics, v.133, n.3, p. 693-709. 1993.

GAKIÈRE, B. *et al.* **Characterization of the cDNA and gene for a cytosolic cobalamin-independent methionine synthase in *Arabidopsis thaliana* (Accession No. U97200). (PGR99-115).** Plant Physiology, n.120, v.4, p.1206. 1999.

HESSE, H. *et al.* **Current understanding of the regulation of methionine biosynthesis in plants.** Journal of Experimental Botany, v.55, n. 404. 2004.

HESSE, H. & HOEFGEN, R. **Molecular aspects of methionine biosynthesis.** TRENDS in Plant Science, v.1.8, n.6, p.259-262. 2003.

KREITMAN, M. **Methods to detect selection in populations with applications to the human.** Annual Review of Genomics and Human Genetics, v.1, p.539-59. 2000.

KRYAZHIMSKIY, S. & PLOTKIN J.B. **The population genetics of dN/dS.** PLoS Genet, v4, n.12. 2008.

- LIBRADO, P. & ROZAS, J. **DnaSP v5: A software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics*, v.25, p.1451-1452. 2009.
- MURRELL, B.S. **Improved models of biological sequence evolution.** Dissertation approved for the degree of Doctor of Philosophy in Computer Science at Stellenbosch University - South Africa. 2012.
- MITTON, J.B. **Molecular approaches to population biology.** *Annual Review of Ecology and Systematics*, v.25, p.45-69. 1994.
- NARITA Y. *et al.* **Characterization of the salt-inducible methionine synthase from barley leaves.** *Plant Science*, v.167, n.5, p. 1009-1016. 2004.
- NYLANDER, J.A.A. **MrModeltest v2.** Program distributed by the author. Evolutionary Biology Centre, Uppsala University. 2004.
- OLIVEIRA, L.O. *et al.* **Molecular phylogeography of *Carapichea ipeacuanha*, an amphitropical shrub that occurs in the understory of both semideciduous and evergreen forests.** *Molecular Ecology*, v.19, p.1410–1422. 2010.
- OLIVEIRA, L.O. & MARTINS, E.R. **A quantitative assessment of genetic erosion in ipecac (*Psychotria ipeacuanha*).** *Genetic Resources and Crop Evolution*, v.49, p.607–617. 2002.
- PEJCHAL, R. & LUDWIG, M. **Colabamin-independent methionine synthase (MetE): A face-to-face double barrel that evolved by gene duplication.** *PloS Biology*, v.3, n.2, p1-12, 2005.
- POND, S.L.K., *et al.* **GARD: a genetic algorithm for recombination detection.** *Bioinformatics Applications Note*, v.22, n.24, p.3096-3098. 2006.
- POND, S.L.K. & FROST, S.D.W. **Not so different after all: A comparison of methods for detecting amino acid sites under selection.** *Molecular Biology and Evolution*, v.22, p.1208-1222. 2005.
- POSADA, D. & CRANDALL, K.A. **Intraspecific gene genealogies: trees grafting into networks.** *TRENDS in Ecology & Evolution*, v.16, n.1, p. 37-45. 2001.
- QUANG, N.D. *et al.* **Nucleotide variation at the methionine synthase locus in an endangered tree species, *Fokienia hodginsii* (Cupressaceae) in Vietnam.** *African Journal of Biotechnology*, v.11, n.88, p.15398-15404. 2012.
- QUANG, N.D. *et al.* **Patterns of nucleotide diversity at the methionine synthase locus in fragmented and continuous populations of a wind-pollinated tree, *Quercus mongolica* var. *crispula*.** *Journal of Heredity*, v.100, n.6, p.762-770. 2009.
- QUEIROZ, C.S. *et al.* **Evolution of the 5.8S nrDNA gene and internal transcribed spacers in *Carapichea ipeacuanha* (Rubiaceae) within a phylogeographic context.** *Molecular Phylogenetics and Evolution*, v.59, p.293–302. 2011.
- RAVANEL, S. *et al.* **Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol.** *The Journal of Biological Chemistry*, v.279, n.21, p.22548-22557. 2004.

- RAVANEL, S. *et al.* **The specific features of methionine biosynthesis and metabolism in plants.** PNAS, v.95, n.13, p.7805-7812. 1998.
- RONQUIST, F. *et al.* **MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space.** Systematic Biology, v.61, p.1-4. 2012.
- ROSSI, A.A.B. *et al.* **Genetic diversity and geographic differentiation of disjunct Atlantic and Amazonian populations of *Psychotria ipecacuanha* (Rubiaceae).** Genetica, v.136, p.57-67. 2009.
- SANG T. **Utility of Low-Copy Nuclear Gene Sequences in Plant Phylogenetics.** Critical Reviews in Biochemistry and Molecular Biology, v.37, n.3, p.121–147. 2002.
- SKORUPA, L.A. & ASSIS, M.C. **Collecting and conserving ipecac (*Psychotria ipecacuanha*, Rubiaceae) germplasm in Brazil.** Economic Botany, v.52, p.209-210. 1998.
- TAJIMA, F. **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** Genetics, v.123, n.3, p.585-595. 1989.
- TAMURA, K. *et al.* **MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** Molecular Biology and Evolution, v.28, p. 2731-2739. 2011.
- XU, G. *et al.* **Divergence of duplicated gene in exon-intron structure.** PNAS, v.109, n.4, p.1187-1192. 2012.
- WHO (World Health Organization). **Radix Ipecacuanhae.** Monographs on selected medicinal plants, v. 3, p.204-218. 2007.
- WRIGHT S.I. & GAUT, B.S. **Molecular Population Genetics and the Search for Adaptive Evolution in Plants.** Molecular Biology and Evolution, v.22, n.3, p.506–519. 2005.
- ZEH, M. *et al.* **Cloning and characterization of a cDNA encoding a cobalamin-independent methionine synthase from potato (*Solanum tuberosum* L.).** Plant Molecular Biology, v.48, p. 255–265. 2002.