

CAMILA DE SOUSA QUEIROZ

VARIAÇÃO INTRAGENÔMICA DE rDNA EM *Carapichea ipecacuanha*
(RUBIACEAE): EVOLUÇÃO EM CONCERTO INCOMPLETA E PSEUDOGENES

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

2010

CAMILA DE SOUSA QUEIROZ

VARIAÇÃO INTRAGENÔMICA DE rDNA EM *Carapichea ipecacuanha*
(RUBIACEAE): EVOLUÇÃO EM CONCERTO INCOMPLETA E PSEUDOGENES

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: 24 de fevereiro de 2010

Prof.^a Karla Suemy Clemente Yotoko
(Co-orientadora)

Prof. Jorge Abdala Dergam dos Santos
(Co-orientador)

Prof.^a Tânia Maria Fernandes Salomão

Prof.^a Juliana Lopes Rangel Fietto

Prof. Luiz Orlando de Oliveira
(Orientador)

*Aos meus pais Vera e Arcanjo,
dedico mais esta conquista.*

AGRADECIMENTOS

Agradeço a Deus pela vida e amparo nos momentos difíceis;

Aos meus pais, Vera e Arcanjo, e aos meus irmãos, Lucas e Laura, pelo amor incondicional e por todos os sacrifícios ao longo de mais esta jornada;

Ao meu namorado Jayme pela compreensão e carinho;

Aos amigos queridos, especialmente Fausto, Talitha, Juliana e Mariana, pelo apoio e entendimento da minha ausência;

Ao Prof. Luiz Orlando de Oliveira pela orientação;

Ao Prof. Lyderson Facio Viccini por auxiliar meus primeiros passos científicos;

À Prof. Karla Suemy Clemente Yotoko pela disposição e amizade;

Às colegas de laboratório, especialmente Magali, Maíra, Roberta, Beatriz, Maria Andréia e Flávia, pelo apoio nas rotinas de laboratório e pelo companheirismo.

Aos demais professores, funcionários e colegas de curso pelo convívio;

À Universidade Federal de Viçosa e ao programa de pós-graduação em Genética e Melhoramento pela estrutura oferecida;

Ao CNPq pela bolsa de estudos concedida;

A todos que de alguma forma acreditaram em mais um sonho e, direta ou indiretamente, contribuíram para sua realização.

BIOGRAFIA

Camila de Sousa Queiroz, filha de Arcanjo Castilho de Queiroz e Veralúcia Bibiana de Sousa Queiroz, nasceu no dia 16 de janeiro de 1986 em Juiz de Fora - MG. Concluiu o ensino fundamental em 2000 no Colégio Militar de Juiz de Fora e o ensino médio na mesma instituição em 2003. Gradou-se no Curso de Ciências Biológicas da Universidade Federal de Juiz de Fora na modalidade licenciatura em julho de 2007 e na modalidade bacharelado em dezembro do mesmo ano. Em março de 2008 ingressou no curso de Mestrado em Genética e Melhoramento pela Universidade Federal de Viçosa.

SUMÁRIO

RESUMO.....	vi
ABSTRACT	vii
INTRODUÇÃO.....	1
MATERIAL E MÉTODOS.....	4
<i>Amostras</i>	4
<i>Amplificação, clonagem e seqüenciamento</i>	6
<i>Alinhamento, delimitação das regiões e haplótipos</i>	7
<i>Análise da região 5.8S</i>	9
<i>Análise da região ITS2</i>	10
<i>Análise da região ITS1</i>	10
RESULTADOS	11
<i>Caracterização da região ITS</i>	11
<i>Identificação de pseudogenes no 5.8S rDNA</i>	12
<i>Identificação de pseudogenes no ITS2</i>	19
DISCUSSÃO.....	23
<i>Estrutura secundária e identificação de pseudogenes do 5.8S em C. ipecacuanha</i>	23
<i>Estrutura secundária e identificação de pseudogenes de ITS2 em C. ipecacuanha</i>	25
<i>Implicações da distribuição de pseudogenes entre os haplótipos amostrados</i>	26
<i>Varição intragenômica e evolução em concerto</i>	27
CONCLUSÕES.....	29
REFERÊNCIAS BIBLIOGRÁFICAS	30

RESUMO

QUEIROZ, Camila de Sousa, M.Sc., Universidade Federal de Viçosa, fevereiro de 2010. **Varição intragenômica de rDNA em *Carapichea ipecacuanha* (Rubiaceae): evolução em concerto incompleta e pseudogenes.** Orientador: Luiz Orlando de Oliveira. Co-orientadores: Karla Suemy Clemente Yotoko e Jorge Abdala Dergam dos Santos.

A investigação da variação intra-individual do cistron de rDNA ITS em *Carapichea ipecacuanha* foi realizada a partir de seqüenciamento direto de fragmentos de PCR de 26 indivíduos e do seqüenciamento de 240 clones, representando outros 51 indivíduos. As estruturas secundárias das regiões 5.8S e ITS2 apresentam as configurações típicas presentes em angiospermas respectivamente em 20 e 48 haplótipos encontrados. A comparação com os motivos conservados em angiospermas e as estruturas secundárias obtidas para a região 5.8S, indicadores mais confiáveis da funcionalidade das seqüências em *C. ipecacuanha*, permitiram a identificação de 13 pseudogenes entre os 26 haplótipos de 5.8S encontrados. A análise da estrutura secundária do ITS2 adicionou um haplótipo ao conjunto de pseudogenes. A perda de funcionalidade em alguns haplótipos deve ser mais antiga do que em outros em razão do acúmulo de maior número de substituições e da formação de grupos de pseudogenes pertencentes a populações distintas.

ABSTRACT

QUEIROZ, Camila de Sousa, M.Sc., Universidade Federal de Viçosa, February 2010. **Intragenomic variability of rDNA in *Carapichea ipecacuanha* (Rubiaceae): incompleted concerted evolution and pseudogenes.** Adviser: Luiz Orlando de Oliveira. Co-advisers: Karla Suemy Clemente Yotoko and Jorge Abdala Dergam dos Santos.

The investigation of intra-individual variation of the rDNA cistron ITS in *Carapichea ipecacuanha* was made from direct sequencing of 26 individuals and from sequencing of 240 clones, representing another 51 individuals. The secondary structures of 5.8S and ITS2 regions show the typical configurations present in angiosperms respectively for 20 and 48 haplotypes. The comparison with the conserved motifs in angiosperms and the secondary structures obtained for the 5.8S region, the most reliable markers of functional sequences in *C. ipecacuanha*, allowed the identification of 13 pseudogenes among 26 haplotypes of 5.8S. The analysis of the secondary structure of ITS2 added one haplotype to the set of pseudogenes. The loss of functionality in some haplotypes has occurred prior to the others copies due to the accumulation of a larger number of substitutions and the formation of groups of pseudogenes that belong to distinct populations.

INTRODUÇÃO

A utilização da região ITS para estudo das relações entre gêneros, espécies e populações é recorrente na literatura. Formada pelos DNAs ribossomais (rDNAs) 18S, 5.8S e 28S e pelos espaçadores internos ITS1 e ITS2 (figura 1), esta unidade de transcrição se repete no genoma de centena a milhares de vezes em tandem (Long e Dawid, 1980) e é um dos componentes essenciais da região organizadora do nucléolo.

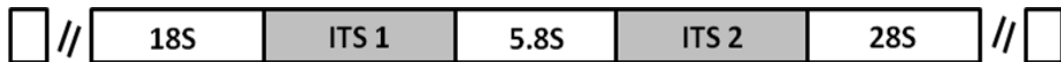


Figura 1 – Organização interna da região ITS em eucariotos, com os rDNAs 18S, 5.8S e 28S e os espaçadores internos ITS1 e ITS2.

Uma característica fundamental deste conjunto de rDNAs é a evolução em concerto das diferentes cópias, em que todas as repetições presentes no genoma compartilham as mesmas substituições (Álvarez e Wendel, 2003), de tal modo que se pode assumir a igualdade entre elas. Embora a evolução em concerto seja reconhecida, o número de trabalhos que indicam a existência de variação intra-individual é crescente (por exemplo, Yang et al., 1999; Hughes et al., 2002; Ochieng et al., 2007; Xu et al., 2009; Oliveira et al., 2010). Os polimorfismos encontrados em um mesmo indivíduo podem ser decorrentes da localização de conjuntos de rDNA em cromossomos não homólogos, da hibridização interespecífica ocorrida na formação de poliplóides ou quando a evolução em concerto é mais lenta do que a taxa de mutação, conforme resumido por Xu et al. (2009). Além disso, tem sido observado que a heterogeneidade de cópias dentro de um mesmo genoma pode incluir tanto cópias funcionais quanto não funcionais (por exemplo, Buckler et al.,

1997; Yang et al., 1999; Hughes et al., 2002; Ochieng et al., 2007; Zheng et al., 2008; Xu et al., 2009).

A rubiácea *Carapichea ipecacuanha* (Brot.) L. Andersson [= *Psychotria ipecacuanha* (Brot.) Stokes; *Cephaelis ipecacuanha* (Brot.) A. Rich.] é uma das espécies para a qual já foi descrita a existência de variação intragenômica na região ITS (Oliveira et al., 2010). Conhecida popularmente por ipeca, esta espécie tem ampla distribuição geográfica, que inclui populações disjuntas na Floresta Atlântica, Floresta Amazônica, Colômbia e América Central (Oliveira et al., 2010). A maior parte dos estudos com *C. ipecacuanha* está concentrada nas propriedades farmacológicas dos alcalóides encontrados em suas raízes, principalmente a emetina e a cefaelina. Seu uso na medicina tradicional iniciou com nativos brasileiros bem antes da chegada dos colonizadores europeus (Veloso 1947), mas o foco atual tem sido o estudo da emetina para uso no tratamento de câncer, uma vez que sua capacidade de induzir a apoptose em linhagens de células humanas cancerígenas já foi reportada (Watanabe et al., 2002; Boon-Unge et al., 2007; Möller et al., 2007).

A degradação das áreas de ocorrência natural e a exploração de *C. ipecacuanha* para fins terapêuticos tornam imprescindíveis os estudos que visam acessar a variabilidade genética e a sua organização dentro da distribuição geográfica. A partir de marcadores ISSR, Rossi et al. (2009) compararam a diversidade genética entre as populações de *C. ipecacuanha* da Floresta Atlântica e da Floresta Amazônica e encontraram a maior porcentagem de polimorfismos nas populações da Floresta Atlântica, indicando a existência de maior variabilidade neste local. A análise filogeográfica para a espécie, baseada na região ITS, revelou a ocorrência de 23 haplótipos não compartilhados entre as áreas de ocorrência, dos quais 20 estão distribuídos entre populações da Floresta Atlântica, organizados em duas linhagens distintas (Oliveira et al., 2010).

Visto que análises prévias da região de ITS de *C. ipecacuanha* revelaram que a distribuição da diversidade genética é desigual entre as áreas de ocorrência natural e que há variação intra-individual em ITS, este estudo pretende investigar mais detalhadamente esta variação, buscando responder às seguintes questões: (1) qual seria a estrutura secundária típica do rRNA 5.8S nesta espécie?; (2) qual seria a conformação da estrutura secundária do RNA de ITS2?; (3) seria possível diferenciar as cópias funcionais das cópias não-funcionais (pseudogenes) a partir de pelo menos uma dessas regiões?; (4) e, se identificados, como estes pseudogenes estariam distribuídos entre os haplótipos amostrados?; (5) finalmente, quais fatores poderiam estar relacionados à manutenção das diferentes cópias desta região no genoma?

MATERIAL E MÉTODOS

Amostras:

Este trabalho utilizou 51 seqüências da região ITS depositadas no GenBank (GU361045-GU361109) e que foram obtidas de 17 indivíduos seqüenciados diretamente e 10 indivíduos clonados (Oliveira et al. 2010). No entanto, parte das seqüências de seqüenciamento direto não continha a região ITS1-5.8S-ITS2 completa. Para estes indivíduos, a amplificação e o seqüenciamento foram refeitos para completar as bases que não estavam presentes nas seqüências depositadas no GenBank. O presente estudo adicionou seqüências de outros 50 indivíduos, totalizando assim 77 indivíduos. A tabela 1 lista as populações selecionadas e suas localizações, a distribuição dos indivíduos por população e por área de ocorrência e o número de indivíduos com seqüências obtidas por seqüenciamento direto ou por clonagem.

Para obtenção das seqüências da região ITS, foram usadas amostras de DNA de *C. ipecacuanha* previamente extraídas e conservadas em TE 0,1 mM à -20°C, pertencentes ao banco de DNA mantido pelo Laboratório de Biologia Molecular e Filogeografia (BIOAGRO – Universidade Federal de Viçosa – Viçosa, Minas Gerais, Brasil). Os indivíduos escolhidos são representativos das populações em toda a distribuição geográfica da espécie. Conforme já exposto, as populações da Floresta Atlântica apresentam maior variabilidade e, por isso, foram amostradas em maior quantidade.

Tabela 1 – Populações amostradas, incluindo as abreviações adotadas, número de indivíduos usados por população e localização.

Populações	Amostras		Coordenadas	
	Oliveira et al. (2010) SD ^a /CL ^b	Este estudo SD ^a /CL ^b	Latitude	Longitude
Floresta Atlântica				
Faria Lemos (FRL)	0/0	0/2	20°48'41"	42°01'49"
Ponte Nova (PTN)	0/0	0/1	20°25'26"	42°53'42"
Dores de Guanhões (DOG)	0/1	0/1	19°02'17"	42°57'27"
Guaraciaba (GUA)	2/0	0/3	20°34'14"	43°00'28"
Raposo (RAP)	0/2	1/10	21°06'40"	42°05'43"
Trilha Lagoa do Meio (TLM)	0/1	0/3	19°38'34"	42°30'56"
Trilha do Vinhático (TVI)	1/0	0/0	19°45'44"	42°28'56"
Trilha Central (TCE)	1/0	2/1	19°33'44"	42°37'57"
Visconde do Rio Branco (VRB)	0/3	6/6	21°00'37"	42°50'26"
subtotal	4/7	9/27		
Floresta Amazônica				
Colorado (COL)	1/0	0/3	11°14'55"	64°24'37"
Exu (EXU)	0/0	0/1	15°40'41"	57°32'03"
Figueirópolis (FIG)	2/0	0/1	15°26'37"	58°47'47"
Lambari D'oeste (LAM)	0/1	0/1	15°19'27"	58°00'10"
Mozart (MOZ)	1/0	0/0	15°04'57"	57°58'00"
Prata (PRA)	1/0	0/0	15°30'26"	58°01'50"
Soroteca (SOR)	2/0	0/0	15°31'34"	58°00'42"
Vila Bela Santíssima Trindade (VBS)	1/0	0/1	15°00'29"	59°57'02"
Abaitará (ABR)	1/0	0/0	11°44'21"	61°31'36"
Sítio Jesus (JES)	1/0	0/0	11°31'29"	61°33'51"
Rio Vermelho (RVE)	1/0	0/2	15°17'38"	57°51'43"
Tangará da Serra (TAN)	0/1	0/3	14°04'38"	57°03'45"
subtotal	11/2	0/12		
América Central				
Panamá (BCI)	2/1	0/2	79°50'39"	9°9'38"
subtotal	2/1	0/2		
Total	17/10	9/41		

^a SD - indivíduos dos quais foram obtidas seqüências por meio de seqüenciamento direto

^b CL - indivíduos dos quais foram obtidas seqüências por meio de clonagem

Amplificação, clonagem e seqüenciamento:

A amplificação de parte do cistron de DNA ribossomal 18S-5.8S-28S, incluindo seus espaçadores internos, foi realizada com o auxílio dos primers ITS4 (5' TCCTCCGCTTATTGATATGC 3') e ITS5 (5' GGAAGTAAAAGTCGTAACAAGG 3') de White et al. (1990). Cada reação foi preparada em um volume total de 25 µl, contendo tampão na concentração 1X, 1,5 mM de MgCl₂, água, DMSO a 5%, 0,8 mM de deoxinucleotídeos (dNTP), 0,5 µM de cada primer, 1,25 unidades de *Taq* DNA polimerase (Phoneutria) e aproximadamente 60 ng de DNA. As reações foram conduzidas em termociclador Mastercycler (Eppendorf), com uma etapa inicial de 4 minutos a 94°C, seguida de 35 ciclos de 1 minuto a 94°C, 1 minuto a 54°C e mais 1 minuto a 72°C, além de uma etapa final de 72°C por 5 minutos.

O volume de 5 µl de cada produto de amplificação foi submetido à eletroforese em gel de agarose 0,8% corado com brometo de etídio para verificação da amplificação do fragmento de interesse. As imagens foram capturadas sob transiluminação no equipamento L•PIX (Loccus Biotecnologia).

As reações que apresentaram fragmentos de DNA com o tamanho esperado foram purificadas com a enzima ExoSAP IT (USB), na proporção de 9 µl de reação para 3 µl da enzima, também conduzida no termociclador Mastercycler (Eppendorf), em duas fases de 15 minutos: a primeira a 37°C e a segunda a 80°C. Para comprovar a eficácia da purificação, 1 µl do material também foi verificado em gel de agarose 0,8% e a captura da imagem realizada sob transiluminação.

Amostras das reações obtidas com a região de ITS foram enviadas para seqüenciamento (Macrogen, Coréia). Os cromatogramas foram inspecionados visualmente no programa Sequencher 4.8 (Gene Codes) para investigar a presença de picos duplos, indicativo da amplificação de dois ou mais fragmentos distintos durante a reação em cadeia da polimerase e, portanto, da existência de variação

intragenômica. Entre os 50 indivíduos adicionais investigados, 36 apresentaram picos duplos em ambas as fitas complementares e foram então submetidos à clonagem. Também 5 indivíduos que não mostraram cromatogramas com picos duplos foram selecionados aleatoriamente e clonados, visando confirmar a ausência de variação intragenômica.

Assim, para 41 indivíduos, o fragmento amplificado foi inserido em vetor pGEM-T Easy (Promega) e este vetor foi utilizado para a transformação das cepas de *Escherichia coli* DH5 α através de choque térmico a 42°C. Após a seleção das colônias transformadas por marcação branco/azul com X-gal, foi realizada a extração de DNA plasmidial com o kit de purificação da Promega. Para cada indivíduo foram escolhidos de 1 a 6 clones que foram enviados para seqüenciamento (Macrogen, Coréia).

Todos os cromatogramas foram inspecionados visualmente no Sequencher 4.8 (Gene Codes) e as correções necessárias foram feitas manualmente.

Alinhamento, delimitação das regiões e haplótipos:

As 26 seqüências obtidas pelo seqüenciamento direto e as 240 do seqüenciamento de clones foram alinhadas em conjunto no Sequencher 4.8 (Gene Codes). As análises para delimitar as regiões ITS1, 5.8S e ITS2 foram conduzidas por comparação com dados presentes em 3 bancos de dados: "Rfam" (Gardner et al., 2008), "The European Ribosomal Database" (Wuyts et al., 2004) e "The ITS2 database" (Selig et al., 2008). As seqüências foram podadas para remover as seqüências de RNA 18S e 28S adjacentes, de forma que a seqüência resultante inicia-se na primeira base do ITS1 e termina na última base do ITS2, contendo exclusivamente o segmento ITS1-5.8S-ITS2.

O programa DnaSP 5.10 (Librado e Rozas, 2009) foi utilizado para determinação dos haplótipos presente entre as 266 seqüências, da diversidade haplotípica e da diversidade nucleotídica. A rede formada por esses haplótipos foi estimada no Network 4.5.1.6 (<http://www.fluxus-technology.com>) com a opção de “Median joining” (Bandelt et al., 1999).

A maioria dos polimorfismos encontrados não parece refletir erros da *Taq* polimerase durante as reações de amplificação, pois ou eles foram compartilhados entre diferentes clones de um mesmo individuo ou entre indivíduos diferentes.

No entanto, substituições singulares foram observadas em 47 dos 240 clones. Cada uma destas substituições representava a única modificação da seqüência em relação ao haplótipo central mais próximo. Ainda, estas substituições estavam distribuídas aleatoriamente ao longo do segmento ITS1-5.8S-ITS2: sendo que 19 estavam presentes em ITS1, 13 em 5.8S e 15 em ITS2. Considerando a menor restrição funcional de ITS1 e ITS2, o esperado seria que estas duas regiões acumulassem maior número de mutações do que 5.8S. Estas substituições singulares foram consideradas erros da *Taq* polimerase que ocorreram durante a amplificação destas 47 seqüências e que foram revelados por meio da clonagem. Por isto, estas mutações foram revertidas para a seqüência consenso.

O conjunto total de seqüências com a região ITS completa foi subdividido em três bancos de dados distintos: o primeiro contendo somente a região ITS1, o segundo contendo apenas o 5.8S e o terceiro contendo somente o ITS2. As análises subseqüentes foram conduzidas em cada um dos três bancos de dados de forma independente.

Análise da região 5.8S:

A estrutura secundária e a energia livre (ΔG) do 5.8S em *C. ipecacuanha* foram estabelecidas na versão online do programa Mfold 3.2 - RNAfolding (Mathews et al., 1999; Zuker, 2003) com as opções padrão de temperatura (37°C) e de condições iônicas. Ainda, foram feitas restrições de pareamento e não pareamento de acordo com a estrutura secundária conservada de 5.8S em eucariotos (F 46 106 3; F 50 61 3; F 70 97 3; F 111 119 3; P 1 0 45; P 63 0 7; P 73 0 22; P 144 0 18). As estruturas secundárias obtidas para os diferentes haplótipos foram alinhadas com auxílio do 4SALE (Seibel et al., 2008) e editadas para impressão no XRNA 2.0 beta (<http://rna.ucsc.edu/rnacenter/xrna/xrna.html>).

Por meio do aplicativo MEGA4 (Tamura et al., 2007), os haplótipos da região 5.8S foram alinhadas com os três motivos conservados em angiospermas indicados por Harpke e Peterson (2008a) e foi calculado o conteúdo CG.

Utilizamos dois critérios para diferenciar cópias funcionais de não-funcionais de 5.8s em *C. Ipecacuanha*. A formação da estrutura secundária com as cinco hélices conservadas para eucariotos foi considerada indicativo de funcionalidade. Os haplótipos que apresentaram estrutura aberrante em uma ou mais hélices de 5.8S foram tidos como cópias não funcionais e denominados pseudogenes. Ainda, foram indicados como supostos pseudogenes os haplótipos que contivessem uma ou mais substituições dentro de pelo menos um dos motivos conservados analisados.

Análise da região ITS2:

Para a determinação da estrutura secundária e energia livre (ΔG) do ITS2 foi usada a predição online no “The ITS2 database” (Selig et al., 2008) com base no modelo existente para *Psychotria ipecacuanha* (GI 4454748). O alinhamento dessa região no MEGA4 (Tamura et al., 2007) foi feito para a medição do conteúdo CG. Já o alinhamento das estruturas secundárias foi feito no 4SALE (Seibel et al., 2008) e as estruturas foram editadas para impressão no XRNA 2.0 beta (<http://rna.ucsc.edu/rnacenter/xrna/xrna.html>). A presença de estrutura secundária aberrante foi definida como parâmetro para a inferência da não funcionalidade.

Análise da região ITS1:

A predição da estrutura secundária da região ITS1 para as seqüências obtidas neste estudo não foi feita porque ainda não há um consenso sobre a estrutura secundária típica em angiospermas (Harpke e Peterson, 2006). Por isso, esta região não foi usada para a detecção de pseudogenes.

RESULTADOS

Caracterização da região ITS:

O seqüenciamento da região compreendida entre os primers ITS4 e ITS5 resultou num fragmento de aproximadamente 700 pb, contendo a parte final do 18S rDNA, o conjunto completo ITS1-5.8S-ITS2 e a parte inicial do 28S rDNA.

A análise preliminar dos cromatogramas do seqüenciamento direto da região ITS mostrou a existência de variação intragenômica em 46 dos 77 indivíduos avaliados, nos quais foram identificados picos duplos no cromatograma de ambas as fitas complementares, o que indica a sobreposição de dois ou mais fragmentos distintos na reação de seqüenciamento. Foram usadas 266 seqüências no total, 26 a partir do seqüenciamento direto e 240 do seqüenciamento de clones. Entre estas 266 seqüências, 172 seqüências são de indivíduos da Floresta Atlântica, 78 da Floresta Amazônica e 16 do Panamá.

Após a delimitação da região ITS1-5.8S-ITS2, foram obtidas 28 seqüências de 581pb e 238 de 582bp. A variação de tamanho encontrada é devido à presença de indels de 1pb únicos em cada seqüência, nas posições 31, 70, 147, 169, 377, 455, 511, 550, 557, 568 e 575 do alinhamento final. A região ITS1 inicia na 1ª base do alinhamento e se estende até a 196ª, variando entre 195 e 196pb; o 5.8S começa na 197ª e termina na 357ª base do alinhamento, com tamanho fixo de 161bp; e o ITS2 inicia na 358ª base alinhada e termina na 582ª, com tamanho entre 224 e 225bp. Foram identificados 185 sítios variáveis entre os 582 sítios alinhados, dos quais 70 estão no ITS1, 41 no 5.8S e 74 no ITS2.

A estrutura secundária padrão de 5.8S, com a formação das cinco hélices conservadas, está presente em 20 dos 26 haplótipos avaliados (76,92%) de *C. ipecacuanha* (figura 3). No entanto, seis haplótipos (H3, H5, H7, H12, H23 E H24) apresentam substituições não compensadas que provavelmente impedem a formação de pelo menos uma das hélices (figura 4A a F). Conseqüentemente, a manutenção da estrutura secundária requer pareamentos energeticamente desfavoráveis (por exemplo, entre duas guaninas) e tais cópias apresentam energias livre (ΔG) diferente do comum encontrado para o 5.8S de *C. ipecacuanha* (-13,70 kcal/mol), com variação entre -0,10 e -14,60 kcal/mol. Deste modo, estes seis haplótipos com formação inadequada das hélices foram considerados portadores de cópias não funcionais de 5.8S.

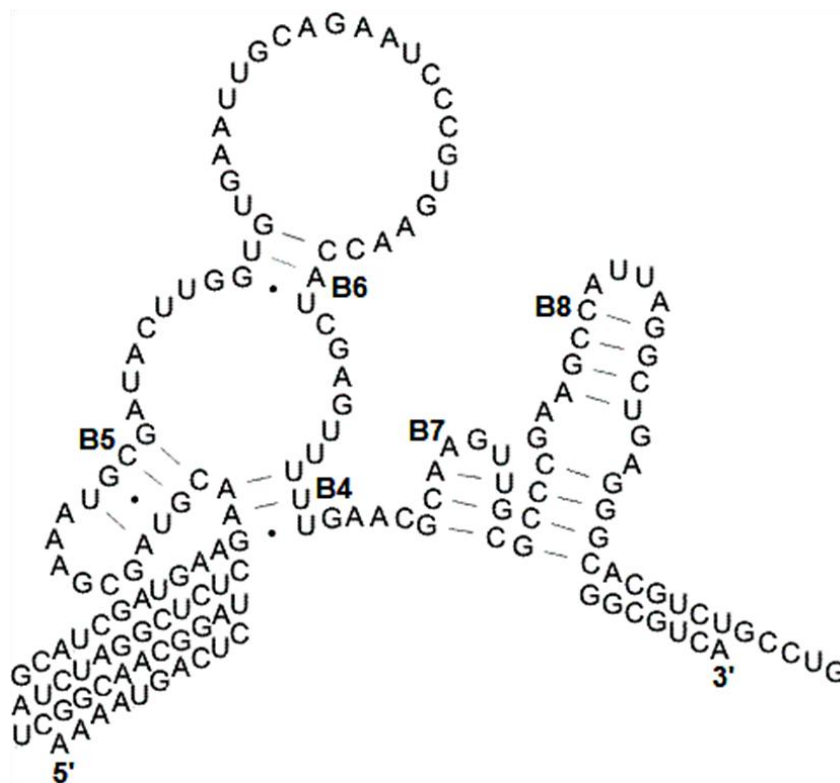


Figura 3 – Estrutura secundária típica do 5.8S em *C. ipecacuanha*, correspondente ao haplótipo H1. Numeração das hélices conservadas de B4 a B8 de acordo com Wuyts et al. (2001).

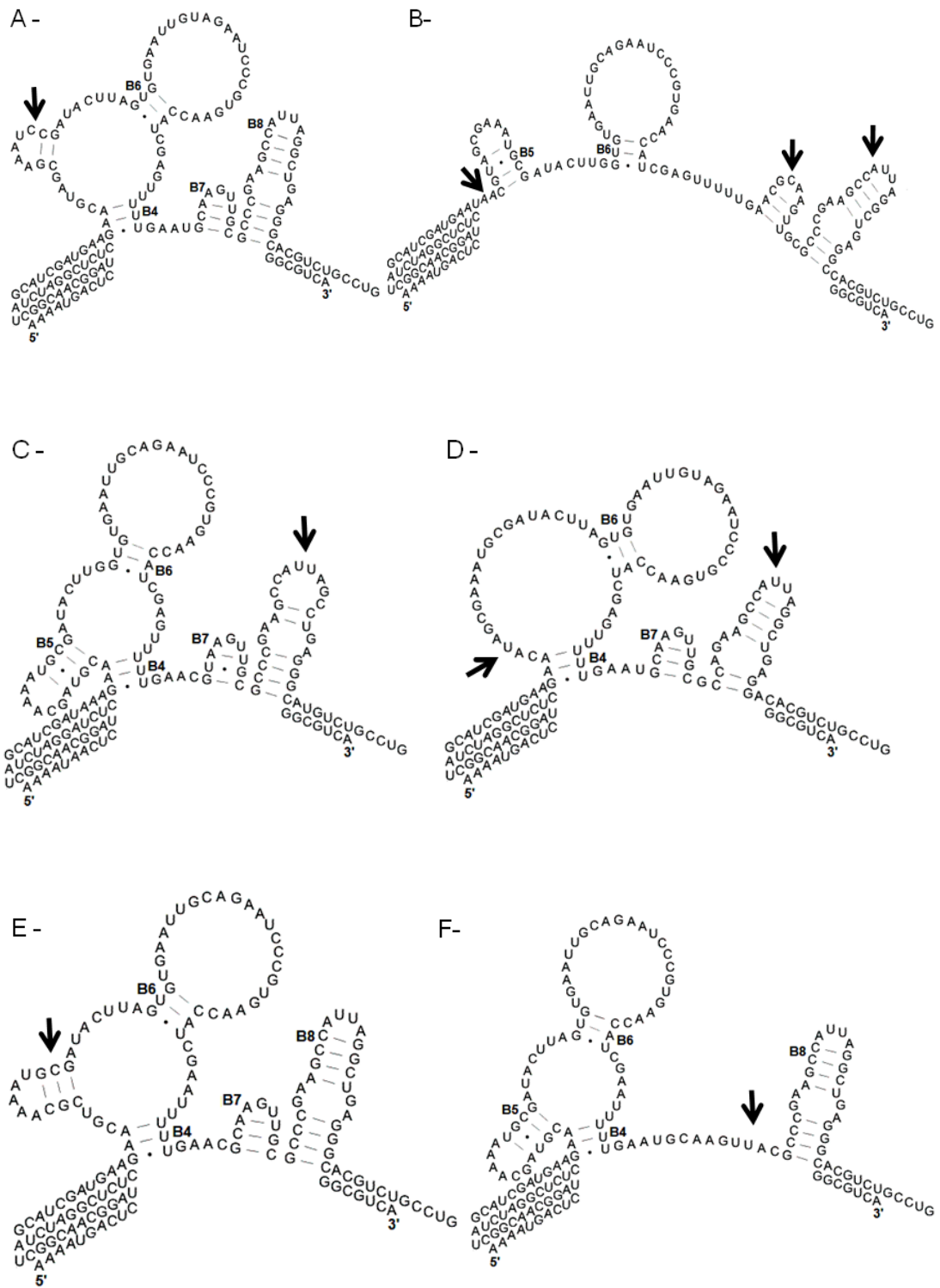


Figura 4 – Estruturas de 5.8S em *C. ipecacuanha* com formação defeituosa das hélices. A – H3; B – H5; C – H7; D – H12; E – H23; e F – H24. Numeração das hélices conservadas de B4 a B8 de acordo com Wuyts et al. (2001). As setas mostram as hélices com formação defeituosa ou inexistente em relação à estrutura típica.

O alinhamento da região 5.8S identificou substituições nos motivos conservados para angiospermas em 12 haplótipos (H3, H4, H5, H6, H7, H11, H12, H14, H16, H17, H21 e H23), incluindo cinco (H3, H5, H7, H12 e H23) dos seis haplótipos que haviam apresentado modificações na estrutura secundária (figura 5). Os sete haplótipos adicionais (H4, H6, H11, H16, H17 e H21), com estrutura secundária estável, mas com substituições dentro dos motivos conservados, também foram considerados supostos pseudogenes.

Haplótipos	Motivos conservados		
	[1]	[2]	[3]
	CGATGAAGAACGTAGC	GAATTGCAGAATCC	TTTGAAYGCA
H1, H2, H8, H9, H10, H13, H15, H18, H19, H20, H22, H25, H26C....
H3*T.....T....
H4*T.....C....
H5*T.....C....
H6*C.....C....
H7*A.....C.T...
H11*G.C....
H12*A.....T.....T....
H14*T.....T....
H16*A.....C....
H17*C.....C....
H21*C.....C....
H23*C.....C....
H24T....

Figura 5 - Alinhamento dos haplótipos de 5.8S em *C. ipecacuanha* com os sítios conservados descritos por Jobes e Thien (1997) e Harpke e Peterson (2008a), onde * indica os haplótipos com substituições dentro destes sítios.

No total, foram encontrados 13 haplótipos contendo cópias potencialmente não funcionais de 5.8S: H3, H4, H5, H6, H7, H11, H12, H14, H16, H17, H21, H23 e H24. Os parâmetros avaliados de estrutura secundária e motivos conservados para o 5.8S estão resumidos na tabela 2. Os haplótipos de pseudogenes têm maior número de sítios variáveis e maior diversidade nucleotídica quando comparados aos haplótipos considerados funcionais (tabela 3).

Tabela 2 – Comparação dos parâmetros avaliados para o 5.8S de *C. ipecacuanha* entre as cópias identificadas como pseudogenes e as cópias funcionais.

Haplótipos	Energia livre ΔG (kcal/mol)	Motivos conservados ^a	Conteúdo CG(%)	Estrutura secundária ^b	Pseudogene ^c
H1	-13,70	xxx	51,60	xxxxx	F
H2	-13,70	xxx	52,20	xxxxx	F
H3	-10,70	x-x	49,70	x-xxx	ψ
H4	-13,70	x-x	51,00	xxxxx	ψ
H5	-14,30	-xx	50,90	-xx--	ψ
H6	-13,40	x-x	52,80	xxxxx	ψ
H7	-4,20	-x-	47,90	xxxx-	ψ
H8	-8,90	xxx	51,00	xxxxx	F
H9	-11,00	xxx	51,00	xxxxx	F
H10	-13,70	xxx	52,20	xxxxx	F
H11	-13,70	xx-	52,20	xxxxx	ψ
H12	-0,10	--x	47,80	x-xx-	ψ
H13	-13,60	xxx	52,20	xxxxx	F
H14	-13,30	x-x	49,00	xxxxx	ψ
H15	-13,70	xxx	50,90	xxxxx	F
H16	-13,70	-xx	51,60	xxxxx	ψ
H17	-13,70	x-x	52,80	xxxxx	ψ
H18	-13,70	xxx	52,80	xxxxx	F
H19	-13,70	xxx	52,80	xxxxx	F
H20	-12,30	xxx	49,00	xxxxx	F
H21	-8,90	-xx	51,60	xxxxx	ψ
H22	-13,70	xxx	50,30	xxxxx	F
H23	-14,60	-xx	50,30	x-xxx	ψ
H24	-14,30	xxx	48,50	xxx-x	ψ
H25	-13,70	xxx	50,90	xxxxx	F
H26	-13,70	xxx	52,20	xxxxx	F

^a Presença (x) ou ausência (-) dos motivos 1, 2 e 3, respectivamente (Harpke e Peterson, 2008a)

^b Formação correta (x) ou incorreta (-) das hélices B4,B5, B6, B7 E B8 respectivamente (numeração de acordo com Wuyts et al., 2001)

^c Haplótipos considerados pseudogenes (ψ) ou funcionais (F)

Tabela 3 – Diversidade nucleotídica de 5.8S entre os haplótipos considerados pseudogenes e os haplótipos funcionais de *C. ipecacuanha*.

Haplótipos	Número de sítios variáveis	Número médio de diferenças par a par	Diversidade nucleotídica
Pseudogenes	27	4,808	0,0299
Funcionais	21	3,846	0,0239
Todos	41	4,292	0,0267

A análise da relação entre a energia livre e o conteúdo CG de 5.8S entre cópias funcionais e pseudogenes está mostrado na Figura 6. Os pseudogenes H7 e H12 apresentam as características típicas de cópias de ITS com perda de função: reduzido conteúdo CG e energia livre aumentada.

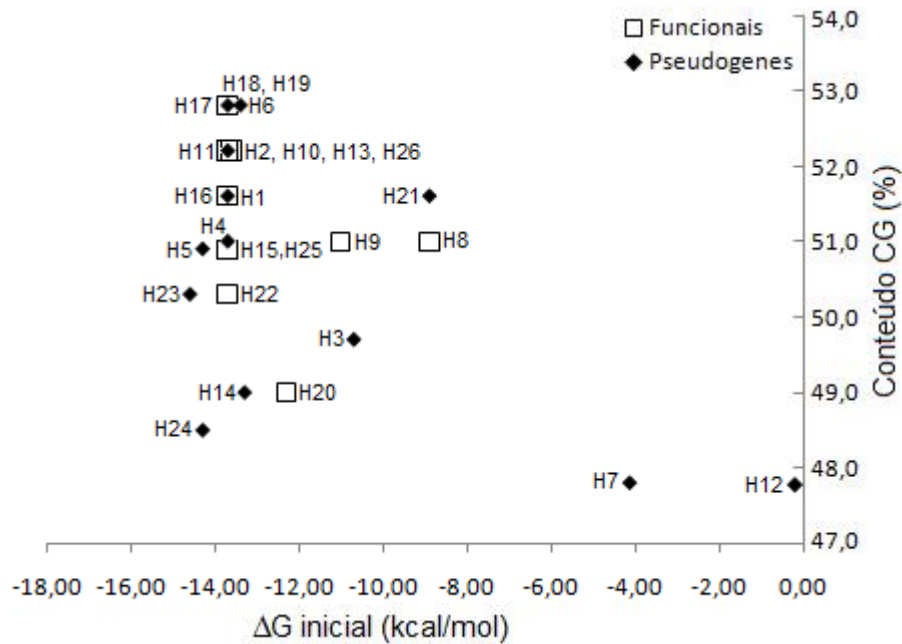


Figura 6 – Relação entre a energia livre e o conteúdo CG calculados para as cópias funcionais e os pseudogenes de 5.8S em *C. ipecacuanha*.

A relação entre os 26 haplótipos encontrados para o 5.8S está demonstrada na figura 7. Claramente, pode-se notar a organização de alguns pseudogenes em linhagens bem definidas, sendo uma delas formada por H6, H16 e H17 e outra por H3, H12, H14, H23 e H24. Entre os 77 indivíduos, 18 contém variação intra-individual de 5.8S, dos quais 9 contém de uma a três cópias de pseudogenes, todos pertencentes a populações da Floresta Atlântica (figura 8).

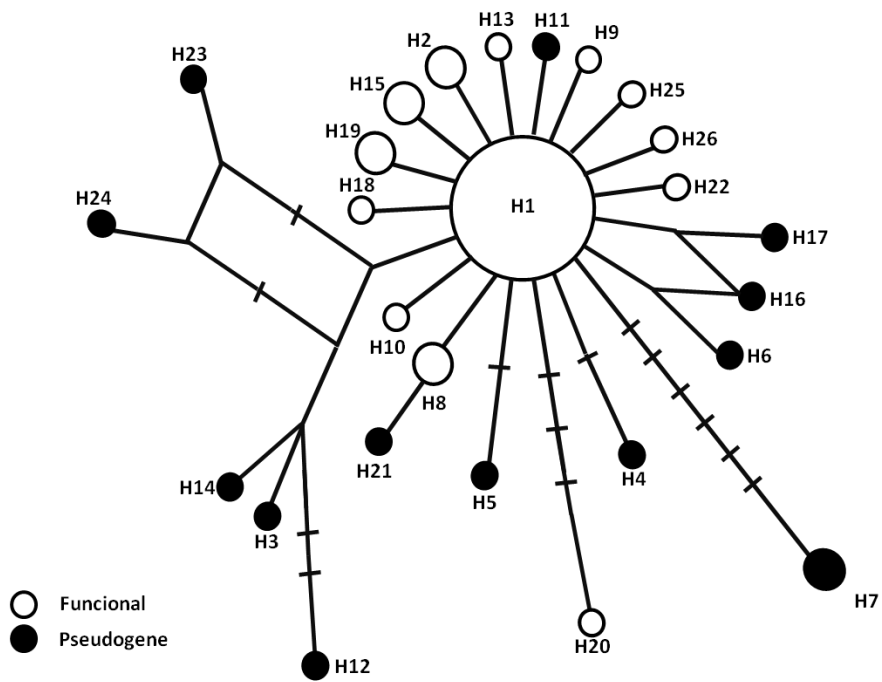


Figura 7 – Rede de haplótipos para a região 5.8S em *C. ipecacuanha*. Cada círculo representa um haplótipo e seu tamanho está diretamente relacionado à quantidade de seqüências agrupadas nele.

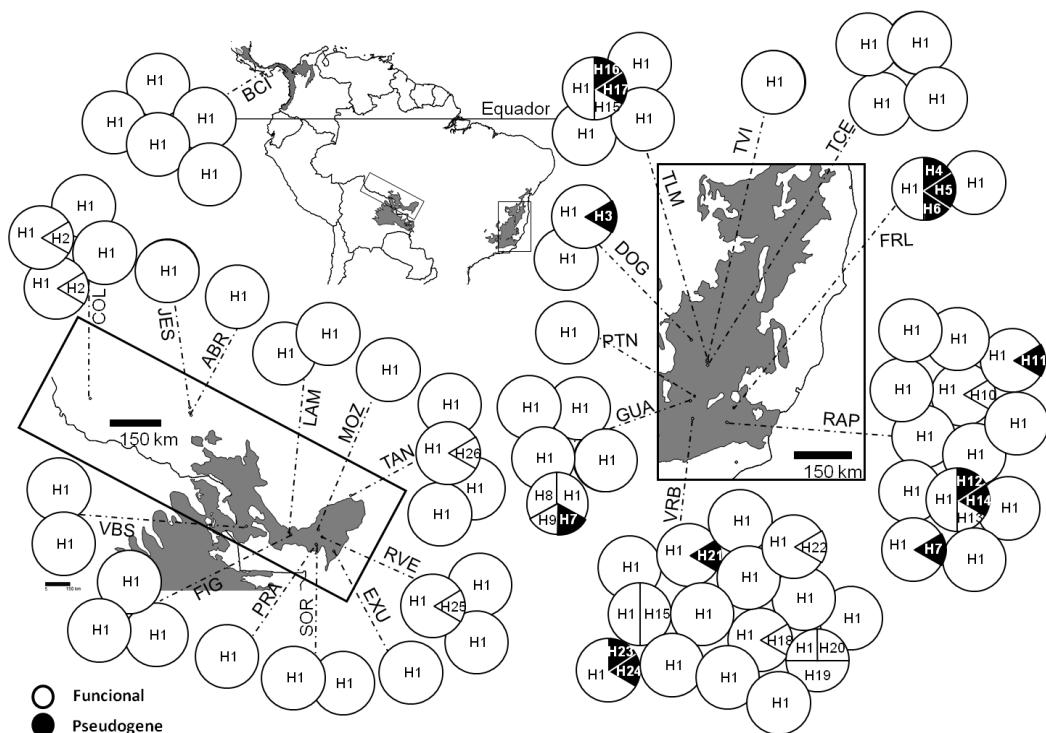


Figura 8 - Distribuição dos haplótipos de 5.8S entre os indivíduos das populações de *C. ipecacuanha* e nas áreas de ocorrência da espécie. A abreviação do nome de cada população segue a tabela 1.

Identificação de pseudogenes no ITS2:

As 266 seqüências foram distribuídas em 51 haplótipos de ITS2, com diversidade haplotípica de 0,8593 (figura 9). A análise de estrutura secundária recuperou a estrutura conservada em 48 dos 51 haplótipos tipos de ITS2. A estrutura padrão de ITS2 para *C. ipecacuanha* inclui as quatro hélices convencionais, com um desencontro de pirimidinas (C-U) na base da hélice II, a hélice III tipicamente maior do que a II e a seqüência conservada 5'NRUGGU3' no lado 5' da hélice III, entre as bases 122 e 127 do alinhamento (figura 10A). Este motivo conservado está presente na forma 5'GGUGGU3' em 100% das seqüências avaliadas. As hélices I e IV apresentaram o maior número de substituições entre os haplótipos, enquanto a hélice II foi a mais conservada. Apesar de a estrutura secundária ser a mesma em 48 haplótipos, a energia livre para sua formação variou entre -44,6 e -71,9 kcal/mol (média -62,40 kcal/mol) e o conteúdo CG entre 59,2 e 62,5% (média 60%).

Os outros três haplótipos de ITS2, presentes nas seqüências correspondentes aos haplótipos de 5.8S H3, H7 e H20, contém modificações não compensadas ou até deleções que alteram a formação correta dos grampos nas hélices (figura 10B a D) e podem ser denominados pseudogenes. Estas seqüências apresentam energia livre menor do que a média encontrada para ITS2 em *C. ipecacuanha*, porém o conteúdo CG está próximo à média (tabela 4). Coincidentemente, os haplótipos H3 e H7 foram considerados pseudogenes de 5.8S, enquanto H20 não é diferenciado como tal nas análises adotadas para o 5.8S.

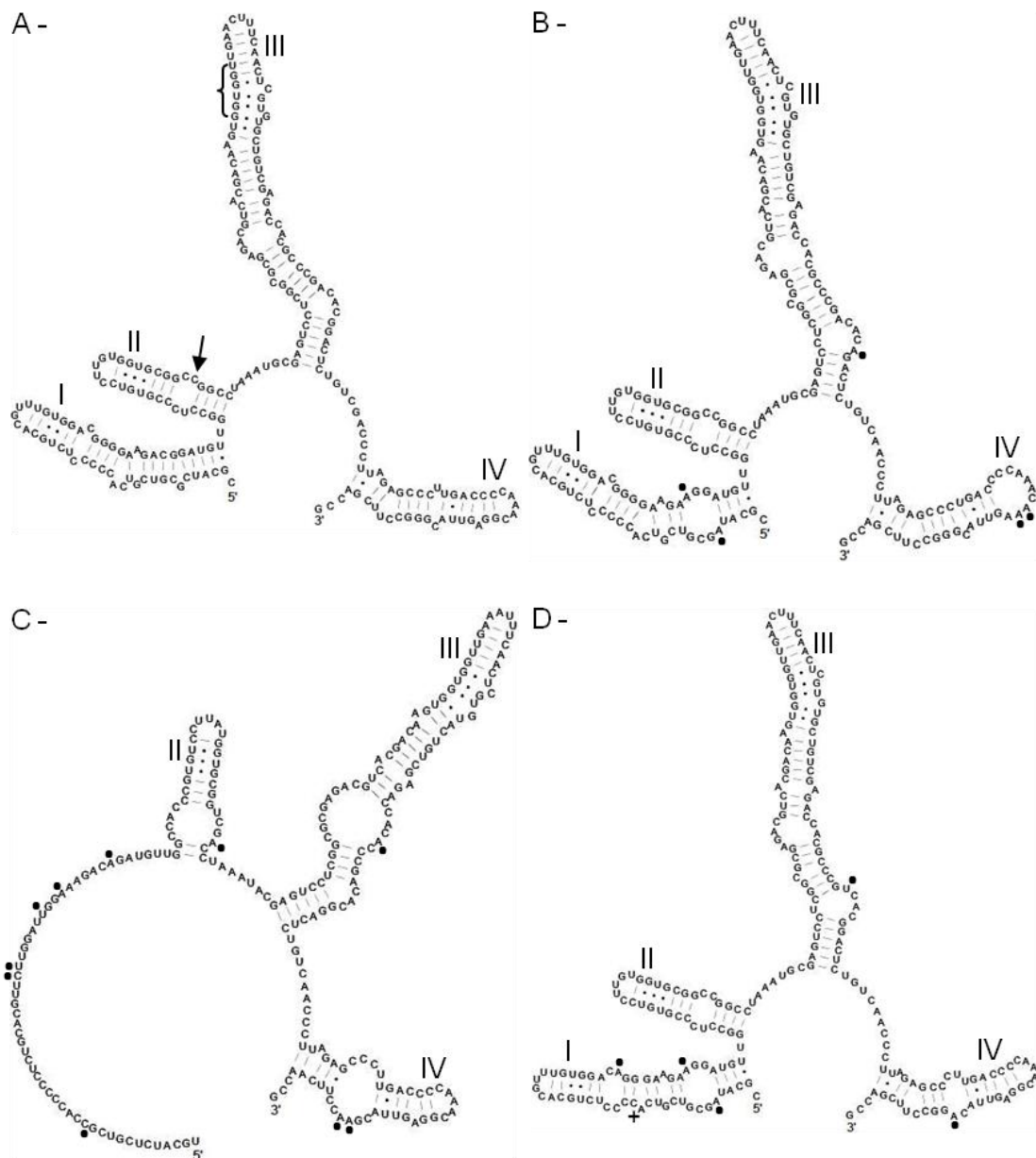


Figura 10 – Estrutura secundária do ITS2 em *C. ipecacuanha*. A – Estrutura padrão, demonstrada com haplótipo 1; B a D – estruturas defeituosas. B – haplótipo 6, correspondente ao haplótipo de 5.8S H3, C – haplótipo 14, correspondente ao haplótipo de 5.8S H7 e D – haplótipo 40, correspondente ao haplótipo de 5.8S H20. A seta em A indica o não-pareamento entre duas pirimidinas na hélice II e a chave, a seqüência conservada NRUGGU na hélice III. Em B a D os pontos ressaltam as substituições não compensadas e + as deleções. Hélices numeradas conforme Coleman (2007).

Tabela 4 - Comparação da energia livre, conteúdo CG e estrutura secundária do ITS2 padrão de *C. ipecacuanha* com os haplótipos de estrutura anômala.

Haplótipo	Energia livre ΔG (kcal/mol)	Conteúdo CG (%)	Estrutura secundária ^b
ITS2 padrão	-62,40 ^a	60,00 ^a	xxxx
6 (equivalente a H3 de 5.8S)	-54,30	59,90	-x--
14 (equivalente a H7 de 5.8S)	- ^c	55,10	----
40 (equivalente a H20 de 5.8S)	-45,80	59,80	-x--

^aMédia entre os valores obtidos para os 48 haplótipos de ITS2 com estrutura conservada.

^bHélices I, II, III e IV sem substituições (x) ou com modificações não compensadas ou deleções (-) (numeração de acordo com Coleman, 2007).

^cNão foi possível calcular devido a não correspondência com os modelos.

DISCUSSÃO

Estrutura secundária e identificação de pseudogenes do 5.8S em C. ipecacuanha:

Detectada a variação intragenômica, torna-se necessária a diferenciação entre as cópias funcionais e as não funcionais (pseudogenes). Os pseudogenes distribuídos pelo genoma podem evoluir em taxas diferentes ou até independentemente de seus correspondentes funcionais. Isto torna a inferência filogenética e outros tipos de análises baseadas na hipótese de ortologia de seqüências ainda mais complexa. São muitos os parâmetros propostos na literatura para a identificação de pseudogenes entre as cópias de ITS: padrão de substituição nucleotídica, variação de tamanho, formação da estrutura secundária e substituições induzidas por metilação (compilados em Bailey et al., 2003).

A estrutura secundária geral de 5.8S é bastante conservada não só entre as angiospermas, como entre todos os eucariotos, já que ela é necessária para o funcionamento correto do ribossomo (Suh et al., 1992). A variação da energia livre observada na formação da estrutura secundária padrão pode ser explicada pelo fato destes haplótipos possuírem seqüências distintas de 5.8S, pois mesmo que as hélices apresentem a mesma configuração, a sua construção depende de pareamentos com diferentes conteúdos energéticos.

Em geral, a energia livre estimada para *C. ipecacuanha* é mais alta do que a determinada para outras espécies. Por exemplo, Harpke e Peterson (2008b) estimaram o valor de -21,40 kcal/mol para as cópias funcionais de *Mammillaria mazatlanensis* e *M. senilis* e Ochieng et al. 2007, -26,44 kcal/mol para *Corymbia intermedia* e -32,62 kcal/mol para *C. eximia*. No entanto, a predição da energia livre para a estrutura secundária está sujeita aos ajustes feitos durante a submissão da

seqüência para dobramento, como a inserção de limitações para pareamentos usadas neste trabalho, o que limita a comparação entre resultados (Harpke e Peterson, 2008b). A comparação dos valores obtidos de energia livre com os de outras espécies mais próximas à *C. ipecacuanha* não foi possível já que, pelo nosso conhecimento, tais dados não estão disponíveis na literatura.

Devido às restrições funcionais da região 5.8S, os seis haplótipos com substituições críticas para o dobramento correto das hélices foram identificados como supostos pseudogenes. Estruturas secundárias incorretas de 5.8S já foram relatadas em outros organismos, como em *Mammillaria* (Harpke e Peterson, 2008b) e em *Corymbia* (Ochieng et al., 2007), e também foram consideradas indicativas de perda da funcionalidade. Entre estes seis haplótipos, cinco contém no mínimo quatro substituições, enquanto o haplótipo H5 possui apenas duas substituições em toda região 5.8S, demonstrando que poucas modificações já são suficientes para desarranjar a formação correta da estrutura secundária.

A comparação da região 5.8S entre os haplótipos e os três motivos conservados em angiospermas (Jobes e Thien, 1997; Harpke e Peterson, 2008a) mostrou substituições críticas em 12 haplótipos (figura 5 e tabela 2). Entre estes, somente cinco haplótipos apresentaram formação instável da estrutura secundária, demonstrando que uma porcentagem muito menor de seqüências do que a prevista por Harpke e Peterson (2008a) perde a capacidade de formar a estrutura secundária adequada quando não possuem um ou mais motivos conservados.

A tendência dos pseudogenes em geral é possuir menor conteúdo CG, resultado de transições G→A e C→T (Grimm e Denk, 2007). Contudo, alguns dos pseudogenes de 5.8S de *C. ipecacuanha* têm conteúdo CG equivalente ou até maior que algumas cópias funcionais de 5.8S. O mesmo já foi observado para algumas espécies de *Mammillaria* (Harpke e Peterson, 2008b), demonstrando que o padrão de composição nucleotídica nem sempre é correlacionado à limitação

funcional (Bailey et al., 2003) e, por isso, não deve ser o único parâmetro de diferenciação entre pseudogenes e cópias funcionais.

Estrutura secundária e identificação de pseudogenes de ITS2 em C. ipecacuanha:

A pressão funcional sobre os espaçadores internos ITS1 e ITS2 é menor do que sobre o 5.8S, embora a maturação dos rRNAs e, conseqüentemente, a formação do ribossomo dependam da estrutura secundária dos espaçadores internos do rRNA (Mai e Coleman, 1997). Deste modo, há predomínio de sítios de nucleotídeos que evoluem de modo essencialmente neutro (Álvarez e Wendel, 2003). Neste caso, as bases que compõe os pareamentos das hélices são objeto de seleção para substituições compensadas que mantenham este pareamento (Álvarez e Wendel, 2003) e, assim, mesmo existindo grandes variações de ITS2 entre os haplótipos amostrados para *C. ipecacuanha* poucos deles apresentam variações na estrutura secundária padrão. Isto também pode explicar o fato de alguns haplótipos identificados pseudogenes para o 5.8S ainda manterem uma estrutura funcional de ITS2.

O haplótipo de 5.8S H20 contém cinco substituições não compensadas e uma deleção na região ITS2 responsáveis por alterações no pareamento para formação das hélices. A sua qualificação como pseudogene de 5.8S não foi possível com as análises utilizadas, mas é notável seu comportamento semelhante a outros pseudogenes de 5.8S em *C. ipecacuanha*, com reduzido conteúdo CG e maior energia livre (figura 6).

Implicações da distribuição de pseudogenes entre os haplótipos amostrados:

As seqüências de pseudogenes divergem progressivamente das seqüências funcionais. Isto ocorre porque, ao escapar das pressões seletivas, elas ficam livres para acumular mutações independentemente das suas contrapartes funcionais. Em *C. ipecacuanha* a análise de diversidade nucleotídica (tabela 3) e a disposição dos haplótipos na rede para o 5.8S (figura 7) corroboram esta afirmação. A maioria dos haplótipos de pseudogenes está distante por duas ou mais substituições do haplótipo de referência H1, enquanto todos os haplótipos funcionais diferem de H1 por apenas uma modificação. Ainda, os haplótipos portadores de cópias não funcionais de 5.8S possuem maior diversidade nucleotídica quando comparados aos haplótipos funcionais.

A quantidade de substituições nucleotídicas acumuladas por um pseudogene é proporcional ao tempo que tiver passado após o escape da limitação funcional. É esperado que os pseudogenes mais recentes apresentem maiores similaridades com as cópias funcionais do que os pseudogenes mais antigos (Echols et al., 2002). Assim, H7 e H12 devem representar os pseudogenes de 5.8S mais antigos entre todos os amostrados, enquanto H11 e H21, por exemplo, devem ter origem bem recente e, portanto ainda são mais semelhantes a H1. Cabe ressaltar que os pseudogenes de origem mais recente, na maior parte dos casos, são justamente aqueles identificados pela comparação com os motivos conservados, indicando a tendência destes haplótipos, em um futuro próximo, acumular mutações adicionais e perder a formação adequada da estrutura secundária.

Alguns dos pseudogenes estão intimamente relacionados na rede de haplótipos para 5.8S, mesmo pertencendo a indivíduos de populações diferentes

(figuras 7 e 8), como é o caso dos haplótipos H6, H16 e H17 (populações TLM e FRL), os haplótipos H3, H12, H14, H23 e H24 (populações DOG, RAP e VRB) e o haplótipo H7 (compartilhado pelas populações GUA e RAP). Dentro destes agrupamentos, estão ainda mais próximos os pseudogenes de um mesmo indivíduo, o que ocorre com os pares H12 e H14 (população RAP), H16 e H17 (população TLM) e H23 e H24 (população VRB). Provavelmente, estes pseudogenes iniciaram sua trajetória antes da subdivisão das populações e, depois da separação destas, ainda foram acrescentadas outras modificações. Esta hipótese também suporta a pressuposição de que estes pseudogenes são os mais antigos. Adicionalmente, os pseudogenes exclusivos de um único indivíduo são justamente os que têm menor número de modificações e foram considerados de origem mais recente.

Outro aspecto importante é que todos os indivíduos contêm a cópia funcional de 5.8S H1 (figura 8), que também é o centro da rede de haplótipos (figura 7), sugerindo que este é o padrão funcional típico de 5.8S na espécie e supostamente para onde deve estar direcionada a homogeneização. Todos os outros haplótipos diagnosticados como funcionais devem, então, caminhar para a igualdade com H1 através da evolução em concerto ou se tornarem pseudogenes. O haplótipo H20, que suscita dúvidas quanto a sua funcionalidade, deve estar num caminho intermediário de pseudogenização a partir de H1.

Varição intragenômica e evolução em concerto:

Neste trabalho foi verificada a existência de variação intragenômica do rDNA ITS em mais da metade dos indivíduos estudados de *C. ipecacuanha*, o que torna evidente a atuação incompleta da evolução em concerto nesta espécie.

Os processos de homogeneização das diferentes cópias dentro do genoma (recombinação e conversão gênica) são mais freqüentes dentro de cromossomos do que entre cromossomos (Schlötterer e Tautz, 1994; Parkin e Butlin, 2004). Uma vez que *C. ipecacuanha* é tetraplóide (Rossi et al., 2008), é plausível supor que a variação intra-individual seja decorrente da localização de conjuntos destes rDNA em cromossomos diferentes, procedentes da hibridização ocorrida na formação da espécie, como é suposto para *Buxus balearica* (Rosseló et al., 2005) e *Leucaena leucocephala* (Hughes et al., 2002).

A reprodução assexuada em *C. ipecacuanha* pode ser considerada outro fator contribuinte para a persistência da variação intragenômica. A propagação vegetativa prolonga o tempo de geração e, logo, retarda a homogeneização promovida pela evolução em concerto (Márquez et al., 2003; Lorenz-Lemke et al. 2005). A agamospermia, relatada na espécie por Rossi et al. (2005), também proporciona este efeito, pois os eventos de recombinação e conversão gênica da evolução em concerto são mais freqüentes durante a meiose, etapa suprimida na reprodução assexuada.

CONCLUSÕES

A estrutura secundária da região 5.8S e da região ITS2 apresentam a conformação típica presente nas demais angiospermas para a maioria das cópias encontradas (cópias funcionais). A comparação com os motivos conservados em angiospermas e a estrutura secundária para a região 5.8S foram os marcadores mais confiáveis de funcionalidade das seqüências em *C. ipecacuanha*, identificando um total de 13 pseudogenes em meio a 26 haplótipos. Já o haplótipo H20 foi considerado um pseudogene adicional em razão das modificações vistas em ITS2 e o padrão de cópia não funcional exibido para o conteúdo CG e a energia livre de 5.8S. Desta forma, 14 pseudogenes foram apontados entre as diferentes cópias de ITS. Devido ao acúmulo de um maior número de modificações e à formação de grupamentos distintos na rede, provavelmente a perda de funcionalidade dos haplótipos H3, H6, H7, H12, H14, H16, H17, H23 e H24 é mais antiga do que nos demais pseudogenes. Ainda, este estudo levanta a hipótese de correlação entre a existência de variação intragenômica em *C. ipecacuanha* e o evento de poliploidização ocorrido durante a especiação, efeito que pode ter sido potencializado em razão da propagação vegetativa e agamospermia.

REFERÊNCIAS BIBLIOGRÁFICAS

- Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417-434.
- Bailey, C.D., Carr, T.G., Harris, S.A., Hughes, C.E., 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Mol. Phylogenet. Evol.* 29, 435-455.
- Bandelt, H-J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37-48.
- Boon-Unge, K., Yu, Q., Zou, T., Zhou, A., Govitrapong, P., Zhou, J., 2007. Emetine regulates the alternative splicing of *Bcl-x* through a protein phosphatase 1-dependent mechanism. *Chem. Biol.* 14, 1386-1392.
- Buckler, E.S.I., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145, 821–832.
- Coleman, A.W., 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res.* 35 (10), 3322-3329.
- Echols, N., Harrison, P., Balasubramanian, S., Luscombe, N.M., Bertone, P., Zhang, Z., Gerstein, M., 2002. Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Res.* 30 (11), 2515-2523.
- Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., Eddy, S.R., Bateman, A., 2008. Rfam: updates to the RNA families database. *Nucleic Acids Res.* 37, D136-D140.
- Grimm, G.W., Denk, T., 2007. ITS evolution in *Platanus* (Platanaceae): Homeologues, Pseudogenes and Ancient Hybridization. *Ann. Bot.* 101 (3), 403-419.

- Harpke, D., Peterson, A., 2006. Non-concerted ITS evolution in *Mammillaria* (Cactaceae). *Mol. Phylogenet. Evol.* 41, 579-593.
- Harpke, D., Peterson, A., 2008a. 5.8S motifs for the identification of pseudogenetic ITS regions. *Botany* 86, 300-305.
- Harpke, D., Peterson, A., 2008b. Extensive 5.8S nrDNA polymorphism in *Mammillaria* with special reference to the identification of pseudogenetic internal transcribed spacer regions. *J. Plant Res.* 121, 261-270.
- Hughes, C.E., Bailey, C.D., Harris, S.A., 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploidy origins and nrDNA polymorphism. *Am. J. Bot.* 89(7), 1057-1073.
- Jobes, D.V., Thien, L.B., 1997. A Conserved Motif in the 5.8S Ribosomal RNA (rRNA) Gene is a Useful Diagnostic Marker for Plant Internal Transcribed Spacer (ITS) Sequences. *Plant Mol. Biol. Rep.* 15, 326-334.
- Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451-1452.
- Long, E.O., Dawid, I.B., 1980. Repeated genes in eukaryotes. *Ann. Rev. Biochem.* 49, 727-764.
- Lorenz-Lemke, A.P., Muschner, V.C., Bonatto, S.L., Cervi, A.C., Salzano, F.M., Freitas, L.B., 2005. Phylogeographic inferences concerning evolution of Brazilian *Passiflora actinia* and *P. elegans* (Passifloraceae) based on ITS (nrDNA) Variation. *Ann. Bot.* 95, 799-806.
- Márquez, L.M., Miller, D.J., MacKenzie, J.B., Oppen, M.J.H., 2003. Pseudogenes contribute to the extreme diversity of nuclear ribosomal DNA in the hard coral *Acropora*. *Mol. Biol. Evol.* 20 (7), 1077-1086.

- Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structures. *J. Mol. Evol.* 288, 911-940.
- Mai, J.C., Coleman, A.W., 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J. Mol. Evol.* 44, 258-271.
- Möller, M., Herzer, K., Wenger, T., Herr, I., Wink, M., 2007. The alkaloid emetine as promising agent for the induction and enhancement of drug-induced apoptosis in leukemia cells. *Oncol. Rep.* 18, 734-744.
- Ochieng, J.W., Henry, R.J., Baverstock, P.R., Steane, D.A., Sheperd, M., 2007. Nuclear ribosomal pseudogenes resolve a corroborated monophyly of the eucalypt genus *Corymbia* despite misleading hypotheses at functional ITS paralogs. *Mol. Phylogenet. Evol.* 44, 752-764.
- Oliveira, L.O., Rossi, A.A.B, Martins, E.R., Batista, F.R.C., Silva, R.S., 2010. Molecular phylogeography of *Carapichea ipecacuanha*, an amphotropical shrub that occurs in the understory of both semideciduous and evergreen forests. *Mol. Ecol.* 19 (7), 1410-1422.
- Parkin, E.J., Butlin, R.K., 2004. Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Mol. Biol. Evol.* 21, 1595–1601.
- Rosseló, J.A., Lázaro, A., Cosín, R., Molins, A., 2005. A phylogeographic split in *Buxus balearica* (Buxaceae) as evidenced by nuclear ribosomal markers: when ITS paralogues are welcome. *J. Mol. Evol.* 64, 143-157.
- Rossi, A.A.B., Oliveira, L.O., Vieira, M.F., 2005. Distyly and variation in floral traits in natural populations of *Psychotria ipecacuanha* (Brot.) Stokes (Rubiaceae). *Rev. Bras. Bot.* 28 (2), 285-294.

- Rossi, A.A.B., Clarindo, W.R., Carvalho, C.R., Oliveira, L.O., 2008. Karyotype and nuclear DNA content of *Psychotria ipecacuanha*: a medicinal species. *Cytologia* (73) 1, 53-60.
- Rossi, A.A.B., Oliveira, L.O., Venturini, B.A., Silva, R.S., 2009. Genetic diversity and geographic differentiation of disjunct Atlantic and Amazonian population of *Carapichea ipecacuanha* (Rubiaceae). *Genetica* 136 (1), 57-67.
- Schlötterer, C., Tautz, D., 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* 4, 777-783.
- Seibel, P.N., Müller, T., Dandekar, T., Wolf, M., 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Res. Notes* 1, 91-98.
- Selig, C., Wolf, M., Müller, T., Dandekar, T. and Schultz, J., 2008. The ITS2 Database II: homology modeling RNA structure for molecular systematic *Nucleic Acids Res.* 36, D377-D380.
- Suh, Y., Thien, L.B., Zimmer, E.A., 1992. Nucleotide sequences of the internal transcribed spacers and 5.8S rRNA gene in *Canella winterana* (Magnoliales; Canellaceae). *Nucleic Acids Res.* 20 (22), 6101-6102.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596-1599.
- Veloso, H.P., 1947. As condições ecológicas da *Cephaelis ipecacuanha* Rich. *Mem. Inst. Oswaldo Cruz* 45, 361-372.
- Xu, J., Zhang, Q., Xu, X., Wang, Z., Qi, J., 2009. Intragenomic variability and pseudogenes of ribosomal DNA in Stone flounder *Kareius bicoloratus*. *Mol. Phylogenet. Evol.* 52, 157-166.

- Watanabe, N., Iwamoto, T., Dickinson, D.A., Illes, K.E., Forman, H.J., 2002. Activation of the mitochondrial caspase cascade in the absence of protein synthesis does not require c-Jun N-terminal kinase. *Arch. Biochem. Biophys.* 405, 231-240
- White, T.J., Bruns, T.D., 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*. Academic Press, California, pp. 315-322.
- Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T., De Wachter, R., 2001. The European Large Subunit Ribosomal RNA database. *Nucleic Acids Res.* 29 (1), 175-177.
- Wuyts, J., Perriere, G., Van de Peer, Y., 2004. The European ribosomal RNA database. *Nucleic Acids Res.* 32, D101-D103.
- Yang, Y.W., Lai, K.N., Tai, P.Y., Ma, D.P., Li, W.H., 1999. Molecular phylogenetics studies of *Brassica*, *Rorippa*, *Arabidopsis* and allied genera based on the internal transcribed spacer region of 18S-25S rDNA. *Mol. Phylogenet. Evol.* 13 (3), 455-452.
- Zheng, X., Cai, D., Yao, L., Teng, Y., 2008. Non-concerted ITS evolution, early origin and phylogenetic utility of ITS pseudogenes in *Pyrus*. *Mol. Phylogenet. Evol.* 48, 892-803.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406-3415.