



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR**  
**CURSO DE PÓS-GRADUAÇÃO EM BIOQUÍMICA**

**CLONAGEM, SEQUENCIAMENTO E CARACTERIZAÇÃO PARCIAL  
DOS GENES RELACIONADOS À LECTINA DE *Vatairea macrocarpa***

**JOÃO GARCIA ALVES FILHO**

**FORTALEZA-CE**

**2008**

**JOÃO GARCIA ALVES FILHO**

**CLONAGEM, SEQUENCIAMENTO E CARACTERIZAÇÃO PARCIAL  
DOS GENES RELACIONADOS À LECTINA DE *Vatairea macrocarpa***

Dissertação apresentada como parte  
dos requisitos para a obtenção do grau  
de Mestre em Bioquímica pela  
Universidade Federal do Ceará.

**FORTALEZA-CE**

**2008**

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

A479c Alves Filho, João Garcia.

Clonagem, sequenciamento e caracterização parcial dos genes relacionados à lectina de *Vatairea macrocarpa* / João Garcia Alves Filho. – 2010.

98 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Bioquímica, Fortaleza, 2010.

Orientação: Prof. Dr. Benildo Sousa Cavada.

1. Bioquímica. 2. Lectina de planta. 3. Sementes. I. Título.

CDD 572

---

Esta dissertação foi apresentada como parte dos requisitos necessários para a obtenção do grau de Mestre em Bioquímica, outorgado pela Universidade Federal do Ceará.

---

João Garcia Alves Filho

Dissertação aprovada em:

---

Benildo Sousa Cavada, Dr.

Orientador

Departamento de Bioquímica e Biologia Molecular-UFC

---

João Paulo Matos, Dr.

Universidade Estadual Vale do Acaraú-UVA, CNPq

---

Cristiane Cunha Frota, Dr.

Departamento de Microbiologia-UFC

*“Existem muitas hipóteses em ciência que estão erradas. Isso é perfeitamente aceitável, eles são a abertura para achar as que estão certas”*

Carl Sagan



## AGRADECIMENTOS

À Deus a quem devo minha gratidão pelo simples fato de minha existência e poder contemplar a grandiosidade de Suas obras.

Ao Prof. Dr. Benildo Sousa Cavada, a quem tenho grande admiração, fico honrado em ser seu orientando uma vez que ele foi responsável pela orientação de vários brilhantes pesquisadores.

Ao Prof. Rodrigo Maranguape Silva da Cunha pelo enorme incentivo, amizade e exemplo de profissional dedicado e humano acima de tudo.

Ao Dr. Celso Shiniti Nagano, pela brilhante forma com que solucionou várias dificuldades encontradas na execução deste trabalho.

Ao Dr. João Paulo Matos, pela disponibilidade em compor a banca e grande contribuição dada à interpretação dos dados de bioinformática além de muitas outras dúvidas tiradas.

A Dra. Cristiane Frota pela disponibilidade em compor a banca.

Ao Dr. Luciano da Silva Pinto, pela grande contribuição dada aos experimentos com clonagem e sequenciamento de DNA.

A todos os amigos do Laboratório de Moléculas Biologicamente ativas, em especial à Kyria Santiago, Raquel Benevides, Ramon, Sâmia, Felipe, Daniela, Cícero, Tales.

A todos os amigos do Núcleo de Biotecnologia de Sobral, em especial à Auxiliadora Oliveira e aos amigos Ricardo, Daniel, Monalysa, Jackson, Mayron, Luiz, Tatiane, Antonio, Fernanda, Gleiciane, Anderson, Viviane, Isana, Ivna, Cíntia, Gisvani, Larisse.

Á toda minha família, em especial meu pai João Garcia, minha mãe Rosália, minha filha Ana Eduarda, ao Sávio, Rosana, Luciana e aos meus padrinhos.

**Este trabalho foi realizado graças ao apoio das seguintes instituições:**

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), através da Bolsa de Mestrado concedida, nos anos de 2006 e 2007.

Universidade Federal do Ceará (UFC).

Departamento de Bioquímica e Biologia Molecular pertencente à UFC.

Laboratório de Moléculas Biologicamente Ativas (Biomol Lab) do Departamento de Bioquímica e Biologia Molecular/UFC.

Faculdade de Medicina de Sobral (FAMED) pertencente à UFC.

Núcleo de Biotecnologia de Sobral (NUBIS) cuja sede fica localizada na Faculdade de Medicina de Sobral.

## SUMÁRIO

LISTA DE FIGURAS .....	x
LISTA DE TABELAS .....	xii
ABREVIATURAS E DEFINIÇÕES .....	xiii
RESUMO.....	xvii
ABSTRACT.....	xviii
<b>1. INTRODUÇÃO.....</b>	<b>1</b>
<b>2. REVISÃO DE LITERATURA .....</b>	<b>2</b>
2.1. Lectinas vegetais .....	2
2.1.1. Breve histórico.....	2
2.1.2. Definição e Classificação .....	3
2.1.3. Função Biológica .....	7
2.1.4. Aplicações das lectinas .....	10
2.2. Lectinas de Leguminosas .....	11
2.2.1. Estrutura e propriedades químicas .....	11
2.2.2. Biossíntese e processamento.....	12
2.2.3. Genes de lectinas de leguminosas.....	18
2.2.4. Evolução molecular de lectinas de leguminosas.....	21
2.2.5. A lectina de <i>Vatairea macrocarpa</i> .....	23
<b>3. OBJETIVOS.....</b>	<b>28</b>
<b>Geral.....</b>	<b>28</b>
<b>Específicos.....</b>	<b>28</b>
<b>4. MATERIAIS E MÉTODOS.....</b>	<b>29</b>
<b>4.1. Isolamento de DNA genômico.....</b>	<b>29</b>
4.2. Eletroforese em gel de agarose 0,8% (m/v), pH 8,0 .....	30
<b>4.3. Amplificação da região central do gene da lectina de <i>V. macrocarpa</i> (VML) por PCR.....</b>	<b>30</b>
<b>4.4. Clonagem do fragmento gênico de <i>Vml</i> amplificado por PCR.....</b>	<b>33</b>
4.4.1. Ligação do produto de PCR ao vetor .....	33
4.4.2. Preparo de células eletrocompetentes.....	33
4.4.3. Transformação .....	35
<b>4.5. Seleção de clones recombinantes para sequenciamento.....</b>	<b>35</b>
4.5.1. PCR de colônia .....	35
4.5.2. Extração de DNA plasmidial ( <i>miniprep</i> ) .....	35
<b>4.6. Isolamento de RNA total .....</b>	<b>37</b>
<b>4.7. Síntese da primeira fita de cDNA.....</b>	<b>37</b>

<b>4.8. Síntese da segunda fita de cDNA por amplificação das extremidades 5' e 3'</b> .....	38
<b>4.9. Sequenciamento de DNA</b> .....	41
4.9.1. Reação de sequenciamento .....	41
4.9.2. Precipitação por etanol .....	41
4.9.3. Análise das sequências no MegaBace 750 .....	42
4.10. Montagem das Sequências Obtidas.....	43
<b>5. RESULTADOS</b> .....	44
5.1. Isolamento de DNA genômico .....	44
5.2. Amplificação da região interna de <i>Vml</i> por PCR.....	44
5.3. Clonagem do fragmento gênico de <i>Vml</i> amplificado por PCR .....	44
5.4. Extração de DNA de plasmídios recombinantes por lise alcalina.....	47
5.5. Sequenciamento da região central do gene <i>Vml</i> .....	47
5.6. Alinhamento das sequências .....	47
5.7. Amplificação Rápida das Extremidades de cDNA (RACE).....	55
5.8. Alinhamento da região 3' de <i>Vml</i> .....	59
<b>6. DISCUSSÃO</b> .....	62
<b>7. SÍNTESE DE RESULTADOS E CONCLUSÕES</b> .....	66
<b>BIBLIOGRAFIA</b> .....	67
<b>ANEXOS</b> .....	76

## LISTA DE FIGURAS

- Figura 1 - Classificação das lectinas de leguminosas..... 6
- Figura 2 – Representação esquemática dos eventos pós-traducionais que ocorrem durante a biossíntese da VML e ConA .....16
- Figura 3 – Árvore filogenética das sequências que codificam lectinas de leguminosas e proteínas relacionadas.....22
- Figura 4 – Características morfológicas de *Vatairea macrocarpa* (Benth.) Ducke.....25
- Figura 5 – Estratégia para a construção de *primers* semi-degenerados para o gene *Vml* de *Vatairea macrocarpa*.....32
- Figura 6 – Estratégia de clonagem dos fragmentos de DNA correspondente à região central do gene *Vml* obtidos por PCR.....34
- Figura 7 – Estratégia utilizada para a amplificação rápida das extremidades de cDNA (RACE) do gene da lectina de *V. macrocarpa*.....39
- Figura 8 – Estratégia utilizada para a amplificação da extremidade 5' do gene que codifica a lectina VML.....39
- Figura 9 - Eletroforese em gel de agarose a 0,8% de DNA genômico de *Vatairea macrocarpa*.....45
- Figura 10 – Eletroforese em gel de agarose a 1% corada com brometo de etídio do gene da lectina da *V. macrocarpa* amplificado por “PCR touchdown” com temperatura de anelamento de 50 a 45°C.....45
- Figura 11 – Eletroforese em gel de agarose 1% da reação de PCR das colônias contendo a região central do gene *Vml*.....46

- Figura 12 – Eletroforese em gel de agarose a 1% corada com brometo de etídio do plasmídio recombinante de *V. macrocarpa* extraído de *E. coli* por lise alcalina. Linhas de 1 a 4: Vml1, Vml2, Vml4, Vml6, respectivamente. Linhas 5 e 6: controle positivo.....50
- Figura 13 – Alinhamento das sequências de DNA obtidas pelo sequenciamento da região central de *Vml* (*contigs* 1, 2 e 3).....51
- Figura 14 – Alinhamento da tradução dos três *contigs* da região central de *Vml* com lectinas de leguminosas depositadas no GenBank. ....53
- Figura 15 – Dendograma obtido pelo método de distância do vizinho mais próximo (Neighbor-Joining), utilizando a distância-p como modelo de substituição a partir das sequências centrais traduzidas das três sequências consenso *contigs* e outras lectinas de leguminosas.....54
- Figura 16 – Gel de agarose a 0,8% corado com brometo de etídio do resultado de extração de RNA total de *V. macrocarpa* pelo kit Plant RNA Reagent.....56
- Figura 17 – Eletroforese em gel de agarose a 1% mostrando a 3'RACE da *Vml* utilizando os *primers* desenhado a partir da região central já sequenciada com os *primers* para o adaptador.....57
- Figura 18 – Eletroforese em gel de agarose a 1% mostrando a amplificação da extremidade 5' da *Vml* utilizando os *primers* desenhado a partir do peptídeo sinal com os *primers* desenhados a partir da região interna já sequenciada e para o *primer* do adaptador da cauda poliA.....58
- Figura 19 – Alinhamento do *ContigA* obtida pela técnica 3'RACE a partir de cDNA de folhas (até o primeiro códon de parada) e sua complementaridade com o *contig3* previamente identificado a partir de DNA genômico.....60
- Figura 20 – Região C-terminal da sequência obtida pela união do *contig3* com o *contigA* até o primeiro códon de parada alinhada com sequências de precursores de outras lectinas de leguminosas.....61

## LISTA DE TABELAS

Tabela 1 – <i>Primers</i> utilizados na amplificação da região central do gene <i>Vml</i> .....	32
Tabela 2 – Condições térmicas da amplificação da região central da <i>Vml</i> por PCR <i>touchdown</i> .....	32
Tabela 3 – <i>Primers</i> utilizados na amplificação da extremidade 5' e 3' do gene <i>Vml</i> ..	40
Tabela 4 – Condições térmicas da amplificação das extremidades 5' e 3' do gene <i>Vml</i> .....	40
Tabela 5 – Quantificação das amostras de DNA plasmidial recombinante de <i>V. macrocarpa</i> (pGemT-Vml) extraído de <i>E. coli</i> por lise alcalina.....	50
Tabela 6 – Heterogeneidade nos resíduos de aminoácidos encontrados na lectina de <i>V. macrocarpa</i> depositada e os resíduos encontrados nas mesmas posições das sequências internas obtidas por clonagem e sequenciamento da região central a partir de DNA genômico. A numeração foi baseada na sequência depositada VML. Het. (Heterogeneidade encontrada nos resíduos da VML), Rp ( <i>Robinia pseudoacacia</i> ), Sj ( <i>Sophora japonica</i> ), Ev ( <i>Erythrina variegata</i> ) .....	52
Tabela 7– Resíduos de aminoácidos importantes na manutenção da estrutura tridimensional e ligação a carboidrato na lectina de <i>V. macrocarpa</i> e outras lectinas relacionadas. A numeração foi baseada na sequência depositada VML. Rp ( <i>Robinia pseudoacacia</i> ), Sj ( <i>Sophora japonica</i> ), Ev ( <i>Erythrina variegata</i> ), M (Sítio de ligação a metais), C (Sítio de ligação a carboidratos), Glc (local onde contem N-glicano) .....	52
Tabela 8 – Quantificação das amostras de RNA total extraídas pelo Plant RNA Reagent. A concentração das amostras varia de 1512 ng/μl a 2540 ng/μl, a absorvância $A_{260}$ varia entre 0,219 a 0,545 e a razão entre $A_{260}/A_{280}$ varia entre 1,707 a 2,025.....	56

## ABREVIATURAS E DEFINIÇÕES

### Resíduos de aminoácidos

Uma letra	Três letras	Aminoácido	Uma letra	Três letras	Aminoácido
A	Ala	Alanina	N	Asn	Asparagina
C	Cys	Cisteína	P	Pro	Prolina
D	Asp	Aspartato	Q	Gln	Glutamina
E	Glu	Glutamato	R	Arg	Arginina
F	Phe	Fenilalanina	S	Ser	Serina
G	Gly	Glicina	T	Thr	Treonina
H	His	Histidina	V	Val	Valina
I	Ile	Isoleucina	W	Trp	Triptofano
K	Lys	Lisina	X	Xxx	Desconhecido
L	Leu	Leucina	Y	Tyr	Tirosina
M	Met	Metionina			

### Bases nitrogenadas/degenerações

Código	Base/degeneração	Código	Base/degeneração
A	Adenina	M	A ou C
B	G, T ou C	N	A, T, G ou C
C	Citosina	R	A ou G
D	G, A ou T	S	G ou C
G	Guanina	T	Timina
H	C, A ou T	V	G, A ou C
I	Inosina	W	A ou T
K	G ou T	Y	C ou T

Lista de lectinas e proteínas relacionadas às lectinas

---

aA1	Inibidor de alfa amilase de <i>Phaseolus vulgaris</i>
Arcelina-1	Arcelina de <i>Phaseolus vulgaris</i>
Blec 1	Lectina putativa de <i>Pisum sativum</i>
BMA	Aglutinina de <i>Browningia milbraedii</i>
BPA	Aglutinina de <i>Bauhinia purpúrea</i>
CLA	Aglutinina de <i>Cladrastis lutea</i>
CLLRP	Proteína relacionada a lectina de <i>Cladrastis lutea</i>
ConA	Lectina de semente de <i>Canavalia ensiformis</i>
CS II	Lectina de <i>Cytisus scoparius</i>
CSA-I	Aglutinina de <i>Cytisus sessilifolius</i>
CSA-II	Aglutinina de <i>Cytisus sessilifolius</i>
DBL	Lectina de semente de <i>Dolichos biflorus</i>
DB58	Lectina de tecido vegetativo de <i>Dolichos biflorus</i>
DGL	Lectina de <i>Dioclea grandiflora</i>
DLA	Aglutinina de <i>Lablab purpureus</i>
EcorL	Lectina de <i>Erythrina corallodendron</i>
GS-II	Lectina de <i>Griffonia simplicifolia</i>
LAA	Aglutinina de <i>Laburnum alpinum</i>
LBL	Lectina de <i>Phaseolus lunatus</i>
LCA	Aglutinina de <i>Lens culinaris</i>
LOL I	Lectina de <i>Lathyrus ochrus</i>
LTA	Aglutinina de <i>Lotus tetragonolobus</i>
MAHb	Hemaglutinina de casca de <i>Maackia amurensis</i>
MAHs	Hemaglutinina de semente de <i>Maackia amurensis</i>
MALb	Leucoaglutinina de casca de <i>Maackia amurensis</i>
MALs	Leucoaglutinina de semente de <i>Maackia amurensis</i>
mtrL 1	Lectina putativa de semente de <i>Medicago truncatula</i>
NGL	Lectina de <i>Arachis hypogaea</i>
NML	Lectina de <i>Arachis hypogaea</i>
OVL	Lectina de <i>Onobrychis viciifolia</i>
PHA-E	Hemaglutinina de <i>Phaseolus vulgaris</i>
PHA-L	Leucoaglutinina de <i>Phaseolus vulgaris</i>

---

---

PNA	Aglutinina de <i>Arachis hypogaea</i>
PSA	Aglutinina de <i>Pisum sativum</i>
RPbAI-a	Aglutinina I de casca de <i>Robinia pseudoacacia</i> (31,5 KDa)
RPbAI-b	Aglutinina I de casca de <i>Robinia pseudoacacia</i> (29 KDa)
RPbA II	Aglutinina II de casca de <i>Robinia pseudoacacia</i>
SBA	Aglutinina de <i>Glycine max</i>
SJAbg	Aglutinina de casca de <i>Sophora japonica</i> (Gal ligante)
SJAbml	Aglutinina I de casca de <i>Sophora japonica</i> (Man/glc ligante)
SJAbmlI	Aglutinina II de casca de <i>Sophora japonica</i> (Man/glc ligante)
SJAsg	Aglutinina de semente de <i>Sophora japonica</i> (Gal ligante)
UEA-I	Aglutinina 1 de <i>Ulex europaeus</i>
UEA-II	Aglutinina 2 de <i>Ulex europaeus</i>
VFA	Aglutinina de <i>Vicia faba</i>
VML	Lectina de <i>Vatairea macrocarpa</i>
WBA I	Aglutinina de <i>Psophocarpus tetragonolobus</i>

---

Outras abreviações:

A	Adenilato
A260	Absorbância a 260 nm
A280	Absorbância a 280 nm
C	Citidilato
C3	Carbono 3
C4	Carbono 4
cDNA	DNA complementar
ConA	Concanavalina A
CTAB	Brometo de cetil trimetil amônio
Da	Dálon
ddNTP	Didesoxi-nucleotídeo-trifosfatado
dNTP	Desoxi-nucleotídeo-trifosfatado
EDTA	Ácido etileno diamino tetra-acético
G	Guanilato
Gal	Galactose
GalNac	N-acetil galactosamina

IPTG	Isopropil- $\beta$ -D-tiogalactopiranosídeo
KDa	Quilodáton
KOAc	Acetato de potássio
KV	Quilovolt
LB	Luria-Bertani
Mg <sup>2+</sup>	Íon magnésio
Mn <sup>2+</sup>	Íon manganês
OD	Densidade óptica
pb	Pares de base
RACE	Amplificação rápida das extremidades de DNA
rcf	Força centrífuga relativa
rDNA	Gene que codifica o RNA ribossômico
RER	Retículo endoplasmático rugoso
RIP	Proteína inativadora de ribossomo
rpm	Rotações por minuto
rRNA	RNA ribossômico
T	Timidilato
Taq	<i>Thermus aquaticus</i>
TBE	Tampão Tris-borato-EDTA
TE	Tampão Tris-EDTA
Tm	Temperatura de melting
Tris	Tris-hidroxiaminometano
tRNA	RNA transportador
X-Gal	5-bromo-4-cloro-3-indoil- $\beta$ -D-galactopiranosídeo

# CLONAGEM, SEQUENCIAMENTO E CARACTERIZAÇÃO PARCIAL DOS GENES RELACIONADOS À LECTINA DE *Vatairea macrocarpa*

**Autor:** João Garcia Alves Filho

**Orientador:** Benildo Sousa Cavada

## RESUMO

No presente trabalho é feita a descrição de três genes distintos que codificam lectinas ou proteínas relacionadas à lectina de *Vatairea macrocarpa*. As sequências foram obtidas pela amplificação de DNA genômico e cDNA de folhas utilizando *primers* semi-degenerados construídos a partir da informação da sequência de aminoácidos da lectina VML depositada no GenBank. O resultado do sequenciamento revelou a presença de três *contigs*. O *contig1* corresponde à lectina VML desde que se assuma que a lectina depositada VML contenha heterogeneidades ou ambiguidades decorrentes na degradação de Edman. A tradução dos *contigs* 2 e 3 mostram identidade de sequência de 77% quando comparadas com VML. As sequências, apesar de apresentar regiões conservadas, mostram diferenças de aminoácidos nos sítios de N-glicosilação, sítios de ligação a carboidrato e metais além da presença de resíduos de cisteína sugerindo que tais proteínas podem ter outras atividades biológicas. A análise da sequência obtida pelo 3' RACE se mostrou complementar ao *contig3*. Sendo assim, a sequência híbrida *contig3/contigA* possui 2 resíduos de cisteína além de revelar diferenças de aminoácidos na região C-terminal quando alinhada com outras lectinas de leguminosas. Análises filogenéticas revelaram que os *contigs* observados formam um grupo monofilético e tem alta similaridade com as lectinas de *Sohora japonica* e *Robinia pseudoacacia*, além da proteína relacionada à lectina de *Cladrastis lutea*.

**Palavras-chave:** Lectina, *Vatairea macrocarpa*, famílias multigênicas, filogenia

# CLONING, SEQUENCING AND PARTIAL CHARACTERIZATION OF *Vatairea macrocarpa* LECTIN RELATED GENES

**Author:** João Garcia Alves Filho

**Adviser:** Benildo Sousa Cavada

## ABSTRACT

In this paper is made a description of three distinct genes that encode *Vatairea macrocarpa* lectin and related proteins. The sequences were obtained by amplification of genomic DNA and cDNA of leaves using semi-degenerate primers constructed from the information of the amino acid sequence of VML lectin deposited in GenBank. The result of sequencing reveals the presence of three different genes, called contig 1, 2 and 3. The VML lectin corresponds to contig1 long as one assumes that the lectin contains heterogeneities deposited VML or ambiguities arising in the Edman degradation. The translation of contigs 2 and 3 show sequence identity of 77% compared to VML. Sequences, despite having conserved regions show differences in amino acid N-glycosylation sites, carbohydrate binding sites and metals and the presence of cysteine residues suggests that these proteins may have other biological activities. The analysis of the sequence obtained by 3'RACE proved complementary to contig3. Thus, the sequence contig3/contigA hybrid has two cysteine residues in addition to revealing differences in amino acid C-terminal region when aligned with other legume lectins. Phylogenetic analysis revealed that the observed contigs form a monophyletic group and has high similarity with lectins from *Robinia pseudoacacia* *Sohora japonica* and, in addition to the lectin-related protein *Cladrastis lutea*.

**Keywords:** Lectin, *Vatairea macrocarpa*, multigenic families, phylogeny

## INTRODUÇÃO

As lectinas são proteínas ou glicoproteínas, que apresentam pelo menos um domínio capaz de se ligar de forma reversível a mono ou oligossacarídeos sem que haja modificação enzimática deles (GOLDSTEIN, HUGHES, *et al.*, 1980); (KOCOUREK e HOREJSI, 1981). Esta classe de proteínas é encontrada em todos os seres vivos, incluindo vírus, mas são mais abundantes e melhor estudadas em leguminosas que contem uma família de lectinas homólogas que ocorrem exclusivamente nesta família.

A lectina de *Vatairea macrocarpa* (VML) é uma proteína ligante de galactose com uma série de efeitos biológicos documentados, como o efeito pró-inflamatório (ALENCAR, ASSREUY, *et al.*, 2004), efeitos nos sítios de ligação a carboidrato no sistema renal (MARTINS, MONTEIRO, *et al.*, 2005) e capacidade de se ligarem ao antígeno *Tn* (DAM, GERKEN, *et al.*, 2007) muito comum em células de câncer maligno de tecido glandular (AVICHEZER e ARNON, 1996).

A clonagem gênica tem levado a descoberta de novas lectinas além de outras proteínas que compartilham similaridade de sequência com lectinas particulares, mas que são desprovidas de atividade de ligação a carboidrato (SHARON e LIS, 2003).

A investigação acerca do gene que codifica a lectina de *Vatairea macrocarpa* VML pode contribuir para o melhor entendimento dos mecanismos pós-traducionais que ocorrem nessa lectina além de fornecer a informação necessária para sua produção em sistemas heterólogos.

## 2. REVISÃO DE LITERATURA

### 2.1. Lectinas vegetais

#### 2.1.1. Breve histórico

O primeiro relato sobre lectinas surgiu na observação da aglutinação de sangue de pombo pelo veneno de cascavel em 1860 por WEIR MITCHELL (KILPATRICK e GREEN, 1992). Em 1888, Hermann Stillmark demonstrou que o extrato de sementes de *Ricinus communis* era capaz de promover a aglutinação de eritrócitos de animais. Ele denominou tal “fator hemaglutinante” de ricina (SHARON e LIS, 1990; VAN DAMME, PEUMANS, *et al.*, 1998). Atualmente, é conhecido que a “ricina” descrita por Stillmark era na verdade uma mistura complexa de moléculas da verdadeira ricina (uma proteína inativadora de ribossomos do tipo 2) e de aglutininas não tóxicas (as primeiras lectinas vegetais encontradas). No entanto, o seu trabalho foi um marco na biologia e na bioquímica, pois foi o primeiro a relacionar a atividade tóxica da ricina a fatores tóxicos de origem protéica e a primeira proteína vegetal a qual se pôde atribuir uma atividade biológica bem definida (VAN DAMME, PEUMANS, *et al.*, 1998).

Após a descoberta do fator hemaglutinante de mamona, outras substâncias tóxicas foram identificadas nas sementes de inúmeras plantas. Em 1898, Elfstrand introduziu pela primeira vez o termo *hemaglutinina* como um nome comum para todas as proteínas vegetais que causam o agrupamento de hemácias (VAN DAMME, PEUMANS, *et al.*, 1998). O público geral passou a conhecer a ricina em 1978 com o assassinato estratégico do jornalista Georgi Markov com uma arma chamada “guarda-chuva” contendo ricina como projétil. Além disso, os Estados Unidos tentaram usar a ricina na construção de armas de destruição em massa durante a Primeira e Segunda Guerra Mundial (SHARON e LIS, 2004).

A história das lectinas de leguminosas começa em 1890 com o isolamento de uma suposta lectina tóxica chamada “robin” das cascas de *Robinia pseudoacacia* (VAN DAMME, PEUMANS, *et al.*, 1998). Em 1907, o médico e biólogo Karl Landsteiner (conhecido pela criação do sistema ABO e descobridor do fator Rh) e

Raubitschek reportaram pela primeira vez a presença de lectinas não tóxicas nos legumes *Phaseolus vulgaris*, *Pisum sativum*, *Lens culinaris* e *Vicia sativa*. Muitos outros trabalhos posteriores relataram a ocorrência de outras hemaglutininas de plantas não tóxicas, tornando evidente a ampla distribuição dessas proteínas no Reino Vegetal e que a toxicidade não era a regra e sim provavelmente a exceção (VAN DAMME, PEUMANS, *et al.*, 1998). A maioria das inovações na pesquisa em lectina vegetal tem sido alcançada em lectinas de leguminosas. Por exemplo, a primeira evidência de que as lectinas são proteínas de ligação a carboidrato surgiram da observação de que a sacarose da cana de açúcar inibia a atividade aglutinante da ConA (SUMMER e HOWELL, 1936; VAN DAMME, PEUMANS, *et al.*, 1998).

No final da década de 40, dois grupos trabalhando separadamente descobriram que algumas hemaglutininas exibem uma clara preferência para eritrócitos de um particular tipo sanguíneo humano, dentro do sistema ABO (RENKONEN, 1948; BOYD e REGUERA, 1949). A descoberta desta especificidade foi o motivo direto da introdução do termo **lectina** (do latim *legere*, que significa escolher), no início da década de 50, para referência a essas proteínas de origem vegetal que aglutinam células (VAN DAMME, PEUMANS, *et al.*, 1998).

Apesar da observação de Summer e Howell, em 1936, de que o açúcar da cana inibia a atividade hemaglutinante da Conavalina A (ConA), foi demonstrado apenas em 1952, por Watkins e Morgan, que as propriedades aglutinantes das lectinas são baseadas na sua atividade específica de ligação a açúcar. As lectinas então passaram a ser reconhecidas como proteínas de ligação a carboidratos e assim puderam ser distinguidas de outras proteínas com base em um critério funcional bem definido (VAN DAMME, PEUMANS, *et al.*, 1998).

### **2.1.2. Definição e Classificação**

Existem três características que satisfazem a definição de lectinas. Primeiro, as lectinas são proteínas ou glicoproteínas que se ligam a carboidratos. Essas (glico)proteínas também podem ter outra atividade biológica como enzimática, por exemplo, mas esse requisito exclui moléculas não protéicas que aglutinam células. Segundo, As lectinas não se enquadram na categoria de imunoglobulinas. Inicialmente, as lectinas eram consideradas substâncias do tipo anticorpos devido

sua especificidade. Contudo, elas não precisam de estímulo antigênico para serem sintetizadas, deste modo, não podendo ser consideradas imunoglobulinas. Terceiro, as lectinas não modificam quimicamente os carboidratos aos quais se ligam. Esse requisito exclui as enzimas que são específicas para carboidratos que eventualmente podem aglutinar células em certas condições (RÜDIGER e GABIUS, 2001 *apud* TRINDADE, 2005).

Um pré-requisito para a definição de lectinas seria a existência de mais de um sítio de ligação a açúcar, excluindo assim várias classes de proteínas monovalentes e enzimas (GOLDSTEIN, HUGHES, *et al.*, 1980; LIENER, SHARON e GOLDSTEIN, 1986). A definição mais recente conferida as lectinas, que já vinha sendo defendida por outros autores tendo como única exigência à presença de pelo menos um domínio sem atividade catalítica, que se ligue reversível e especificamente a mono ou oligossacarídeos (BARONDES, 1988; PEUMANS e VAN DAMME, 1995). Desta forma são consideradas como lectinas muitas proteínas com diferentes propriedades de aglutinação e precipitação. Uma posterior subdivisão das lectinas em merolectinas, hololectinas, quimerolectinas, superlectinas e Multilectinas foi proposta, baseada na estrutura geral das lectinas maduras (VAN DAMME, PEUMANS, *et al.*, 1998; MONTEIRO-MOREIRA, 2002) (FIGURA 1).

As **merolectinas** consistem de um único domínio de ligação a carboidrato. Por definição elas são monovalentes e, portanto, não podem precipitar glicoconjugados ou aglutinar células (FIGURA 1-A). Um exemplo de uma típica merolectina é pequena proteína de ligação à quitina, heveína, extraída do látex da seringueira (*Hevea brasiliensis*) (VAN DAMME, PEUMANS, *et al.*, 1998).

As **hololectinas** são também exclusivamente constituídas de domínios de ligação a carboidratos, no entanto, contendo pelo menos dois destes domínios, que podem ser tanto idênticos ou muito similares e se ligam ao mesmo açúcar ou a açúcares estruturalmente similares (FIGURA 1-B). Por definição são di- ou multivalentes e, portanto precipitam glicoconjugados. A maioria das lectinas vegetais pertence a este subgrupo (VAN DAMME, PEUMANS, *et al.*, 1998).

As **quimerolectinas** são proteínas fusionadas consistindo de um ou mais domínios de ligação a carboidratos aleatoriamente arranjados a um domínio não relacionado (FIGURA 1-C). Tal domínio pode ter uma atividade enzimática bem definida ou outra qualquer, no entanto, agindo de forma independente do domínio de ligação a carboidrato. Dependendo do número de sítios de ligação a carboidratos, as

quimerolectinas podem se comportar como merolectinas ou hololectinas. Por exemplo, RIPSs do tipo 2 com mais de um sítio de ligação a carboidratos por protômero, se comporta como uma verdadeira hololectina, enquanto que quitinases vegetais da classe 1 com um único domínio atua como uma merolectina (VAN DAMME, PEUMANS, *et al.*, 1998).

As **superlectinas** são consideradas um tipo especial de quimerolectinas. Elas seriam proteínas de fusão, com dois domínios de ligação a carboidratos (FIGURA 1-D). Estes domínios são estruturalmente distintos e reconhecem carboidratos estruturalmente diferentes. Apenas um exemplo de superlectinas foi, até hoje, descrito, a lectina de bulbo de tulipa, com dois domínios, um específico para manose e outro específico para N-acetil-glicosamina (VAN DAMME, PEUMANS, *et al.*, 1998).

Mais recentemente, (MONTEIRO-MOREIRA, 2002) ampliou a classificação, criando o termo **multilectinas**, ao mostrar a existência de lectinas possuindo dois ou mais domínios ligantes a carboidratos, idênticos, mas que podem se ligar a açúcares diferentes (FIGURA 1-E). É o caso da frutalina, uma lectina de sementes de *Artocarpus incisa* - (MOREIRA, CASTELO-BRANCO, *et al.*, 1998) que ligam tanto D-galactose como D-manose.

As lectinas também são classificadas de acordo com os monossacarídeos a qual elas se ligam podendo formar quatro grupos, baseado na configuração do C3 e C4 do anel piranosídico (MAKELA, 1957). Ao grupo I pertencem as lectinas específicas para L-fucose, como a de *Lotus tetragonolobus* e *Ulex europaeus*. O grupo II é correspondente as lectinas ligam lactose e N-acetil-glicosamina (*Glycyne max* e *Phaseolus lunatus*). As lectinas que se ligam a manose e glicose, como as de *Pisum sativa*, *Lens culinaris*, *Vicia faba*, *Canavalia ensiformis* e *Dioclea grandiflora* etc., são ditas do grupo III. Lectinas que interagem com açúcares do grupo IV (idiose, gulose, L-xylose, L-glicose) ainda não foram relatadas. Nesta classificação não estavam incluídas as lectinas específicas por ácido siálico e aquelas específicas por estruturas complexas (GOLDSTEIN e PORETZ, 1986). Esta subdivisão é certamente muito útil para a utilização de lectinas com ferramentas, no entanto é provavelmente artificial e irrelevante com respeito às possíveis relações evolutivas (VAN DAMME, PEUMANS, *et al.*, 1998).

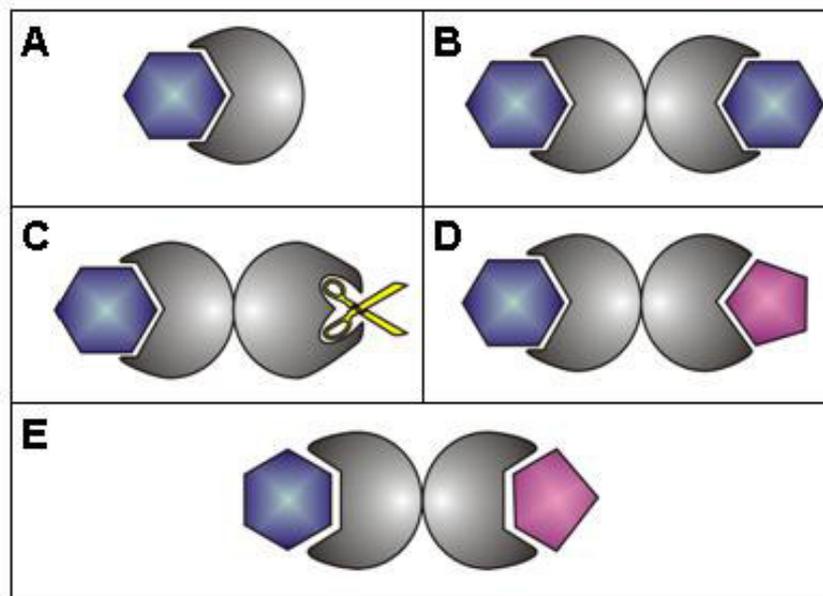


Figura 1 - Classificação das lectinas segundo (VAN DAMME, PEUMANS, *et al.*, 1998) e MONTEIRO-MOREIRA (2002). A – Merolectinas capazes de se ligar a apenas um carboidrato (aqui representado em hexágono), B – Hololectinas capazes de se ligar a dois carboidratos simultâneos, C – Um exemplo de quimerolectina (RIP), D – Superlectinas capazes de se ligarem simultaneamente a carboidratos diferentes com pelo menos duas subunidades distintas e E – Multilectinas onde diferentes carboidratos podem ocupar o mesmo sítio.

O recente aumento do número de sequências disponíveis de lectinas de plantas permitiu inúmeros estudos de comparação e relacionamento evolutivo, tanto a nível primário (sequência linear dos aminoácidos ou nucleotídeos) como a nível estrutural (comparação entre estruturas secundárias, terciárias e quaternárias), que promoveram a distinção de sete famílias de lectinas evolutivamente relacionadas: lectinas de leguminosas, lectinas ligante de manose de monocotiledôneas, lectinas de quitina composta por domínios heveínicos, RIPs (*Ribosome Inactivating Proteins*) tipo 2, lectinas relacionadas à jacalina, lectinas da família das amarantinas e lectinas de floema de Cucurbitaceae, essas três últimas compostas por poucos membros (VAN DAMME, PEUMANS, *et al.*, 1998).

### **2.1.3. Função Biológica**

Devido a sua habilidade de decifrar glico-códigos codificados pela grande variedade de glicanos anexados a glicoconjugados de membrana solúveis e integrados, as lectinas participam em numerosos processos celulares tais como: comunicação, defesa, fertilização, desenvolvimento, etc (CALVETE, THOLEM, *et al.*, 1999). As lectinas provavelmente desempenham papéis biológicos fundamentais, porque elas são encontradas em diversas espécies, em diferentes órgãos e tecidos (CHRISPEELS e RAIKHEL, 1991).

Praticamente todas as células carregam carboidratos na sua superfície na forma de glicoproteínas, glicolipídeos e polissacarídeos. Vários relatos indicam que estes carboidratos têm a habilidade de codificar informações biológicas (SHARON, 1980); (COOK, 1986). Tem se tornado gradualmente claro que as lectinas desempenham dois principais papéis nas plantas. Primeiro, elas são proteínas de armazenamento que podem ser mobilizadas para o crescimento e desenvolvimento vegetal. Segundo, elas são defesas dos vegetais contra herbívoros e patógenos (MURDOCK e SHADE, 2002).

Quando as lectinas ocorrem em altas concentrações em tecidos vegetais, elas indubitavelmente representam uma forma de armazenamento de proteínas (VAN DAMME, PEUMANS, *et al.*, 1998). Evidências para este fato são: Primeiro, as lectinas geralmente ocorrem em maior abundância em sementes e outros órgãos vegetativos de armazenamento; Segundo, elas são encontradas em organelas sub-celulares que são prováveis sítios de armazenamento; Terceiro, elas se acumulam

durante a fase de crescimento e desenvolvimento ou reprodutiva do ciclo de vida do vegetal e são mobilizadas e utilizadas posteriormente (VAN DAMME, PEUMANS, *et al.*, 1998; MURDOCK e SHADE, 2002).

Alguns autores defendem que o fato das lectinas serem proteínas de armazenamento é perfeitamente compatível com o seu provável papel de defesa. Tubérculos e especialmente sementes, são repositórios de energia e aminoácidos reservados para o crescimento da planta. Sementes representam adicionalmente, o veículo pelo quais muitas plantas passam seus genes às gerações posteriores. Portanto, esses órgãos vitais precisam de proteção. As lectinas são sintetizadas, durante o desenvolvimento das sementes, juntamente com as mais abundantes proteínas de armazenamento, compartilhando a mesma localização intracelular e sistemas de regulação temporal e espacial de sua expressão (VAN DAMME, PEUMANS, *et al.*, 1998). Durante a germinação e o crescimento da plântula, ambas as proteínas de armazenamento e lectinas são degradadas para compor o suprimento de aminoácidos (CHRISPEELS e RAIKHEL, 1991). Portanto, as lectinas ajudam na defesa da planta e, no tempo certo, as mesmas podem suprir o material e a energia necessária para o crescimento, desenvolvimento e reprodução (MURDOCK e SHADE, 2002).

Entre outros papéis importantes das lectinas nos vegetais, podem ser citados: transporte de carboidratos nas sementes e ligação de bactérias fixadoras de nitrogênio aos pêlos radiculares. Nas leguminosas em particular, evidências experimentais têm demonstrado que as lectinas expressas na raiz atuam como mediadores da simbiose entre bactérias fixadoras de nitrogênio do gênero *Rhizobium* e as raízes da planta. Estudos como o de (DIAZ, MELCHERS, *et al.*, 1989), defendiam a teoria de que as lectinas eram responsáveis pela específica interação entre as leguminosas e sua bactéria fixadora simbiote. No entanto, a descoberta que a especificidade do hospedeiro (Leguminosa) das cepas de *Rhizobium* depende da estrutura dos fatores *nod* (também chamados de oligossacarídeos de lipoquitina) que estas secretam quase que exclui um envolvimento das lectinas das leguminosas no processo (VAN DAMME, PEUMANS, *et al.*, 1998). Anteriormente, então, presumiu-se que as lectinas eram determinantes na aglutinação de grande número de bactérias na superfície do pêlo radicular, um papel que teria evoluído da habilidade destas moléculas de aglutinar e imobilizar bactérias com uma reação de defesa (CHRISPEELS e RAIKHEL, 1991). No entanto,

tal teoria também foi questionada pela descoberta de um tipo de proteína de ligação a carboidrato que possui uma alta afinidade para os fatores *nod* nas raízes de *Dolichos biflorus* (ETZLER, MURPHY, *et al.*, 1997).

O envolvimento das lectinas de leguminosas em processos específicos de reconhecimento dentro da planta é discutível, devido a duas principais observações: Primeiro, a abundância das lectinas não corresponde ao provável número de receptores específicos (GOLDSTEIN, 2002); Segundo, que a maioria das lectinas de leguminosas possui uma especificidade preferencial a carboidratos que estão ausentes nas plantas, que, no entanto, são importantes componentes dos glicoconjugados animais (VAN DAMME, PEUMANS, *et al.*, 1998). Essas observações favorecem ainda mais o papel de defesa das lectinas.

A existência de sítios hidrofóbicos dentro da estrutura de algumas lectinas de leguminosas, que se liga a fitohormônios como o ácido indolacético e citocininas, que modulam sua atividade hemaglutinante e sua interação com carboidratos, sugere um possível papel para as lectinas em respostas reguladas por hormônios (CAVADA, MOREIRA, *et al.*, 1993), e em importantes processos fisiológicos.

As dúvidas sobre a possibilidade de lectinas de plantas desempenharem função endógena também têm recebido atenção e gerado idéias sobre o provável envolvimento de lectinas na regulação e sinalização celular em plantas. Estudos relatam que plantas como arroz, *Helianthus tuberosus*, e *Iponema batatas*, quando sujeitas ao estresse biótico ou abiótico, reagem sintetizando pequenas quantidades de lectinas ligantes de manose relacionadas à jacalina citoplasmáticas e nucleares. Experimentos com *Nicotiana tabacum* também revelaram que o éster metil jasmonato induz a expressão de uma lectina específica para oligômeros de N-acetilglicosamina nas folhas, que não pode ser detectada na amostra-controle (VAN DAMME, BARRE, *et al.*, 2004). Essa lectina se localiza no núcleo e citoplasma, ocorrendo em todas as células de parênquima, mas não nos tecidos vasculares. Lectinas ligantes de manose de monocotiledôneas também foram encontradas no citoplasma de plantas, com uma característica comum: nenhuma delas possuía peptídeo-sinal. Assim, estes estudos indicam que lectinas de plantas também podem estar envolvidas em interações endógenas proteína-carboidrato e que também estariam implicadas na evolução das lectinas de plantas clássicas, fazendo desta abordagem, um campo fértil para novas descobertas na área de fisiologia de plantas (TRINDADE, 2005).

#### 2.1.4. Aplicações das lectinas

Esta grande variedade de estrutura dos carboidratos associada a glicoconjugados de superfície tem o potencial de modificar a atividade de proteínas aos quais estão anexados. Portanto, estas moléculas podem desencadear processos de diferenciação celular, desenvolvimento, caracterização de doenças, endereçamento de drogas e identificação de estados patológicos (HUSSAIN, JANI e FLORENCE, 1997). Por estas razões, o reconhecimento celular pelas lectinas parece ser um evento central em uma grande variedade de fenômenos biológicos em tecidos normais ou enfermos (SINGH, TIWARY e KENNEDY, 1999).

Lectinas ligantes de quitina, por exemplo, têm sido isoladas de diversas fontes e muitas delas apresentam atividade antifúngica e inseticida, uma vez que a quitina é o componente-chave da parede celular de fungos e presente no exoesqueleto de insetos (TRINDADE, 2005). Lectinas de plantas, em particular, têm sido estudadas sob vários aspectos, por exemplo, seu potencial antifúngico e inseticida para proteger as plantas que as produzem naturalmente ou pela introdução de material genético que codifique a expressão desse tipo de lectina atóxica ao homem (PEUMANS e VAN DAMME, 1996; CARLINI e GROSSI-DE-SA, 2002).

Heveína é uma proteína, bastante estável, ligante de quitina encontrada na seringueira (*Hevea brasiliensis*) e apontada como envolvida na defesa da planta ao ataque de insetos e fungos (NEWMAN, LEHR, *et al.*, 2004 *apud* TRINDADE, 2005). Além disso, o domínio heveínico é o ancestral comum de muitas lectinas vegetais evolutivamente relacionadas (TRINDADE, 2005).

Muitos dos antígenos associados aos processos tumorais constituem estruturas glicídicas expressas na forma de glicoproteínas ou glicolipídeos na superfície de células tumorais. Dentre essas estruturas glicídicas o Antígeno *T* (Gal $\beta$ 1 $\rightarrow$ 32-GalNAc $\alpha$ O-Serina/Treonina) e o Antígeno *Tn* (2-GalNAc $\alpha$ O-Serina/Treonina) foram identificados em 90% dos adenocarcinomas (câncer maligno que se origina em tecido glandular), inclusive aqueles no ovário e mama (AVICHEZER e ARNON, 1996) *apud* (SANTANA, 2004).

Os antígenos *T* e *Tn* podem ser reconhecidos por lectinas Gal/2-Gal Nac específicas tais como a lectina da *Vicia villosa* B4 (VVL B4) e a lectina da *Arachis hypogaea* (PLN). Essas lectinas foram utilizadas como ferramentas de estudo desses antígenos em células de carcinoma ovariano, sendo significativamente

reativas em linhagens desses carcinomas (AVICHEZER e ARNON, 1996). Nesses estudos, metodologias de marcação de lectinas com compostos radioativos, tais como o  $^{125}\text{I}$ , foram utilizadas para procedimento de radioimunoensaios e autoradiografias.

A lectina da *Dolichos biflorus* ou DBL possui afinidade muito baixa para a Gal e muito maior para a 2-GalNAc, portanto é capaz de distinguir Gal de 2-GalNAc. Mesmo com essa especificidade a DBL não se liga ao antígeno *Tn*, no entanto a DBL possui afinidades de ligação maiores para o dissacarídeo de Forssman (D-2-NacGal- $\alpha$ [1 $\rightarrow$ 3]-D2-NacGal) e para o Antígeno A (sistema *ABO*); sua ligação com o antígeno *Tn* é muito baixa ou inexistente (WU e SUGII, 1991).

## 2.2. Lectinas de Leguminosas

### 2.2.1. Estrutura e propriedades químicas

As lectinas desta família se assemelham em suas propriedades fisiológicas embora se diferenciem em suas especificidades a carboidratos devido a variações de aminoácidos envolvidos no sítio de ligação a carboidratos (SHARMA e SUROLIA, 1997). Apresentam um variado e amplo grau de especificidades e além das lectinas que se ligam a manose e das lectinas específicas por manose/maltose, são encontrados todos os outros grupos com especificidades diversas por monossacarídeos. Frequentemente são metaloproteínas onde cada subunidade contém um íon de  $\text{Mn}^{2+}$  e um de  $\text{Ca}^{2+}$ , que são essenciais para a atividade de ligação a carboidrato da lectina. Algumas lectinas de leguminosas são N-glicosiladas (SHARON e LIS, 1990).

A arquitetura deste monômero é usualmente descrita como constituída de duas folhas-  $\beta$  pregueadas. No entanto, (BANERJEE, DAS, *et al.*, 1996) mostraram que a correta descrição envolve três folhas- $\beta$  pregueadas: um conjunto posterior de 6 segmentos, um anterior de 7 segmentos e uma folha menor de 5 segmentos que tem o papel de manter as duas primeiras unidas. Um núcleo hidrofóbico está localizado entre as folhas frontal e posterior. Não existem regiões de  $\alpha$ -hélice e 50% dos resíduos estão nas regiões de *loops*. Um desses *loops* se curva sobre a folha anterior resultando na formação de um segundo núcleo hidrofóbico entre eles.

O domínio de ligação a carboidratos das lectinas de leguminosas é uma depressão rasa sobre os *loops* associados com a face côncava da folha- $\beta$  pregueada de 7 segmentos (SHARMA e SUROLIA, 1997). Dois resíduos altamente conservados, Asp e Asn, são cruciais para o reconhecimento dos carboidratos já que eles, juntamente com um NH da cadeia principal de um resíduo Gly ou Arg, participam em quatro ligações de hidrogênio chaves com o monossacarídeo. A disposição espacial idêntica destes resíduos conservados tem sido atribuída a impedimentos impostos pela coordenação com  $\text{Ca}^{2+}$  e outro metal de transição, normalmente o  $\text{Mn}^{2+}$ . A atividade de ligação a carboidrato das lectinas de leguminosas depende da presença destes metais e seus sítios de ligação são extremamente conservados em todas as lectinas desta família. A conservação dos resíduos de aminoácidos nas lectinas de leguminosas que mostram especificidades diferentes frente a carboidratos explicita a habilidade única destas proteínas de utilizarem uma sequência comum de resíduos para ligação e ainda discriminar entre diferentes monossacarídeos.

Como as lectinas são um grupo de proteínas muito heterogêneas, várias tentativas já foram feitas para agrupá-las em subgrupos de acordo com suas características bioquímicas/biofísicas, estruturas moleculares, especificidade a carboidrato e propriedades biológicas.

### **2.2.2. Biossíntese e processamento**

As lectinas de leguminosas são sintetizadas no retículo endoplasmático rugoso como pré-pro-proteínas. A pré-pro-lectina possui na sua extremidade N-terminal uma sequência de 20 a 30 resíduos de aminoácidos, denominada de peptídeo sinal, o qual é removido co-traducionalmente durante o transporte da cadeia polipeptídica nascente para o interior do RER. A forma resultante, chamada de pro-lectina ou precursor, pode ou não receber, nesse momento, uma ou mais cadeias laterais de oligossacarídeos em sequências específicas de sua estrutura (os sítios de glicosilação). A pro-lectina é então transportada para os corpos protéicos, organela onde a lectina madura (ou seja, completamente processada) irá ser estocada. Durante o transporte para os corpos protéicos, a porção glicídica dos precursores glicosilados é geralmente modificada, quando da sua passagem pelo complexo de Golgi, pela adição e/ou remoção de resíduos específicos do

oligossacarídeo. Se depois de tais modificações o oligossacarídeo for removido quando da chegada da proteína aos corpos protéicos, a lectina madura será não-glicosilada, apesar de ter sido originada de um precursor glicosilado. A não remoção do oligossacarídeo resulta numa lectina madura glicosilada (NOGUEIRA, 1999).

Tanto os precursores glicosilados como os não glicosilados sofrem concomitantemente, ao longo do transporte para o seu destino final, clivagens e/ou ligações proteolíticas, o que resulta numa lectina madura com uma cadeia polipeptídica menor em relação a pro-lectina. Assim, o processamento das lectinas de leguminosas é bastante complexo. Não há exemplos de lectinas que não apresentem processamento após a remoção co-traducional do peptídeo sinal, ou que apresentem somente uma clivagem proteolítica. A conversão do produto traducional primário na lectina madura requer a remoção proteolítica de um ou mais peptídeos, geralmente com 5 a 10 resíduos (VAN DAMME e PEUMANS, 1996). Vários tipos diferentes de processamento podem ser observados em lectinas de leguminosas, os mais comuns são: remoção de um peptídeo C-terminal; remoção de um peptídeo interno e na extremidade C-terminal; remoção de um peptídeo interno.

A perda dos aminoácidos da região carboxi-terminal por proteólise das “*ragged ends*” é um evento comum na biossíntese de lectinas de leguminosas, sugerindo que essas extremidades estão expostas na superfície da lectina e livremente estruturadas (YOUNG, WATSON, *et al.*, 1995). O resultado da heterogeneidade pode ser o principal contribuinte para o aparecimento de isolectinas. Em SBA, quatro formas truncadas foram identificadas, com 240, 243 e 252 aminoácidos, respectivamente. Em adição, a forma nativa com 250 resíduos. No caso da PNA, o cDNA que codifica 250 resíduos, entre 12 e 15 resíduos C-terminal são removidos. A perda da extremidade C-terminal foi também observada em formas recombinantes dessas e de outras lectinas de uma cadeia quando expressas em *E. coli* (SHARON e LIS, 2003).

As lectinas de *Vicia faba* e *Pisum sativum*, representantes da subtribo Vicieae, são sintetizadas como pro-lectinas com massas moleculares aparentes de 29 e 25 KDa, respectivamente. A cadeia polipeptídica dos precursores sofre clivagem co- e pós-traducional, dando origem às cadeias polipeptídicas beta (21 e 17 KDa, respectivamente) e alfa (5,7 e 6 kDa, respectivamente), que constituem os dois tipos de cadeias polipeptídicas encontradas nas lectinas maduras. As cadeias beta e alfa estão ligadas covalentemente na pró-lectina, de modo que a

cadeia beta ocupa a porção N-terminal seguida da cadeia alfa (HEMPERLY, MOSTOV e CUNNINGHAM, 1982; HIGGINS, CHANDLER, *et al.*, 1983).

Um tipo de processamento também encontrado nas lectinas de leguminosas compostas de uma cadeia é observado na região C-terminal. A sequência do cDNA de algumas lectinas indicam cadeias polipeptídicas maiores do que as encontradas pela sequência direta da proteína. Este processamento é frequentemente impreciso, resultando em heterogeneidade que pode explicar a presença de isoformas ou mesmo múltiplas formas com extremidades C-terminal de tamanhos variados (YOUNG, WATSON, *et al.*, 1995). Um exemplo interessante de processamento pós-traducional na região C-terminal é observado nas lectinas de semente (DBL) e de tecidos vegetativos (DB58) de *Dolichos biflorus*. Cada uma destas lectinas é sintetizada como uma subunidade de massa molecular maior (subunidade I no caso da lectina de sementes e a no caso da DB58). Estas subunidades sofrem uma clivagem na porção C-terminal produzindo as cadeias II e b, respectivamente. Entretanto, essa clivagem ocorre somente na metade das subunidades, resultando na produção de lectinas heterooligoméricas (ELTZER, 1996).

Na família das lectinas de leguminosas o processamento observado na ConA é o mais complexo, diferindo dos tipos de processamento pós-traducionais relacionados anteriormente (VAN DAMME e PEUMANS, 1996). De fato, o primeiro exemplo da ocorrência de excisão de sequências internas e religação dos fragmentos formados (*protein splicing*) em proteínas foi encontrado durante o processamento pós-traducional da ConA (HERMAN, SHANNON e CHRISPEELS, 1985; CHRISPEELS, HARTL e FAYE, 1986). A FIGURA 2 mostra de maneira esquemática os eventos co- e pós-traducionais que ocorrem durante a biossíntese da ConA e lectinas relacionadas.

Experimentos de marcação radioativa e sequenciamento de aminoácidos das formas intermediárias detectadas em sementes em desenvolvimento de *C. ensiformis* permitiram desvendar o mecanismo do processamento proteolítico pós-traducional da forma precursora da ConA (FIGURA 2-B). O produto primário de tradução do mRNA da ConA é uma proteína com 290 resíduos de aminoácidos, a pré-pro-ConA. O peptídeo sinal com 30 resíduos de aminoácidos é removido co-traducionalmente durante o trânsito da cadeia polipeptídica nascente para o lúmen do RER. A forma resultante, denominada pro-ConA ou precursor, é então glicosilada pela ligação covalente de um oligossacarídeo rico em manose ao resíduo Asn153,

localizado na região mediana da pro-ConA. Durante o transporte para os corpos protéicos, via complexo de Golgi, o glicopeptídeo central é removido originando dois polipeptídeos. Esses dois polipeptídeos (metades N- e C-terminais da pro-ConA) são então processados proteoliticamente pela remoção das sequências C-terminais com 11 e 9 resíduos de aminoácidos, respectivamente. Após o processamento C-terminal, as duas metades sofrem uma transposição e são religadas na ordem inversa. Uma clivagem proteolítica adicional remove 4 resíduos da extremidade N-terminal da cadeia polipeptídica já religada, originando a lectina madura não glicosilada, cuja subunidade possui 237 resíduos de aminoácidos (cadeia alfa). Todas as clivagens ocorrem do lado carboxil de resíduos de asparagina da pro-ConA. Uma asparaginil endopeptidase com tais características foi purificada de sementes de *Canavalia ensiformis* (ABE, SHIRANE, *et al.*, 1993), e parece ser a enzima responsável *in vitro* pelo processamento proteolítico da pro-conA (MIN e JONES, 1992).

A religação mencionada anteriormente ocorre na forma de uma nova ligação peptídica entre os resíduos 118 e 119 da lectina madura. Entretanto, a formação de uma nova ligação peptídica não ocorre com uma eficiência completa e, como consequência, algumas metades não são religadas. Assim, quando submetida a eletroforese em condições desnaturantes, a ConA madura apresenta um padrão de três bandas com massas moleculares de 29-30 kDa, (cadeia polipeptídica intacta com 237 resíduos, denominada cadeia alfa), 16-18 kDa (metade N-terminal, os resi-

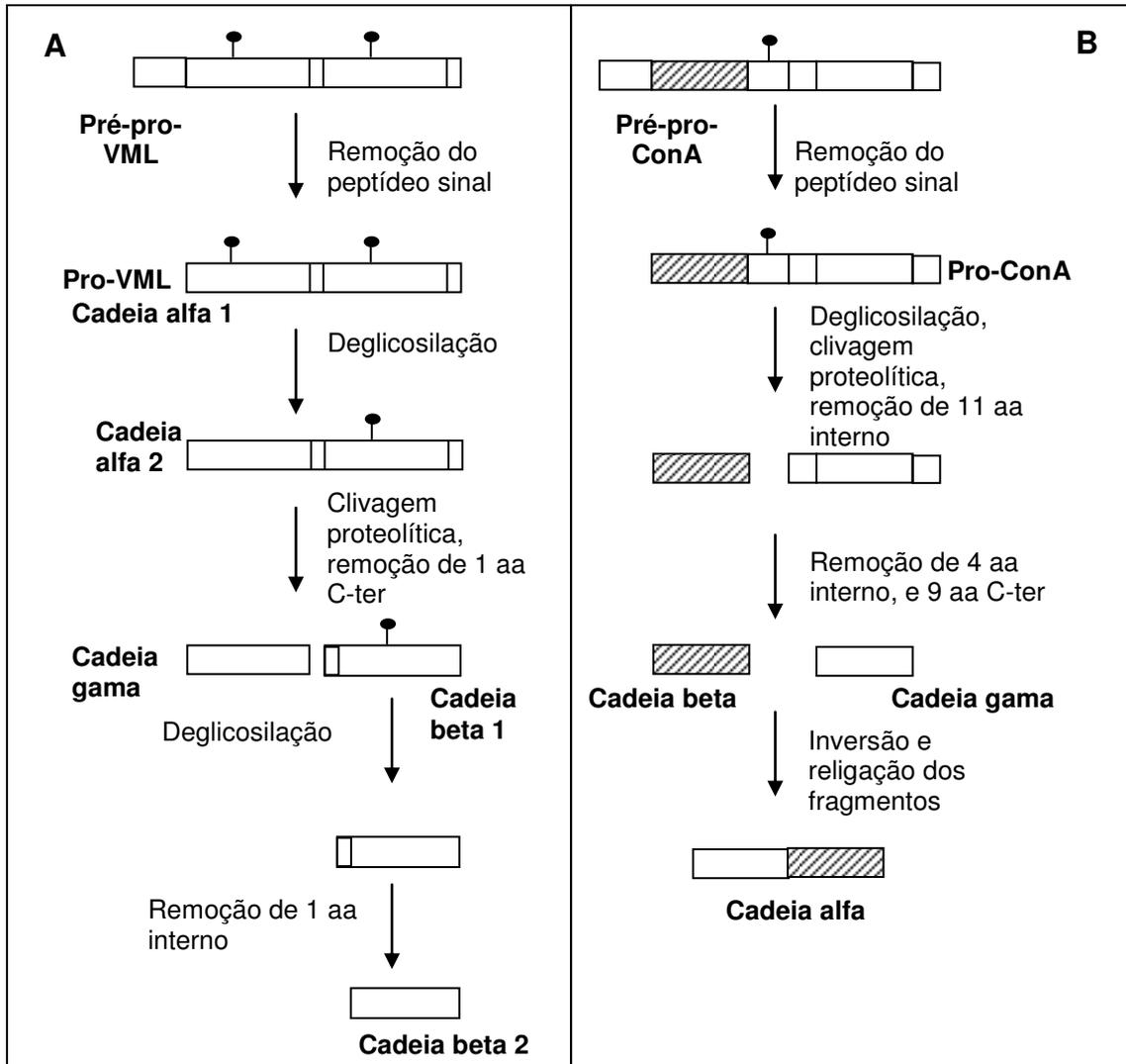


Figura 2 – Representação esquemática dos eventos pós-traducionais que ocorrem durante a biossíntese da VML (painel A) e ConA (painel B). Modificado de: (VAN DAMME, PEUMANS, *et al.*, 1998; CALVETE, SANTOS, *et al.*, 1998).

duos 119-237, denominada de fragmento gama). Como consequência, na semente madura existe uma mistura de subunidades intactas, formadas pela cadeia alfa com 237 aminoácidos, e subunidades fragmentadas, formadas pelos fragmentos beta (resíduos 1-118) e gama (resíduos 119-237). Apesar de não estarem ligados covalentemente, os fragmentos beta e gama são mantidos juntos por ligações não covalentes, formando um protômero com uma estrutura tridimensional que é essencialmente a mesma subunidade intacta formada pela cadeia polipeptídica íntegra, sem descontinuidade em sua estrutura primária (BECKER, REEKE, *et al.*, 1975).

A lectina de *Vicia faba* é sintetizada similarmente a lectina de ervilha (ETZLER, 1992). O produto de transcrição primário é uma pré-pro-lectina (29 kDa), no qual a sequência sinal amino terminal de 29 resíduos é seguido pela cadeia beta e então a cadeia alfa da lectina madura. O processamento envolve a remoção da sequência sinal e a clivagem da ponte peptídica entre a cadeia alfa e beta na posição 183, sem posteriores mudanças nessas cadeias (SHARON e LIS, 2003).

O transporte correto da pro-conA do retículo endoplasmático para os corpos protéicos, via complexo de Golgi, depende da presença do oligossacarídeo (FAYE e CHRISPEELS, 1987). Por outro lado, a lectina de *Phaseolus vulgaris* (PHA) é também sintetizada na forma de um precursor glicosilado, porém os oligossacarídeos (em número de 2, ligados aos resíduos Asn12 e Asn60) não são removidos durante o transporte da PHA para os corpos protéicos, e a PHA madura é uma glicoproteína. Assim, ao contrário do que ocorre com a ConA, a glicosilação da pro-PHA não é necessária para o seu transporte para os corpos protéicos (BOLLINI, CERIOTTI, *et al.*, 1985; FAYE, STURM, *et al.*, 1986; VOELKER, HERMAN e CHRISPEELS, 1989; *apud* NOGUEIRA, 1999).

A determinação da sequência de aminoácidos dos precursores das lectinas de *Canavalia gladiata* (YAMAUCHI, NAKAMURA, *et al.*, 1989) e *C. brasiliensis* (GRANGEIRO, SCHRIEFER, *et al.*, 1997), por dedução a partir da sequência de nucleotídeos dos respectivos cDNAs, e sua comparação com as sequências de aminoácidos das lectinas maduras, também revelou que as metades N- e C-terminais estão ligadas numa ordem inversa no precursor, sugerindo que o mesmo processamento pós-traducional da ConA também ocorre nessas espécies.

### 2.2.3. Genes de lectinas de leguminosas

A maioria dos tecidos vegetais contém uma única lectina, embora ocasionalmente duas (ou mais) lectinas que diferem em sua especificidade por açúcar e outras propriedades são encontradas no mesmo tecido (PEUMANS, ZHANG, *et al.*, 2000). Lectinas únicas, em particular de leguminosas, ocorrem ocasionalmente como uma mistura de isoformas referidas como **isolectinas**. Tipicamente, isolectinas tem uma estrutura molecular similar, embora possa diferir fracamente na sua especificidade ou alguma propriedade física como carga elétrica. Elas podem, então, ser separadas por cromatografia de troca iônica, como a que foi encontrada em WGA ou por cromatografia de afinidade em açúcares imobilizados (SHARON e LIS, 2003).

Duas lectinas distintas ocorrem nas sementes de *Ulex sp*, *Artocarpus sp* e *Vicia cracca*, enquanto em sementes de *Griffonia simplicifolia* três diferentes lectinas (em adição ao número de isolectinas) estão presentes. Lectinas com especificidade diferente são também encontradas em *Sophora japonica*. Em uma mesma planta, lectinas não são necessariamente confinadas ao mesmo tecido. Casos são conhecidos quando lectinas encontradas em partes vegetativas da planta são idênticas com aquelas encontradas em sementes (por exemplo, as três lectinas de semente de *Griffonia simplicifolia* estão também presentes em folhas da mesma planta), mas isto não é sempre assim. Por exemplo, em *Dolichos biflorus*, a planta mais extensivamente estudada com respeito à distribuição de lectinas em vários tecidos, as folhas contem uma lectina (DB58) homóloga àquela da semente (DBL), mas com algumas diferenças em sua especificidade fina (ETZLER, MURPHY, *et al.*, 1997). Adicionalmente, uma lectina de raiz (LNP) tem sido encontrada na mesma parte da planta que é distinta das lectinas de semente tanto na composição de aminoácidos, peso molecular, ponto isoelétrico e especificidade (ETZLER, KALSI, *et al.*, 1999 *apud* SHARON e LIS, 2003).

Algumas lectinas se originam de genes distintos, como no caso daquelas da WGA. Outras resultam das modificações pós-traducionais diferenciais de uma única lectina (YOUNG, WATSON, *et al.*, 1995) ou da diferente reunião de subunidades proximamente relacionadas. As duas isolectinas de *Psophocarpus tetragonolobus* WBA-I e WBA-II diferem principalmete no seu ponto isoelétrico, um sendo ácido e outro básico, enquanto que as isolectinas de semente de *Maackia*

*amurensis* diferem em sua especificidade por célula, uma (MAH) sendo uma hemaglutinina e outra (MAL) tendo uma atividade leucoaglutinante (por exemplo, a habilidade para aglutinar preferencialmente células brancas do sangue tais como linfócitos) (VAN DAMME, PEUMANS, *et al.*, 1998). GLS-I consiste de uma família de cinco isolectinas, cada um tetrâmero de um ou dois tipos de subunidades, A e B que são muito similares em tamanho molecular e composição de aminoácidos, mas diferem em especificidade por carboidrato.

As isolectinas de PHA também representam uma família de cinco proteínas tetraméricas com variedade nas proporções de duas classes de subunidades, E e L. Eles diferem em carboidrato e especificidade celular., tanto quanto nas suas propriedades biológicas: E4 (E-PHA) é uma potente hemaglutinina, L4 (L-PHA) tem uma atividade leucoaglutinina e tem efeito mitogênico; as formas intermediárias (E1L3, E2L2 ou E3L) possuem níveis inferiores das atividades citadas. Misturas similares de isoformas são encontradas nas sementes de *Vicia villosa* e na casca de *Maackia amurensis*, *Robinia pseudoacacia* e *Sophora japonica* (VAN DAMME, PEUMANS, *et al.*, 1998). Nas sementes de *Datura stramonium* três isolectinas estão presente, duas das quais são homodímeros, compostas também de subunidades A ou B, enquanto que a terceira é um heterodímero abrangendo ambas subunidades. Os dois tipos de lectinas (toxina e aglutinina) encontradas, entre outras, em sementes de *Ricinus communis* e de *Abrus precatorius* podem ser consideradas como um caso especial de isolectinas (SHARON e LIS, 2003).

Considerável informação está disponível sobre os genes e biologia molecular de várias lectinas de leguminosas bem caracterizadas (ETZLER, 1992). A lectina de semente de ervilha, por exemplo, é codificada por uma pequena família de genes relacionados, o qual somente um é expresso. No caso da soja, dois genes relacionados, L1 e L2, estão presentes. L1 codifica para SBA, a lectina de semente, enquanto que a função da L2, o qual é expressa em baixos níveis, é desconhecida. Ambos os genes estão também presente em uma linhagem de soja deficiente de lectina. Nesta linhagem, contudo, o gene L1 contem um grande inserto (3,4 kb) na região codificadora que bloqueia sua transcrição. As subunidades E e L de PHA são codificadas por dois genes que estão localizados no mesmo cromossomo separados por 4kb. Há um alto grau de similaridade entre a região codificada pelo mRNA dos dois genes, sugerindo que eles são homólogos e derivados da duplicação de um gene ancestral (SHARON e LIS, 2003).

Em leguminosas que produzem ambas lectinas, de semente e tecido vegetativo, os dois são frequentemente codificados por genes distintos, como no caso da lectina de semente de *Dolichos biflorus* (DBL) e a lectina de folha estreitamente relacionada DB58. Esses dois genes estão localizados na mesma orientação transcricional separados aproximadamente por 3kb e mostram mais de 90% de similaridade de sequência, indicando que eles também podem ter surgido por duplicação gênica. Sua similaridade se estende as regiões flangeadoras dos genes, uma principal exceção sendo a presença de um segmento de 116 pb na região que flangeia a extremidade 5' do gene DBL mas faltando a mesma região no gene DBL58. O inserto contém uma porção de um elemento regulatório específico da semente de 22 pb identificado no gene da lectina da ervilha. Dos achados acima surge a possibilidade que as diferenças, tanto espaciais como temporais, na expressão das duas lectinas de *Dolichos biflorus* podem ter surgido por inserção, ou deleção, desses fragmentos durante a evolução (SHARON e LIS, 2003).

A formação de dois tipos diferentes de subunidades de DBL e DB58 é devido ao processamento C-terminal. Cada uma dessas lectinas é primeiro sintetizada como subunidades intactas e então metade das subunidades sofre remoção proteolítica pós-traducional de um peptídeo carboxi-terminal. O mecanismo que assegura que as subunidades intactas e truncadas se agrupem em uma razão equimolar (2:2 em DBA e 1:1 em DB58) não foi estabelecido. Uma proposta é que as subunidades se agrupam antes da etapa de processamento, levando somente uma das subunidades na lectina homotetramérica ou homodímero acessível ao processamento C-terminal (HAMELRYCK, LORIS, *et al.*, 1999 *apud* SHARON e LIS, 2003).

Genes de lectinas de leguminosas são desprovidos de íntrons. A síntese dessas lectinas é primeiramente regulada a nível transcricional (ETZLER, 1986). Por exemplo, nos cotilédones de *Phaseolus vulgaris*, soja e ervilha, os transcritos de mRNA hibridando com o respectivo cDNA da lectina começa a se acumular durante a meia-maturação da semente, então diminui na maturação tardia e desaparece depois da germinação. O aumento do mRNA coincide com o período de máxima produção das lectinas de semente (SHARON e LIS, 2003).

#### 2.2.4. Evolução molecular de lectinas de leguminosas

A árvore filogenética das lectinas de leguminosas (FIGURA 3) consiste de quatro ramificações principais. O ramo S (Sophoreae) agrupa as lectinas de *Maackia amurensis*, *Cladrastis lutea* e *Sophora japonica*. Uma posterior subdivisão desse ramo separa as lectinas de *Maackia amurensis* de *Sophora japonica* e *Cladrastis lutea*. As lectinas de especificidade diferente de *Sophora japonica* e *Maackia amurensis* parecem ter surgido por duplicação gênica partindo de um ancestral comum (VAN DAMME, PEUMANS, *et al.*, 1998).

O ramo P (Phaseoleae) inclui um extenso grupo de lectinas de três diferentes subtribos. O subgrupo Pb (Phaseoleae-Phaseoleae) agrupa as lectinas da tribo Phaseoleae, indicando que as lectinas de *Phaseolus vulgaris*, *Phaseolus lunatus*, *Dolichos biflorus*, e *Glycine max* provavelmente surgiram de um ancestral comum (FIGURA 3). *Phaseolus vulgaris* possui três tipos de genes distintos que codificam lectinas, arcelinas e inibidores de  $\alpha$ -amilase. Ambos os genes provavelmente vieram de um único ancestral gênico e se divergiram através de duplicação gênica. O subgrupo Pd (Phaseoleae-Diocleae) agrupa lectinas glucose/manose ligantes das tribos Diocleae e Sophoreae. O subgrupo Pe (Phaseoleae-Erythrina) corresponde a um pequeno grupo que inclui as lectinas galactose ligante de *Psophocarpus tetragonolobus* e *Erythrina corallodendron* (VAN DAMME, PEUMANS, *et al.*, 1998).

O ramo V (Vicieae) também inclui um extenso grupo de lectinas e diferentes tribos (FIGURA 3). O primeiro subgrupo Vv (Vicieae-Vicieae) agrupa as lectinas glucose-ligante da tribo Vicieae (incluindo os gêneros *Lens*, *Lathyrus*, *Medicago*, *Pisum* e *Vicia*) e a lectina manose/glucose ligante do nódulo de *Arachis hypogaea* (tribo Hedysareae). O subgrupo Vc (Vicieae-Caesalpinoideae) agrupa as lectinas de Caesalpinoideae como *Bauhinia purpurea* e *Griffonia simplicifolia*. Devido as Caesalpinoideae serem consideradas mais primitivas do que as Papilionaceae, esse grupo de lectinas podem ser os ancestrais comuns das lectinas de leguminosas.

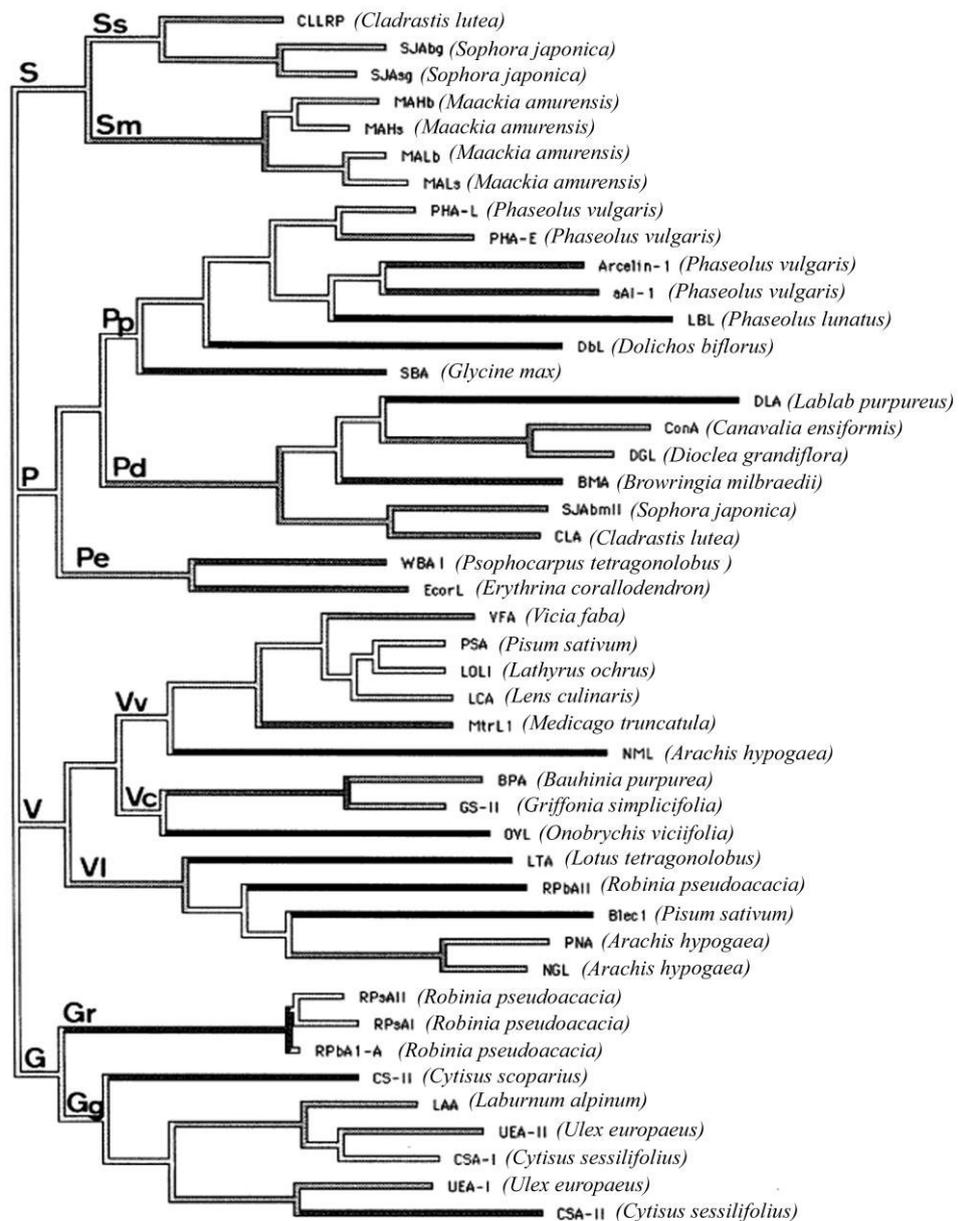


Figura 3 – Árvore filogenética das seqüências que codificam lectinas de leguminosas e proteínas relacionadas. As siglas das lectinas estão mostradas na lista de abreviações. O dendrograma também inclui seqüências de lectinas putativas das sementes de *Medicago truncata* (MtrL1) e bulbo de *Pisum sativum* (Blec1) assim como proteínas relacionadas às lectinas de *Cladrastis lutea* (CLLRP) e *Phaseolus vulgaris* (arcelina-1 e inibidor de alfa-amilase1(aA1)). (Fonte: VAN DAMME *et al*, 1998).

O ramo G (Galegeae/Genisteae) agrupa lectinas de um grupo taxonômico relativamente estreito incluindo as espécies das tribos Galegeae (*Robinia pseudoacacia* e *Laburnum alpinum*) e Genisteae (*Cytisus* e *Ulex*) (VAN DAMME, PEUMANS, *et al.*, 1998).

A ocorrência de lectinas de uma mesma espécie que pertencem a grupos diferentes no dendograma permite construir um perfil da evolução molecular de lectinas de leguminosas. Por exemplo, a expressão de lectinas no ramo S e P de *Sophora japonica* e *Cladrastis lutea* indica que esses dois ramos envolvem um ancestral comum (ramo SP). Similarmente, a ocorrência de lectinas no ramo V e G como ocorrem em *Robinia pseudoacacia* sugere um ancestral comum (ramo VG). Finalmente, como todas as lectinas de leguminosas devem ter um ancestral comum, deve haver um domínio de lectinas de leguminosas (LLD) (VAN DAMME, PEUMANS, *et al.*, 1998).

Recentemente, um receptor quinase putativo tem sido descoberto em *Arabidopsis thaliana* que apresenta a mesma estrutura tridimensional e 39% de similaridade de sequência com lectinas de leguminosas, indicando que o domínio de lectinas de leguminosas pode ocorrer fora da família das leguminosas. Não se sabe sobre a atividade carboidrato-ligante dessa proteína, mas especula-se que ela possa desempenhar um importante papel na transdução de sinais (VAN DAMME, PEUMANS, *et al.*, 1998).

### **2.2.5. A lectina de *Vatairea macrocarpa***

*Vatairea macrocarpa* é uma planta exclusiva dos cerrados, campos cerrados e cerradões, onde apresenta baixa frequência com dispersão irregular e descontínua. Floresce entre os meses de agosto e setembro e seus frutos amadurecem no período de dezembro e janeiro. Produz de maneira irregular e, possivelmente a cada dois anos, moderada quantidade de sementes viáveis (FIGURA 4d) (LORENZI, 2002). Possui uma altura de 5 a 10m, dotada de copa irregular e rala (FIGURA 4a). Seu tronco é cilíndrico e geralmente tortuoso, de 30 a 50 cm de diâmetro, com casca muito grossa, corticosa (FIGURA 4e), e partida em pequenas placas retangulares. Apresenta folhas alternas compostas e imparipinadas. Os folíolos são alternos e opostos, em número de 5 a 7, coriáceos,

discolores, glabros na face superior e denso-tomentosos na inferior, de 5 a 8 cm de comprimento por 4 a 5 cm de largura, sobre pecíolo tomentoso de 4 a 7 mm de comprimento. Possui Inflorescências em panículas terminais amplas, de 14 a 20 cm de comprimento (FIGURA 4b). O fruto é do tipo sâmara alada, glabra e de base esponjosa (FIGURA 4c) (LORENZI, 2002).

A madeira (FIGURA 4f) é indicada para construção civil, como caibros e vigas, para o fabrico de móveis, batentes de portas e janelas, esquadrihas e para lenha e carvão. As flores são apícolas. A casca é reputada como de valor medicinal. A árvore é extremamente ornamental quando em flor, podendo ser empregada com sucesso na arborização paisagística (LORENZI, 2002).

A lectina de sementes de *Vatairea macrocarpa* VML é uma glicoproteína que se apresenta na forma de um tetrâmero formado por quatro subunidades idênticas de 25 kDa cada, independente de pH. Essa lectina induz aglutinação de hemácias de coelhos e humanos e mostra especificidade por galactose, N-acetilgalactosamina e outros açúcares derivados (CAVADA, SANTOS, *et al.*, 1998). Estudos de cromatografia de afinidade de oligossacarídeos e glicoproteínas em coluna imobilizada de VML demonstraram que esta lectina também reconhece sequências de N-acetilgalactosamina não reduzidas (RAMOS, BOMFIM, *et al.*, 2000 *apud* ALENCAR, 2001).

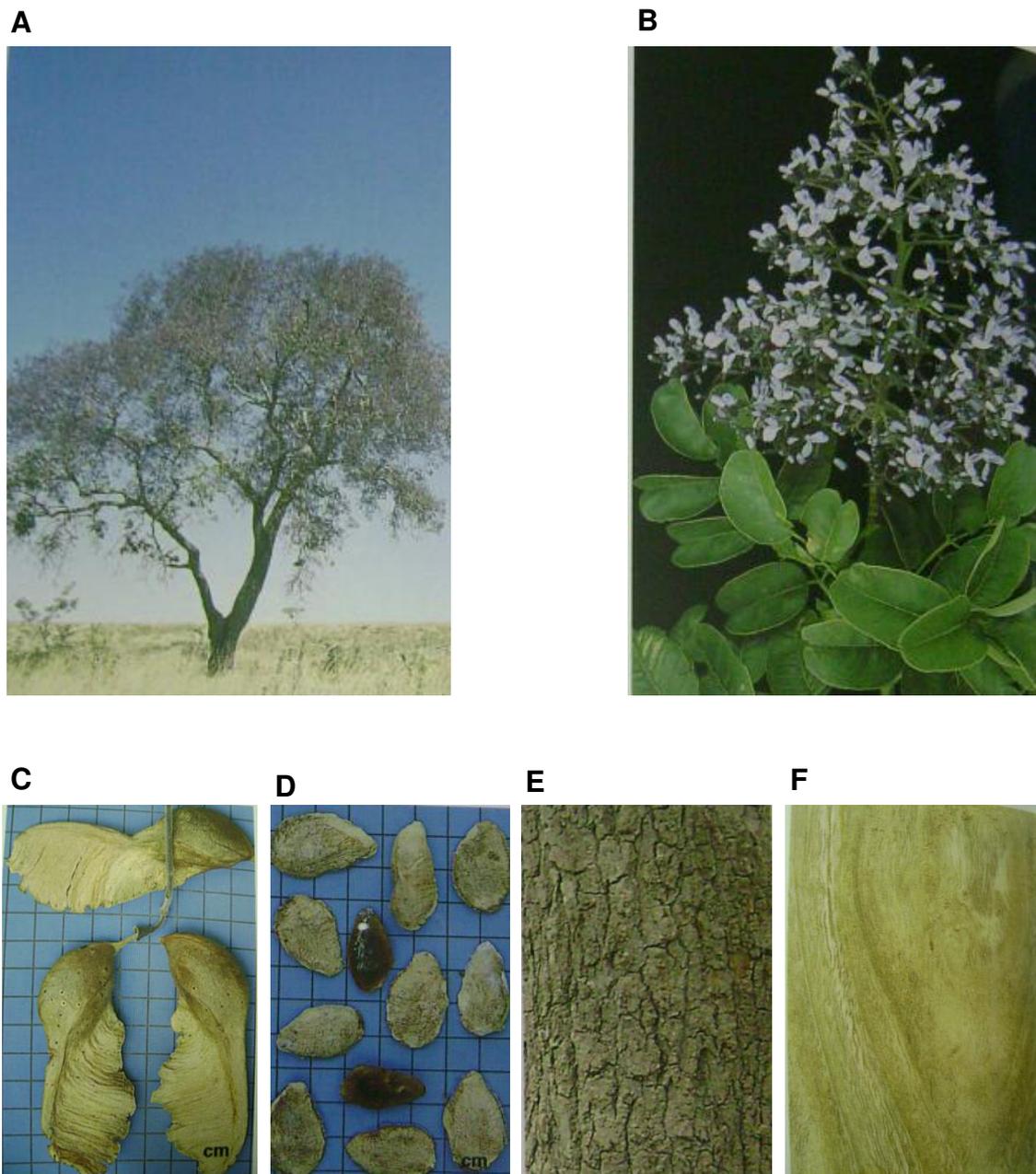


Figura 4 – Características morfológicas de *Vatairea macrocarpa* (Benth.) Ducke. A – Planta inteira com altura de 5-10 m; B – Folhas e inflorescência; C – Fruto sâmara alada; D – Sementes; E – Casca; F – Madeira. Fonte: (LORENZI, 2002).

Estudos de SDS-PAGE mostram que VML é composta de uma banda dupla de 32-34 kDa (cadeia alfa) e duas bandas menores de 22 kDa (cadeia beta) e de 13 kDa (cadeia gama). A lectina de *V. macrocarpa* é N-glicosilada nos resíduos de asparagina nas posições 111 e 183. O sequenciamento N-terminal da proteína revelou que os fragmentos gama e beta correspondem a parte N- e C- terminal, respectivamente, da cadeia alfa (CALVETE, SANTOS, *et al.*, 1998). Além disso, a heterogeneidade nas posições 117 (I/V), 148 (M/K), 154 (G/A), e 168 (E/Q) sugere que a lectina de semente de *V. macrocarpa* pode conter uma mistura de lectinas altamente relacionadas assim como ocorre com as lectinas de *Robinia pseudoacacia* (VAN DAMME, BARRE, *et al.*, 1995) e várias outras lectinas de leguminosas.

O mecanismo proposto para o processamento proteolítico da VML é que ocorre uma clivagem da ponte asparagina-114 e lisina-115 da cadeia alfa depende da deglicosilação na asparagina-111. Isso gera os fragmentos gama (N-terminal deglicosilado) e beta (C-terminal glicosilado). Além disso, o fragmento beta pode sofrer deglicosilação na asparagina-183 seguido de processamento N- e C-terminal gerando isoformas. Tais fragmentos são reunidos por ligações não covalentes fazendo com que ela tenha estrutura tridimensional praticamente idêntica àquela da cadeia alfa (não processada) (CALVETE, SANTOS, *et al.*, 1998).

A lectina de semente de *Vatairea macrocarpa* estimula o fluxo de H<sup>+</sup> em *Rhizobium tropici* com ou sem Galactose. O mesmo não foi observado quando albumina sérica bovina foi testada. A adição de 16 µg de VML por ml resultou em um aumento de 700% da concentração extracelular de Na<sup>+</sup> sugerindo que VML ativa *R. tropici* resultando em alteração no potencial de membrana da bactéria (MARTINEZ, CHANWAY, *et al.*, 2004).

A lectina de *V. macrocarpa* tem um importante efeito nos sítios de ligação a carboidrato do sistema renal (MARTINS, MONTEIRO, *et al.*, 2005). Além disso, várias outras atividades biológicas têm sido encontradas na VML como efeito pró-inflamatório (ALENCAR, ASSREUY, *et al.*, 2004); indução da migração de neutrófilos por mecanismo indireto (ALENCAR, ASSREUY, *et al.*, 2003); ativação de cultura de macrófagos para liberar mediadores quimiotáticos (ALENCAR, ASSREUY, *et al.*, 2007); indução da formação de edema em pata de rato com infiltração de leucócito (ALENCAR, ASSREUY, *et al.*, 2004). Mais recentemente, Dam e colaboradores (2007) mostraram por microcalorimetria que *V. macrocarpa* assim como *Glycine max*

(SBA) se ligam ao antígeno *Tn* da mucina em toda a extensão da cadeia polipeptídica num modelo conhecido como “*bind and jump*” antes da dissociação.

## OBJETIVOS

### Geral

1. Analisar o gene que codifica a lectina de *Vatairea macrocarpa*;
2. Identificar a possível presença de outros genes relacionados à lectina VML;

### Específicos

1. Amplificar o gene que codifica a lectina de *V. macrocarpa* utilizando *primers* desenhados a partir de sequências conservadas de lectinas obtidas pelo *GenBank*;
2. Clonar os fragmentos gênicos de *Vml* obtidos por PCR em células de *E. coli* tipo JM109;
3. Sequenciar os clones obtidos e utilizar a sequência para o desenho de novos *primers* que serão utilizados para a amplificação das extremidades 5' e 3';
4. Amplificar as extremidades 5' e 3' do gene *Vml* utilizando a técnica de RACE;

## 5. MATERIAIS E MÉTODOS

### 4.1. Isolamento de DNA genômico

Para a obtenção do DNA genômico de *V. macrocarpa*, folhas jovens foram coletadas de uma planta localizada próximo ao Parque de Desenvolvimento Tecnológico do Ceará - PADETEC, no Campus do Pici, Fortaleza-CE. O método utilizado para purificação de DNA é uma variação da metodologia originalmente descrita por (MURRAY e THOMPSON, 1980) usando-se o reagente CTAB, conforme descrito a seguir.

As folhas coletadas foram lavadas com águas destilada e o limbo das folhas separadas das nervuras. Após esse procedimento, pesou-se 0,4 g do limbo das folhas. O material biológico obtido foi macerado em pistilo e almofariz estéril, em presença de nitrogênio líquido. O pó resultante foi misturado a 4 ml do tampão de extração (Tris-HCl 100 mM; pH 8,0; NaCl 1,4 M EDTA 20 mM; CTAB 2% e 2-mercaptoetanol 0,2%) pré-aquecido a 60°C, em tubo polipropileno (Tipo Falcon®) estéril. A amostra foi incubada a 60 °C por 16 horas (“*overnight*”) e então extraída com 1 volume de clorofórmio: álcool isoamílico, (25:1, v/v). O tubo foi invertido várias vezes e então centrifugado a 2.500 x g em centrífuga Centribio Mod. 80-2B por 15 minutos a temperatura ambiente.

Os ácidos nucleicos presentes na fase superior aquosa foram transferidos para um tubo Falcon limpo e precipitado pela adição de 2/3 do volume de isopropanol 100%. Ao DNA precipitado foi adicionado 2 ml de NaCl 1 M e precipitado novamente pela adição de 2,5 volumes de etanol 100% por 16 horas a 4 °C.

O precipitado foi lavado com etanol 70% (1 ml por 10 minutos) ressuspendido em 2 ml de TE (Tris-HCl 10 mM, EDTA 1 mM, pH 8,0) e armazenado a uma temperatura de 4 °C, para uso posterior.

A concentração do DNA das várias amostras foi estimada a partir dos valores de absorbância a 260 nm (A260) usando-se espectrofotômetro *GeneQuant pro* (GE Healthcare). A razão A260/A280 foi usada como parâmetro para se estimar a pureza de cada preparação (SAMBROOK, FRITSCH e MANIATIS, 1989).

#### 4.2. Eletroforese em gel de agarose 0,8% (m/v), pH 8,0

A eletroforese de DNA em gel de agarose foi realizada segundo o protocolo descrito por (SAMBROOK, FRITSCH e MANIATIS, 1989). Resumidamente, a quantidade de agarose 0,4 g foi adicionada a 50 ml de tampão TBE (Tris-Ácido Bórico 0,045 M, EDTA 0,001 M). Após ser completamente fundida, a mistura foi resfriada a 60 °C e o brometo de etídio adicionado, a partir de uma solução estoque de 10 mg/ml, em água (Amersham, USB), de maneira que sua concentração final foi de 0,25 mg/ml. O gel permaneceu em repouso a temperatura ambiente até a sua completa solidificação.

A amostra a ser submetida à corrida eletroforética foi previamente misturada, 20% do seu volume da amostra, com uma solução de azul de bromofenol 0,25% (m/v), glicerol 30% (v/v), preparada em tampão TE. O gel, contendo brometo de etídio, foi submerso em tampão TBE na mesma concentração do gel, em cuba Hoefer HE (mini horizontal submarine unit) Amersham Biosciences™.

A amostra foi, então, aplicada ao gel e a corrida eletroforética foi realizada a uma voltagem constante de 100V, fornecida pela fonte *Electrophoresis Power Supply* EPS 301, (Amersham Biosciences™). A voltagem foi desligada, encerrando-se a corrida quando o azul de bromofenol visualizado percorresse mais de 2/3 do comprimento do gel. O gel foi levado para um sistema de fotodocumentação em gel L-Pix contendo transluminador (*transluminator*) e uma câmera *Canon A620* ligado diretamente a um computador onde as fotodocumentações ficam armazenadas.

#### 4.3. Amplificação da região central do gene da lectina de *V. macrocarpa* (VML) por PCR

Para a amplificação do gene precursor de *V. macrocarpa* foram desenhados *primers* semi-degenerados a partir de sequências de aminoácidos conservados de lectinas leguminosas (ANEXO 2 e 3). O *primer* direto (*forward*) foi desenhado a partir da sequência **PI(H/Q)IWE** (CCN GTS CAV ATT TGG GA) e o *primer* reverso (*reverse*) a partir da sequência **PEWVRV** (WAC YCT HAC CCA YTG NGG). O gene que codifica a proteína VML aqui chamado *Vml* (o qual é desprovido de íntrons) deve conter 720pb. Contudo, o fragmento a ser amplificado com os

*primers* citados deve possuir cerca de 430 pb e se localiza na região interna do gene *Vml* (FIGURA 5). Também se tentou amplificar a partir da região mais externa do gene da *Vml* utilizando os *primers forward* na sequência **KFNPNP** (VaiF1 e VaiF2) e *reverse* na sequência **QAPSDD** (VaiR1 e VaiR2). A sequência dos *primers* contendo outras informações é mostrada na TABELA 1.

Para a reação de amplificação através da reação em cadeia da polimerase (PCR) foi utilizado (em uma reação de 25 µl): 300 ng de DNA genômico, 0,2 µM de *primers* semidegenerados, 0,2 mM de dNTPs e 1 unidade de *Taq* DNA polimerase *Platinum* da Invitrogen™ (acompanhando tampão 10X e MgCl<sub>2</sub> conforme instruções do fabricante), quando necessário o volume foi completado com água Tipo 1 estéril (água Mili-Q). Além da reação da amostra, foi realizada uma reação adicional contendo todos os reagentes, à exceção de DNA genômico (controle negativo).

As condições térmicas de amplificação foram feitas utilizando uma variante da PCR tradicional conhecida como PCR “*touchdown*” que consiste em uma amplificação com uma temperatura de anelamento inicial alta que reduz gradativamente em cada ciclo (ROUX, 2002). Tal estratégia visa o aumento do rendimento dos amplicons sem perda da especificidade. Sendo assim, as seguintes condições térmicas foram utilizadas: desnaturação do DNA a 95 °C durante 3 minutos, seguido de 10 ciclos de amplificação (95 °C durante 1 minuto, 50 °C durante 0,5 minutos e 72 °C durante 2 minutos). Nesta etapa houve uma redução de 0,5 °C por ciclo sendo que o primeiro ciclo tinha uma temperatura de anelamento de 50 °C e após os 10 ciclos passa a ter 45 °C. Após essa redução, houve uma amplificação de 32 ciclos de (95 °C por 1 minuto, 45 °C por 1 minuto e 72 °C por 2 minutos), seguido por extensão final a 72 °C durante 8 minutos (TABELA 2). As reações de amplificação foram feitas em termociclador MasterCycler Gradient (Eppendorf). Após a reação de PCR, Os produtos foram analisados em gel de agarose 1%.

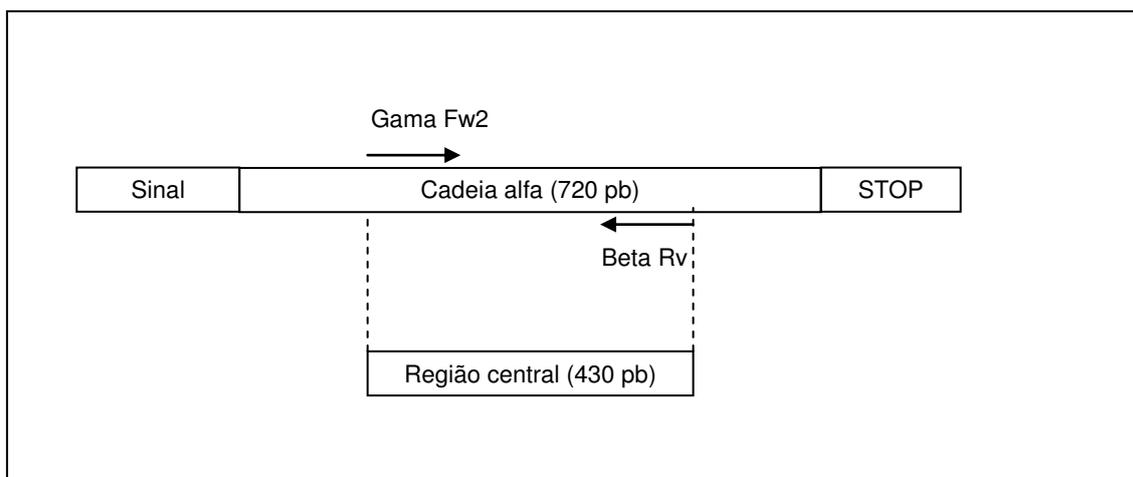


Figura 5 – Estratégia para a construção de *primers* semi-degenerados para o gene *Vml* de *Vatairea macrocarpa*. Sinal – Região do gene que codifica o peptídeo sinal; STOP – Região onde fica o primeiro códon de terminação.

Tabela 1 – *Primers* utilizados na amplificação da região central do gene *Vml*

<b>Primers</b>	<b>Sequência (5' para 3')</b>	<b>Tm</b>	<b>GC</b>	<b>Nº Deg.</b>
Gama Forward II	CCN GTS CAV ATT TGG GA	50,0°C	35,9%	16
Beta Reverse	WAC YCT HAC CCA YTG NGG	52,1°C	44,0%	96
VaiF1	AAR TTY AAY CCY AAY CC	44,7°C	38,2%	24
VaiF2	AAR TTY AAY CCY AAI CC	46,2°C	35,6%	12
VaiR1	RTC RTC NGA RGG NGC YTG	56,5°C	61,1%	256
VaiR2	RTC RTC NGA RGG NGC ITG	57,7°C	58,3%	128

Tabela 2 – Condições térmicas da amplificação da região central da *Vml* por PCR *touchdown*.

	<b>Temperatura</b>	<b>Tempo</b>
	95°C	3 min
10X	95°C	1 min
(-0,5°C/ciclo)	50°C	30 seg
(3,0°C/s)	72°C	2 min
	95°C	1 min
32X	45°C	1 min
	72°C	2 min
	72°C	8 min

#### **4.4. Clonagem do fragmento gênico de *Vml* amplificado por PCR**

##### 4.4.1. Ligação do produto de PCR ao vetor

O fragmento de DNA produto da PCR foi diretamente ligado no plasmídeo PGEM-T *easy vector*® (Promega) conforme as instruções do fabricante (FIGURA 6). A reação foi incubada durante 1 hora a 25°C e em seguida ficou a 4°C por 12 horas.

##### 4.4.2. Preparo de células eletrocompetentes

As células de *E.coli* (cepa JM109) conservadas em glicerol a -20°C foram crescidas em meio LB Broth (20 g/L) *overnight* a 200 rpm em agitador orbital. Seu crescimento foi acompanhado pela quantificação da turbidez bacteriana, medido em espectrofotômetro com OD de 600 nm. Foi deixado atingir a fase *Log* de crescimento onde a absorbância deveria estar por volta de 0,4 até 0,6.

As bactérias crescidas em LB Broth foram transferidas para tubos tipo falcon e foram concentradas por centrifugação a 4000 x g por 10 minutos a 4°C. O sobrenadante foi descartado e ao precipitado foram adicionados 50 ml de glicerol a 10% até ressuspender. Em seguida, os tubos foram centrifugados a 4000 x g por 10 minutos e o sobrenadante foi descartado. Em seguida, foi adicionado 50 ml de glicerol por alguns minutos. Em seguida os tubos foram novamente centrifugados a 4000 x g por 10 minutos a 4°C. O sobrenadante foi descartado e por fim foi adicionado ao sedimento de células 4 ml de glicerol 10%. As células eletrocompetentes foram armazenadas em freezer -80°C.

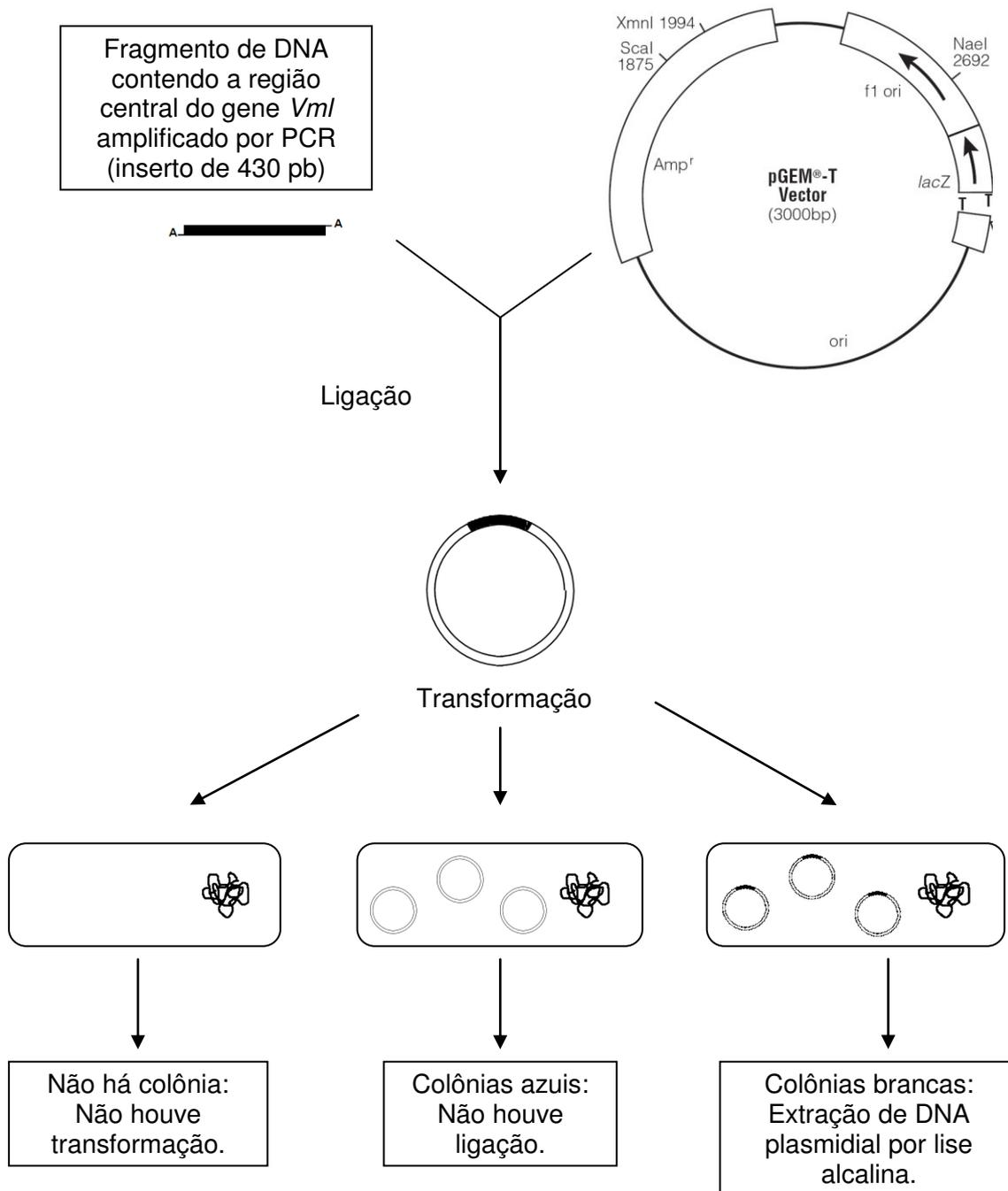


Figura 6 – Estratégia de clonagem dos fragmentos de DNA correspondente à região central do gene *Vml* obtidos por PCR.

### 4.4.3. Transformação

A transformação por eletroporação foi realizada em um eletroporador Multiporator-Eppendorf. Inicialmente, foi adicionado 100 µl de células eletrocompetentes juntamente com 1 µl do produto da reação de ligação nas cubetas de 2 mm para eletroporação. Em seguida, a cubeta foi submetida a um pulso elétrico de 1700 V em 5 milissegundos. As células transformadas foram transferidas para um novo tubo de polipropileno de 1,5 ml onde foi adicionado mais 200 µl de meio LB Broth. A mistura foi incubada a 37°C por 30 minutos para que houvesse uma recuperação das células tratadas com choque elétrico.

Em uma placa de Petri contendo meio LB-agar com ampicilina (100 µg/ml), X-Gal (32 µg/ml) e IPTG (80 µM). Em seguida, foram plaqueadas 100 µl de células transformadas e as placas foram então incubadas a 37°C por 12 horas. Após esse período espera-se observar o surgimento de colônias azuis (sem inserto) e colônias brancas (com inserto) (FIGURA 6).

## 4.5. Seleção de clones recombinantes para sequenciamento

### 4.5.1. PCR de colônia

Uma reação de PCR de colônia utilizando os primers para *Vml* foi feita para avaliar a eficiência e triagem da transformação. Adicionalmente, a reação pode ser feita com *primers* que se anelem no vetor, como os *primers* para os sítios M13 (M13F: TGT AAA ACG ACG GCC AGT e M13R: CAG GAA ACA GGT ATG ACC) existentes no pGEM-T. A reação foi feita utilizando os mesmos componentes de uma reação de PCR convencional exceto pelo DNA que foi obtido diretamente da colônia com auxílio de alça de platina.

### 4.5.2. Extração de DNA plasmidial (*miniprep*)

As colônias transformadas incubadas em meio LB-Agar foram crescidas em 100 ml de meio LB-Broth sob agitação orbital constante a 37 °C para posterior extração de DNA plasmidial.

As bactérias foram concentradas por centrifugação (a 2.000 x g por 5 minutos a temperatura ambiente em centrífuga Centribio MOD 80-2B). O sobrenadante foi descartado com auxílio de pipeta deixando o sedimento em apenas 2 ml de meio restante no tubo. Os 2 ml restantes (contendo o sedimento) foram então misturados com auxílio de pipeta e transferidos para tubos de polipropileno com capacidade para 2ml e centrifugados a 6.000 x g por 10 minutos a 4°C em Micro Centrífuga Refrigerada 14000DR. O sobrenadante foi descartado e ao sedimento foi adicionado 400 µl da solução I (glicose 20%; 0,5M EDTA pH 8,0; 1M Tris-HCl pH 7,4). Em seguida, foi adicionado 800 µl da solução II (NaOH 0,2 N; SDS 1%). Os tubos foram misturados por inversão. Posteriormente, foi adicionado 600 µl da solução III (60% KOAc 5M; 11,5% ácido acético glacial) sendo novamente misturada por inversão.

O lisado foi centrifugado a 10.000 x g (em Micro Centrífuga Refrigerada 14000DR) por 10 minutos a 4°C. O sobrenadante foi transferido para um novo tubo onde foram adicionados 1,2 ml de isopropanol. Os tubos foram misturados por inversão e mantidos a temperatura ambiente por 10 minutos. Em seguida, as amostras foram centrifugadas a 12.000 x g (em Micro Centrífuga Refrigerada 14000DR) por 10 minutos a 4°C, o sobrenadante foi descartado. Ao precipitado foi adicionado 400 µl de TE (Tris 10 mM; EDTA 1 mM pH 8,0) contendo RNase (20 µg/ml) e incubado por 30 minutos a 37°C.

Após esse tempo, foi adicionado 0,15 volume de acetato de sódio 2M e 3 volumes de etanol 100%. As amostras foram então centrifugadas a 14.000 x g (em Micro Centrífuga Refrigerada 14000DR) por 5 minutos a 4°C. O sobrenadante foi descartado e o precipitado foi lavado com 100ul de etanol 70%. Os tubos foram mantidos abertos até completa evaporação do etanol. Finalmente, o precipitado foi dissolvido em 100 µl de TE e armazenado a -20 °C para uso posterior. Os plasmídios foram quantificados por eletroforese em gel de agarose a 0,8% e por espectrofotometria antes de serem submetidos à reação de sequenciamento.

#### 4.6. Isolamento de RNA total

O RNA total foi obtido pelo *Plant RNA Reagent* (Invitrogen™). Brevemente, folhas jovens de *V. macrocarpa* foram coletadas e imediatamente congeladas em nitrogênio líquido. As amostras foram então maceradas e incubadas com 1 ml de tampão de extração. Os tubos foram levemente invertidos e incubados a temperatura ambiente por 15 minutos. A solução foi centrifugada por 2 minutos a 12.000 x *g* a temperatura ambiente e o sobrenadante foi transferido para um novo tubo. Os tubos foram misturados por inversão após adição de 0,1 ml de NaCl 5M e 0,3 ml de clorofórmio. As amostras foram centrifugadas a 10.000 x *g* por 10 minutos a 4°C para separar as fases. A fase superior aquosa foi transferida para um novo tubo livre de RNase seguido pela adição de um volume igual de isopropanol. As amostras foram misturadas por inversão e deixadas a temperatura ambiente por 10 minutos.

Os tubos foram centrifugados a 12.000 x *g* por 10 minutos. O sobrenadante foi descartado e ao precipitado foi adicionado 1 ml de etanol 75%. Os tubos foram centrifugados a temperatura ambiente por 1 minuto a 12.000 x *g*. O líquido foi descartado e ao precipitado foi adicionado 30  $\mu$ l de água livre de RNase misturando levemente com pipeta e deixando em repouso até o precipitado se dissolver por completo. As amostras foram armazenadas em freezer -80°C para uso posterior.

#### 4.7. Síntese da primeira fita de cDNA

A cauda poli A do RNA mensageiro foi usada como molde para o anelamento do adaptador Qt que tem a seguinte sequência de 5' para 3' (CCA GTG AGC AGA GTG ACG AGG ACT CGA GCT CAA GCT TTT TTT TTT TTT TTT T). A reação de síntese de cDNA foi feita utilizando-se 3  $\mu$ g de mRNA, 1  $\mu$ g de Qt, 0,5 mM de dTNPs e 500 U da enzima M-MLV RT (*Moloney-Murine Leukemia Virus Reverse Transcriptase*, USB) contendo tampão 5X (250 mM Tris-HCl, pH 8,3, 395 mM KCl, 15 mM MgCl<sub>2</sub>, 50 mM DTT) em uma reação com um volume final de 50  $\mu$ l.

Antes da reação, o RNA tinha sido incubado a 80°C por 3 minutos para desnaturação de possível estrutura secundária. Em seguida, as amostras foram

incubadas a 37°C por 30 minutos e em seguida por 75°C por 10 minutos para desnaturação da enzima. O material ficou armazenado a -20°C para uso posterior.

#### 4.8. Síntese da segunda fita de cDNA por amplificação das extremidades 5' e 3'

As extremidades não amplificadas da *Vml* podem ser obtidas pela técnica de RACE (*Rapid Amplification of cDNA ends*) (FROHMAN e MARTIN, 1989). O 3' RACE consiste na adição de uma sequência âncora à extremidade 3' do mRNA. Posteriormente, um *primer* universal (juntamente com um *primer* específico à região já sequenciada) é utilizado para amplificar a região desconhecida.

A estratégia utilizada para amplificar a região 5' foi a utilização de um *primer* reverso desenhado na região interna já sequenciada e um *primer* direto com a sequência **LNKVNS** (GCT ACT CAA CAA GGT GAA YTC) desenhado para o peptídeo sinal em uma região bastante conservada em lectinas de leguminosas (FIGURA 8). A região 3' foi amplificada utilizando os *primers* Q0 ou Qi (específico para o adaptador Qt) e *primers* desenhados na região interna do gene *Vml* chamados 3RACEVml e 3RACEVml2 conforme mostra a Figura 7. Os *primers* utilizados para o RACE estão listados nas tabelas 3.

A reação foi feita em um volume final de 25 µl contendo 300 ng de DNA, 0,4 µM de *primers*, 0,2 µM de dNTPs e 1 unidade de *Taq* DNA polimerase *Platinum* (Invitrogen™) contendo tampão 10 X (200 mM Tris-HCl, pH 8,4, 0,5 M KCl) e MgCl<sub>2</sub> conforme instruções do fabricante.

As reações de amplificação foram feitas em termociclador *MasterCycler Gradient* (Eppendorf). As condições térmicas estão mostradas na tabela 4. Após a reação de PCR, foi feita uma eletroforese em gel de agarose a 1%.

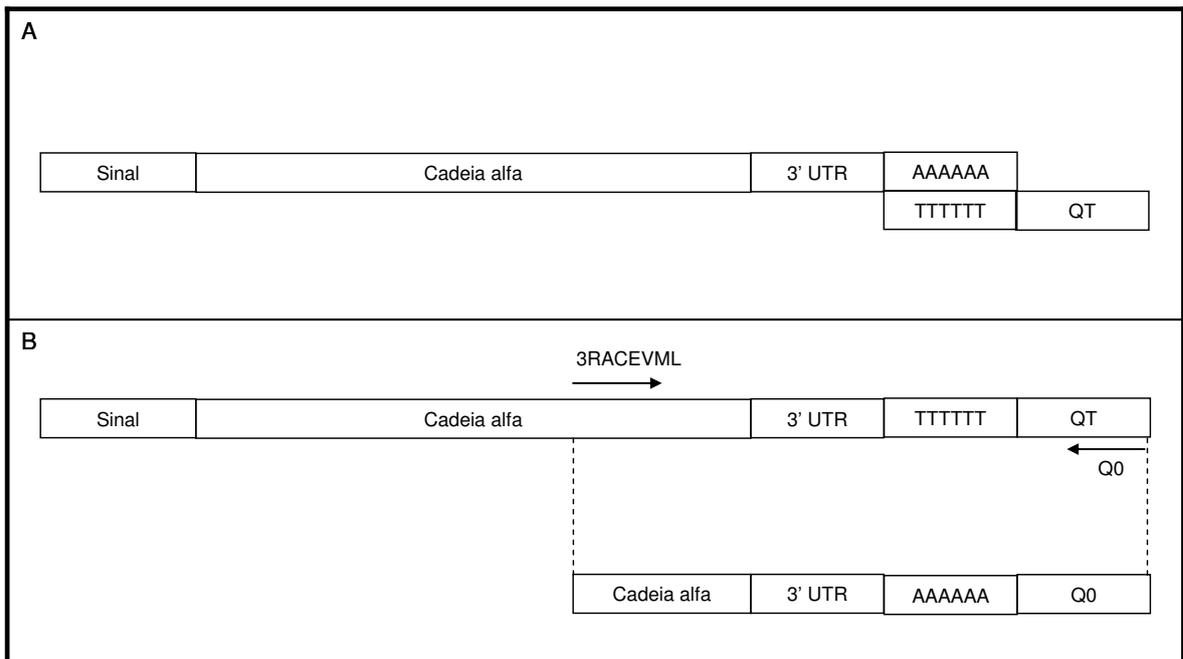


Figura 7 – Estratégia utilizada para a amplificação rápida das extremidades de cDNA (RACE) do gene da lectina de *V. macrocarpa*.

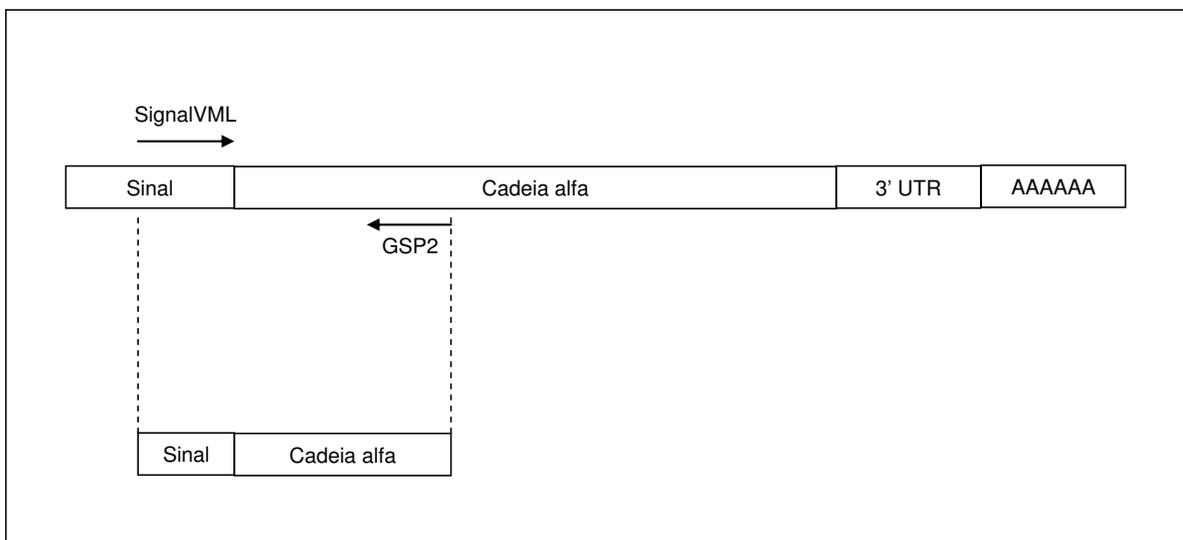


Figura 8 – Estratégia utilizada para a amplificação da extremidade 5' do gene que codifica a lectina VML.

Tabela 3 – *Primers* utilizados na amplificação da extremidade 5' e 3' do gene *Vml*

<b>Primers</b>	<b>Sequência (5'para 3')</b>	<b>Tm</b>	<b>GC %</b>
Signal_vml (F)	GCT ACT CAA CAA GGT GAA YTC	54	43
3race_vml (F)	GGG TTG GGA GAA TGG GGA AGT AGC	69	58
3race_vml2 (F)	GCA CCT CCA GAT ACT CAA CC	57	55
Gsp2vml (R)	GGA GGT GCA AGA AAG AAG GC	61	55
Qi (R)	GAG GAC TCG AGC TCA AGC	55,2	61,1
Q0 (R)	CCA GTG AGC AGA GTG ACG	55,4	61,1

Tabela 4 – Condições térmicas da amplificação das extremidades 5' e 3' do gene *Vml*

	<b>Temperatura</b>	<b>Tempo</b>
	95°C	5 min
30X	95°C	1 min
	58°C	30 seg
	72°C	1 min
	72°C	8 min

## 4.9. Sequenciamento de DNA

### 4.9.1. Reação de sequenciamento

Na reação de incorporação dos dideoxinucleotídeos marcados foi utilizado 4 µl de premix (contendo *Taq* DNA polimerase, tampão, dNTPs e ddNTPs marcados) fornecido pelo kit *DYEnamic™ ET dye terminator kit* (*MegaBace™ Amersham Biosciences*). Além disso, foi adicionado 100 ng de DNA plasmidial (ou 300 ng de produto de PCR) e 5 pmoles de *primer* para um volume final de 10 µl por reação conforme indicado pelo fabricante.

A reação foi feita em placa com 96 poços e submetida posteriormente ao programa do termociclador com 25 ciclos cujas temperaturas são: desnaturação (95°C por 20 segundos); anelamento (50°C por 15 segundos) e extensão (60°C por 1 minuto).

### 4.9.2. Precipitação por etanol

Inicialmente, foi adicionado 1 µl de acetato de amônio (7,5M) em cada poço seguida pela adição de 27,5 µl de etanol 100%, a placa foi selada com adesivo resistente a álcool e misturada por inversão várias vezes. A placa foi centrifugada a 2.500g por 30 minutos (ou 2.000 g por 1 hora). O sobrenadante foi removido por inversão e a placa foi invertida e centrifugada rapidamente (*spin*) para a completa remoção do álcool. Os precipitados foram lavados com etanol 70% (100 µl/poço). O sobrenadante foi novamente removido por inversão e a placa foi invertida e centrifugada rapidamente (*spin*) para a completa remoção do álcool.

Para a ressuspensão das amostras foi adicionado a cada poço 10 µl de *MegaBace loading solution* do kit *DYEnamic ET terminator kit* (*MegaBace™*). As amostras foram misturadas por vortex durante 20 segundos e centrifugada rapidamente (*spin*).

#### 4.9.3. Análise das sequências no MegaBace 750

Os fragmentos marcados foram submetidos a uma corrida eletroforética utilizando o Analisador genético MegaBace750 com auxílio do programa ICM (*Instrument Control Manager*). O equipamento requer em cada corrida a injeção de matriz MegaBace™ *Long Read Matrix* (GE Healthcare) e uma placa contendo 100 µl de tampão 10x MegaBace™ LPA Buffer diluído 10 vezes. O procedimento de injeção de matriz foi feito conforme instruções do fabricante.

Para análise dos fragmentos obtidos diretamente por reação de PCR, os parâmetros eletroforéticos utilizados foram 3 kv por 60 segundos de injeção e 9 kv por 120 minutos. Aos fragmentos obtidos por miniprep, os parâmetros foram 2 kv por 60 segundos de injeção e 9 kv por 120 minutos de corrida, todos a uma temperatura constante de 44°C.

O eletroferograma obtido foi armazenado no computador e pôde ser interpretado e editado utilizando o programa MegaBace *Sequence Analyser*™. Que pode excluir sequências consideradas de baixa qualidade (causadas por sobreposição de picos no eletroferograma ou picos de baixa intensidade de sinal). Possíveis erros da *Taq* DNA polimerase na reação de PCR assim como regiões de baixa qualidade de sequenciamento também podem ser resolvidas em programas de alinhamento global entre as duplicatas obtidas no sequenciamento.

#### 4.10. Montagem das Sequências Obtidas

Os dados brutos provenientes do sequenciamento de todas as amostras foram submetidos a um pacote de programas para montagem e obtenção de sequências consenso. O pacote de programas utilizado para a leitura dos dados brutos e montagens das sequências consenso foi o do **Phred** (versão 0.020425) (EWING e GREEN, 1998) e do **Phrap** (versão 0.990329) PHIL GREEN, *University of Washington*). O programa Phred lê todos os dados brutos do sequenciamento utilizando análise de Fourier, e calcula a probabilidade de erro para cada *base call*. Os escores do Phred são na forma de logaritmo negativo da probabilidade de que aquele *base call* é um erro. Portanto, quanto maior for o escore do Phred, menor é a probabilidade que um erro foi cometido (BROWN, 2000); (GIBAS e JAMBECK, 2001). O programa Phrap, então utiliza essas probabilidades de erro para determinar sequências consenso altamente precisas, pela utilização da sequência de mais alta qualidade em cada posição para construir o consenso (BROWN, 2000). As sequências consenso montadas pelo programa Phrap, foram então visualizadas e editadas pelo programa de interface gráfica Consed (versão 16.0); (GORDON, ABAJIAN e GREEN, 1998).

As sequências consenso obtidas das amostras foram traduzidas em *Translate tool* (Expasy) e comparados com as sequências do GenBank NCBI (*National Center for Biotechnology Information*) através da ferramenta BLAST (*Basic Local Alignment Search Tool*) (ALTSCHUL, GISH, *et al.*, 1990). O BLASTn foi utilizado para a comparação de sequências de nucleotídeos com o banco de dados. A comparação entre as sequências de DNA traduzidas com as sequências de proteínas disponíveis foi feita utilizando a ferramenta BLASTp e BLASTx.

As análises filogenéticas foram feitas utilizando o programa **MEGA** (*Molecular Evolutionary Genetics Analysis*) versão 4 (TAMURA, DUDLEY, *et al.*, 2007). O método para construção de árvores utilizado foi o do vizinho mais próximo (Neighbor-Joining NJ) e o modelo matemático de Tamura-Nei.

## 5. RESULTADOS

### 5.1. Isolamento de DNA genômico

A análise do DNA genômico no espectrofotômetro mostrou uma concentração de 105 ng/ $\mu$ l e relação  $A_{260}/A_{280}$  de 1,5. A eletroforese em gel de agarose mostrou uma banda uniforme com tamanho de aproximadamente 23.000 pb indicando boa qualidade na extração (FIGURA 9).

### 5.2. Amplificação da região interna de *Vml* por PCR

Os *primers* semi degenerados (gama Forward II e Beta Reverse) utilizados para a amplificação da região interna do gene *Vml*, a partir de DNA genômico, se mostraram eficientes na amplificação do gene *Vml* utilizando touchdown-PCR. A eletroforese revelou a presença de uma banda de aproximadamente 430 pb, conforme o esperado. Nenhuma banda foi vista no controle negativo da reação. Além disso, os *primers* degenerados VaiF1, VaiF2, VaiR1 e VAIr2 não foram eficientes na amplificação da lectina de *V. macrocarpa* (FIGURA 10).

### 5.3. Clonagem do fragmento gênico de *Vml* amplificado por PCR

A transformação de bactérias com o pGEMT-*Vml* resultou na obtenção de colônias brancas (contendo o plasmídeo ligado ao inserto de DNA) e azuis (contendo apenas o plasmídeo circular). A presença de colônias azuis mostra que a reação de ligação não garante que todos os plasmídeos serão ligados ao inserto de DNA e servem como controle negativo mostrando a eficiência do X-Gal e IPTG.

Então foram realizadas PCR's de colônia com todas as colônias obtidas. O resultado foi visto em uma corrida eletroforética revelando bandas de aproximadamente 500 pb conforme o esperado (FIGURA 11).

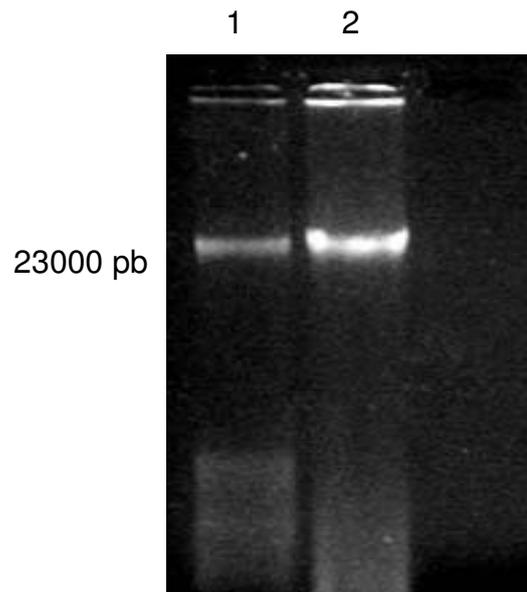


Figura 9 - Eletroforese em gel de agarose a 0,8% de DNA genômico de *Vatairea macrocarpa*. Linha 1: Amostra utilizada como referência do tamanho aproximado de 23000pb previamente marcada com marcador lambda. Linha 2: DNA genômico de *V. macrocarpa*.

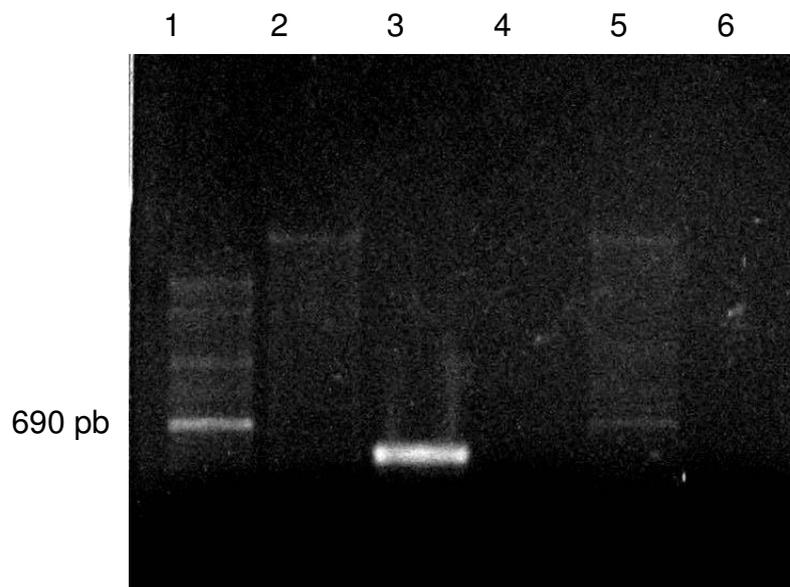


Figura 10 – Eletroforese em gel de agarose a 1% corada com brometo de etídio do gene da lectina da *V. macrocarpa* amplificado por “PCR touchdown” com temperatura de anelamento de 50 a 45°C. Linha 1: *primers* VaiF1-VaiR1; Linha 2: *primers* VaiF2-VaiR2; Linha 3: *primers* Gama Forward II-Beta Reverse, Linha 4: *primers* VaiF2- Beta Reverse; Linha 5: *primers* Gama Forward II -VaiR2; Linha 6: controle negativo.

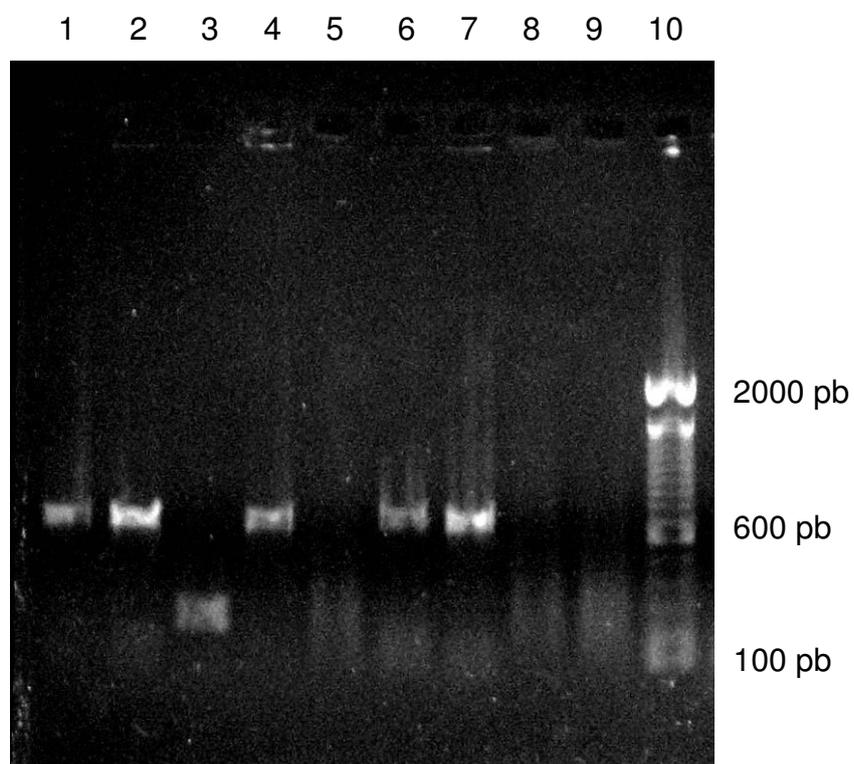


Figura 11 – Eletroforese em gel de agarose 1% da reação de PCR das colônias contendo a região central do gene *Vml*. Linhas de 1 a 9: pGemT-*Vml* amplificados com os *primers* M13F e M13R; As linhas 8 e 9 representam o controle negativo das respectivas amostras. Na linha 10 está o marcador de peso molecular *Ladder* 100 pb.

#### 5.4. Extração de DNA de plasmídios recombinantes por lise alcalina

Após a triagem dos clones recombinantes, procedeu-se a extração de DNA plasmidial utilizando o método da lise alcalina. A quantificação dos plasmídios mostrou concentrações que variaram de 348 a 1740 ng/μl (TABELA 5). O resultado revelado pela eletroforese em gel de agarose a 1% é mostrado na Figura 12.

#### 5.5. Sequenciamento da região central do gene *Vml*

As condições de reação e purificação bem como os parâmetros de injeção e corrida no analisador genético se mostraram eficientes na obtenção de sequências de qualidade. As análises posteriores e desenho dos *primers* foram realizados a partir da sequência consenso entre as quadruplicatas obtidas, referentes a cada amostra sequenciada. Vale salientar que foram consideradas apenas regiões com valor de qualidade Phred maior que 20 e que o programa Phrap gerou três sequências gênicas distintas, denominados de *contig1*, *contig2* e *contig3*.

#### 5.6. Alinhamento das sequências

As sequências consensos dos 3 *contigs* obtidos foram submetidas à tradução utilizando o *Translate Tool (Expasy)*, que realiza a tradução nos diferentes *frames* e nas duas direções para obter a sequência de aminoácidos a partir de uma sequência de DNA. As seis sequências de proteínas resultantes de cada *contig* foram comparadas com as sequências de proteínas depositadas no *GenBank* do *National Center of Biotechnology Information* (<http://www.ncbi.nlm.nih.gov/>) através do *Basic Local Alignment Search Tool - BLAST* (ALTSCHUL, GISH, *et al.*, 1990). Essa comparação permitiu a identificação da verdadeira janela de leitura e demonstrou as semelhanças entre a estrutura primária dos três *contigs* obtidos com a lectina VML determinada por (CALVETE, SANTOS, *et al.*, 1998).

As sequências de aminoácidos traduzidas a partir das três sequências consenso obtidas quando alinhadas com a lectina VML mostram que as sequências, apesar de homólogas, codificam proteínas diferentes.

A Figura 13 mostra o alinhamento das três seqüências consenso em pauta de leitura correto juntamente com as sequências traduzidas, mostradas abaixo das sequências de DNA. O alinhamento foi feito incluindo a VML cuja sequênciã foi editada para mostrar somente a região central que vai dos resíduos 55 a 210.

O *contig1* foi o que se apresentou mais próximo de sequênciã da lectina VML com 99% de similaridade de sequênciã havendo substituição em apenas quatro resíduos de aminoácidos. Os resíduos da VML substituídos foram: histidina por glutamina posição 58; aspartato por asparagina na posição 94; lisina por glutamato na posição 160 e glutamato por glutamina na posição 169 (FIGURA 13). Tal sequênciã pode ser considerada o gene que efetivamente codifica VML desde que se admita a possibilidade de haver ambigüidades ocorridas na degradação de Edman que gerou a sequênciã de aminoácidos de VML.

Os *contigs* 2 e 3 apresentam maior quantidade de resíduos de aminoácidos do que a sequênciã da VML, contendo 2 e 5 resíduos adicionais respectivamente. O *contig3* é o que apresenta maior tamanho contendo 5 inserções de aminoácidos na região entre os resíduos 113 e 114 da VML (FIGURA 13). Além disso, os *contigs* 2 e 3 mostram 77% de identidade de sequênciã quando comparados com VML. O *Contig3* contém um resíduo de cisteína na posição 103.

Os resíduos de asparagina no sítio de glicosilação da VML foram substituídos nos *contigs* 2 e 3. A asparagina 111 foi substituída por prolina e arginina nos *contigs* 2 e 3, respectivamente enquanto que a asparagina 183 foi substituída por serina.

Segundo (CALVETE, SANTOS, *et al.*, 1998) os sítios de ligação a carboidratos da VML são: glicina 88, fenilalanina 127, asparagina 129 e aspartato 132, ao nível da região central. Os *contigs* 2 e 3 mostram conservação do aspartato 132 e da glicina 88 (mesmo havendo mudança de códon).

Também foi observada a presença de 18 resíduos de aminoácidos conservados, mas resultante de códons diferentes. Esse dado enfatiza a possível importância desses resíduos na manutenção da estrutura tridimensional da proteína (FIGURA 13) (ANEXO 1). As três sequências consenso mostraram também diferentes resíduos de aminoácidos nas posições 117, 148, 154 e 168 podendo confirmar casos de heterogeneidade de sequência descritos inicialmente por (CALVETE, SANTOS, *et al.*, 1998).

Além de comparar as três sequências consenso entre si e com a VML, também tentamos alinhar essas sequências com outras lectinas de leguminosas. Para isso, as sequências de aminoácidos das lectinas de leguminosas similares à VML foram alinhadas e foi recortada manualmente a região que flanqueasse os sítios **PIQ(H)IWD** e **PEWVRV** que são bastantes conservados. Tal estratégia permitiria que as sequências de aminoácidos das lectinas de leguminosas juntamente com as sequências traduzidas das três sequências obtidas no experimento pudessem ser alinhadas em um programa de alinhamento global Clustal sendo visualizado graficamente pelo programa Bioedit como mostra a Figura 14. O alinhamento mostrou que as três sequências consenso apresentam alta similaridade com as outras lectinas de leguminosas. Alguns resíduos de aminoácidos presentes nas três sequências consenso puderam ser observados nas mesmas posições em *Robinia pseudoacacia*, *Sophora japonica* e *Erythrina corallodendron* (TABELA 6) (FIGURA 17). Alguns sítios de ligação a metais se mantiveram conservados nas três sequências consenso bem como nas lectinas de *R. pseudoacacia*, *S. japonica* e *E. corallodendron* (TABELA 7).

As relações filogenéticas entre as sequências consenso e outras lectinas de leguminosas estão esquematizadas na Figura 15. O dendograma obtido pelo método de distância do vizinho mais próximo (*Neighbor joining*) visualizado pelo programa MEGA mostrou que as três sequências consenso juntamente com a VML formam um grupo monofilético e o grupo é muito similar à lectina de *Sophora japonica* e a proteína relacionada à lectina de *Cladrastis lutea*.

Tabela 5 – Quantificação das amostras de DNA plasmidial recombinante de *V. macrocarpa* (pGemT-Vml) extraído de *E. coli* por lise alcalina

<b>Amostra</b>	<b>Conc.</b>	<b>A<sub>260</sub></b>	<b>A<sub>260</sub>/A<sub>280</sub></b>
Vml1	438 ng/ul	0,175	1,786
Vml 2	1740 ng/ul	0,696	1,928
Vml 3	975 ng/ul	0,390	1,893
Vml 4	348 ng/ul	0,139	1,527
Vml 6	1175 ng/ul	0,470	1,918
Vml 7	988 ng/ul	0,395	1,863

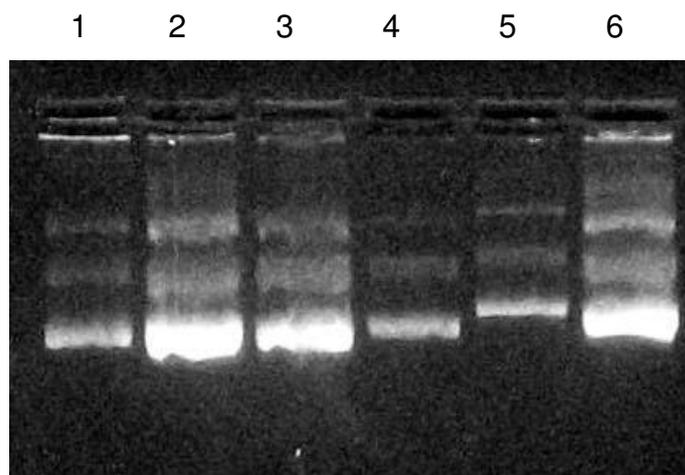


Figura 12 – Eletroforese em gel de agarose a 1% corada com brometo de etídio do plasmídio recombinante de *V. macrocarpa* extraído de *E. coli* por lise alcalina. Linhas de 1 a 4: Vml1, Vml2, Vml4, Vml6, respectivamente. Linhas 5 e 6: controle positivo.

	<i>contig3</i>	CCG	ATT	CAG	ATT	TGG	GAT	AAT	ACC	ACA	GGC	AGG	GTG	GCA	AGC	TTT	GCC		
	<i>contig2</i>	CCG	ATA	CAG	ATT	TGG	GAT	GAT	AGC	ACA	GCA	AGC	TTT	GCA	AGC	TTT	GCC		
	<i>contig1</i>	CCG	ATT	CAG	ATT	TGG	GAT	GAT	AGC	ACA	GCA	AGC	TTT	GCA	AGC	TTT	GCC		
	<i>contig3</i>	P	I	Q	I	W	D	N	T	T	G	R	V	A	S	F	A		
	<i>contig2</i>	P	I	Q	I	W	D	D	S	T	G	R	V	A	S	F	A		
	<i>contig1</i>	P	I	Q	I	W	D	D	S	T	D	R	V	A	S	F	A		
	VML	P	I	H	I	W	D	D	S	T	D	R	V	A	S	F	A		
																	70		
ACT	TCC	TTC	TCC	TTT	GTT	GTG	GAA	GCA	CCT	GAT	GCT	TCC	CAA	ATT	GCA	GAT	GGT	ATT	GCC
ACT	TCC	TTC	TCC	TTT	GTT	GTG	GAA	GCA	CCT	GAT	GCT	TCC	CAA	ATT	GCA	GAT	GGC	ATT	GCG
ACT	TCC	TTC	TCC	TTT	GTT	GTG	GAA	GCA	CCT	GAC	GAA	TCC	AAA	ACT	GCA	GAT	GGT	ATT	GCC
T	S	F	S	F	V	V	E	A	P	D	A	S	Q	I	A	D	G	I	A
T	S	F	S	F	V	V	E	A	P	D	A	S	K	T	A	D	G	I	A
T	S	F	S	F	V	V	E	A	P	D	E	S	K	T	A	D	G	I	A
T	S	F	S	F	V	V	E	A	P	D	E	S	K	T	A	D	G	I	A
																			90
TTC	TTT	CTT	GCA	CCA	CCA	GAT	ACT	CAA	ACC	GAA	TCA	TGT	GGG	GGA	TAC	CTT	GGA	CTT	TTC
TTC	TTT	CTT	GCA	CCA	CCA	GCT	ACT	CAA	CCC	CAA	GGT	CCT	GGG	GGT	CTC	CTT	GGA	CTT	TTC
TTC	TTT	CTT	GCA	CCT	CCA	GAT	ACT	CAA	CCC	CAA	AAA	AAT	GGG	GGT	TTC	CTT	GGA	CTT	TTC
F	F	L	A	P	P	D	T	Q	T	E	S	C	G	G	Y	L	G	L	F
F	F	L	A	P	P	A	T	Q	P	Q	G	P	G	G	L	L	G	L	F
F	F	L	A	P	P	D	T	Q	P	Q	K	N	G	G	F	L	G	L	F
F	F	L	A	P	P	D	T	Q	P	Q	K	D	G	G	F	L	G	L	F
																			110
CGC	AAC	AGT	AGT	GAC	GAC	AAT	TAC	AAT	AAA	TCC	AAC	CAA	ACT	GTT	GCA	GTG	GAA	TTT	GAT
CCT	GAC	AGT	GGT	-	-	-	TAC	AAT	GAC	TCC	AAC	CAA	ATT	GTT	GCA	GTG	GAA	TTT	GAC
AAT	GAT	AGT	-	-	-	-	-	AAT	AAA	TCC	ATC	CAA	ACT	GTT	GCA	GTG	GAA	TTT	GAT
R	N	S	S	D	D	N	Y	N	K	S	N	Q	T	V	A	V	E	F	D
P	D	S	G	-	-	-	Y	N	D	S	N	Q	I	V	A	V	E	F	D
N	D	S	-	-	-	-	-	N	K	S	I	Q	T	V	A	V	E	F	D
N	D	S	-	-	-	-	-	N	K	S	I	Q	T	V	A	V	E	F	D
																			130
ACT	TAC	TCT	AAC	GAC	TGG	GAT	CCA	AGT	ATT	AAA	CAT	ATT	GGA	ATT	GAT	GTC	AAC	TCT	ATT
ACT	TAC	CCT	GAC	GCA	TGG	GAT	CCA	AGT	TAT	ATA	CAT	ATT	GGA	ATT	GAT	GTC	AAC	TCT	ATT
ACT	TTC	TCT	AAC	ACT	TGG	GAT	CCA	AGT	GCT	AGA	CAT	ATT	GGA	ATT	AAT	GTT	AAC	TCT	ATT
T	Y	S	N	D	W	D	P	S	I	K	H	I	G	I	D	V	N	S	I
T	Y	P	D	A	W	D	P	S	Y	I	H	I	G	I	D	V	N	S	I
T	F	S	N	T	W	D	P	S	A	R	H	I	G	I	N	V	N	S	I
T	F	S	N	T	W	D	P	S	A	R	H	I	G	I	N	V	N	S	I
																			150
GTG	TCA	AAC	AAT	AAT	GTG	ATG	TGG	GTT	TGG	GAG	AAT	GGG	GAA	GTA	ACC	AAT	GTT	TAC	ATA
GTG	TCA	ACA	AAA	ATT	GTG	AAA	TGG	GTT	TGG	GAG	AAT	GGG	AAA	GTA	GCC	AAT	GCT	CTC	ATA
GAG	TCT	ATG	AAA	TAT	GTG	AAA	TGG	GGT	TGG	GAG	AAT	GGG	GAA	GTA	GCC	AAT	GTT	TAC	ATA
V	S	K	N	N	V	M	W	V	W	E	N	G	E	V	T	N	V	Y	I
V	S	T	K	N	V	K	W	V	W	E	N	G	K	V	A	N	A	L	I
E	S	M	K	Y	V	K	W	G	W	E	N	G	E	V	A	N	V	Y	I
E	S	M	K	Y	V	K	W	G	W	E	N	G	K	V	A	N	V	Y	I
																			170
AGC	TAT	AAA	GCT	TCC	ACC	AAA	ACT	TTA	ACT	GCC	TCT	CTA	ACT	TAC	CCT	TCA	AGT	GCA	ACA
AGC	TAT	GTA	GCT	TCC	ACC	AAA	ACC	TTA	ACT	GTC	TCT	CTA	ACT	TAC	CCT	TCA	AGT	GAC	ACA
AGC	TAT	CAA	GCT	TCC	ACC	AAA	ACA	TTA	ACT	GCC	TCT	TTA	ACT	TAC	CCT	TCG	AAT	GCA	ACA
S	Y	K	A	S	T	K	T	L	T	A	S	L	T	Y	P	S	S	A	T
S	Y	L	A	S	T	K	T	L	T	V	S	L	T	Y	P	S	S	D	T
S	Y	Q	A	S	T	K	T	L	T	A	S	L	T	Y	P	S	N	A	T
S	Y	E	A	S	T	K	T	L	T	A	S	L	T	Y	P	S	N	A	T
																			190
AGT	TAT	ATC	GTC	TCT	GCA	AAT	GTG	GAT	TTG	AAG	AGT	GCT	CTT	CCG	GAA	TGG	GTA	AGA	GTT
AGT	TAT	ATC	CTC	TCT	GCA	TAT	GTG	GAT	TTG	AAG	GAA	ATT	CTT	CCT	GAA	TGG	GTA	AGG	GTT
AGT	TAT	ATC	GTC	TCT	GCA	AAT	GTG	GAT	TTG	AAG	AGT	GCT	CTT	CCT	GAA	TGG	GTG	AGA	GTT
S	Y	I	V	S	A	N	V	D	L	K	S	A	L	P	E	W	V	R	V
S	Y	I	L	S	A	Y	V	D	L	K	S	A	L	P	E	W	V	R	V
S	Y	I	V	S	A	N	V	D	L	K	S	A	L	P	E	W	V	R	V
S	Y	I	V	S	A	N	V	D	L	K	S	A	L	P	E	W	V	R	V
																			210

Figura 13 – Alinhamento das sequências de DNA obtidas pelo sequenciamento da região central de *Vml* (*contigs* 1, 2 e 3). Abaixo, alinhamento da tradução dos *contigs*(1, 2 e 3) e sua tradução com a lectina de sementes de *V. macrocarpa* depositada no GenBank.

Tabela 6 – Heterogeneidade nos resíduos de aminoácidos encontrados na lectina de *V. macrocarpa* depositada e os resíduos encontrados nas mesmas posições das sequências internas obtidas por clonagem e sequenciamento da região central a partir de DNA genômico. A numeração foi baseada na sequência depositada VML. Het. (Heterogeneidade encontrada nos resíduos da VML), Rp (*Robinia pseudoacacia*), Sj (*Sophora japonica*), Ev (*Erythrina variegata*)

Posição	Het. VML	VML	Contig1	Contig2	Contig3	Rp	Sj	Ev
57		H	Q	Q	Q	Q	H	H
103		D	N	P	C	L	P	G
117	I/V	V	I	N	N	N	Y	Y
148	M/K	M	M	T	K	V	T	T
154	G/A	G	G	V	V	N	G	Q
159		K	E	K	E	E	E	Q
168	E/Q	E	Q	L	K	E	Q	D

Tabela 7– Resíduos de aminoácidos importantes na manutenção da estrutura tridimensional e ligação a carboidrato na lectina de *V. macrocarpa* e outras lectinas relacionadas. A numeração foi baseada na sequência depositada VML. Rp (*Robinia pseudoacacia*), Sj (*Sophora japonica*), Ev (*Erythrina variegata*), M (Sítio de ligação a metais), C (Sítio de ligação a carboidratos), Glc (local onde contem N-glicano)

Posição	VML	Contig1	Contig2	Contig3	Rp	Sj	Ev	Tipo
87	D	D	D	D	D	D	D	M/C
111	N	N	R	P	K	H	N	Glc
123	E	E	E	E	E	D	E	M
125	D	D	D	D	D	D	D	M
127	F	F	Y	Y	F	H	F	M
129	N	N	N	D	N	N	N	M/C
132	D	D	D	D	D	D	D	M/C
137	H	H	H	H	H	H	H	M
145	I	I	I	I	I	I	I	M
147	S	S	S	S	S	S	S	M
183	N	N	S	S	L	S	N	Glc

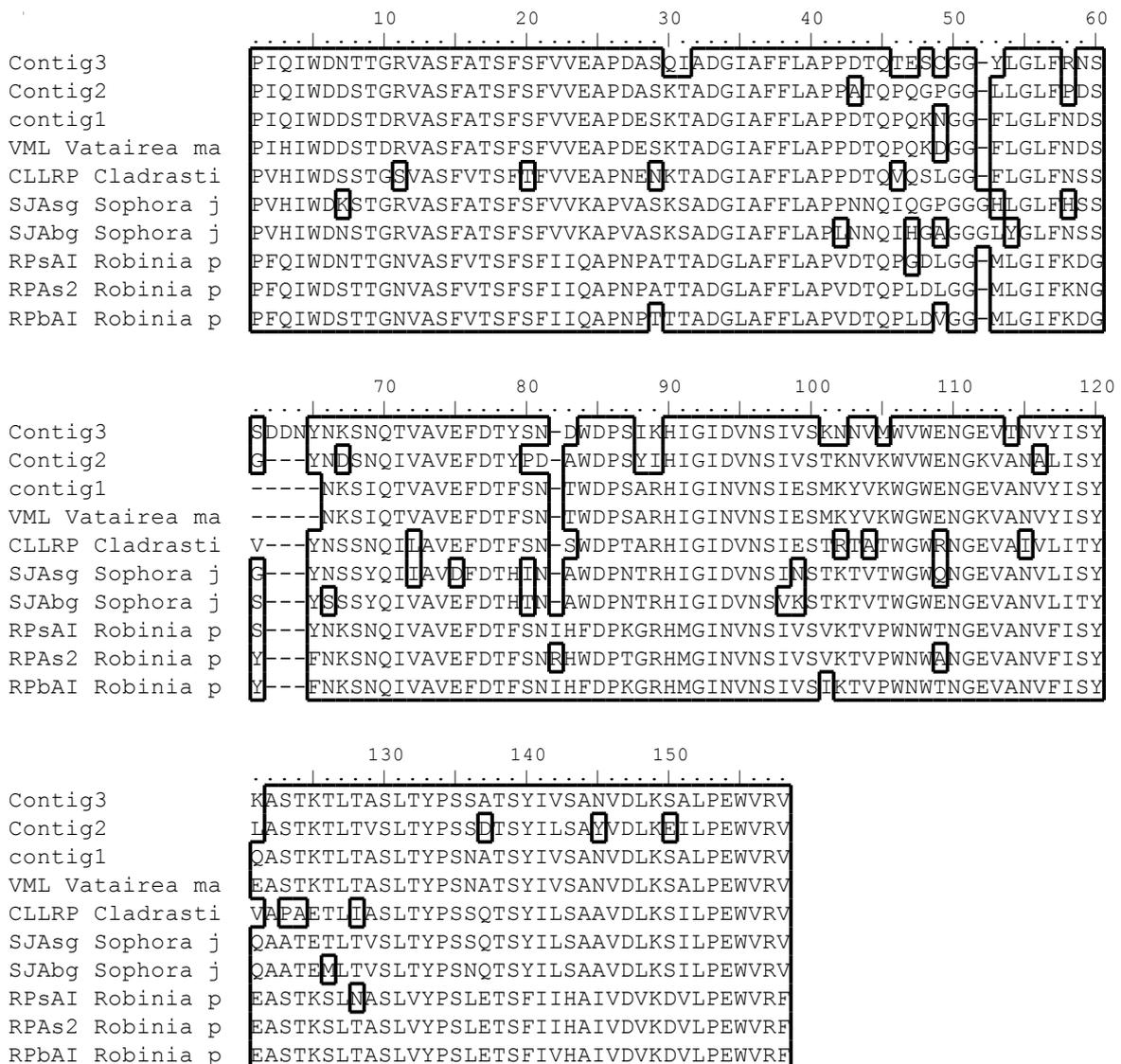


Figura 14 – Alinhamento da tradução dos três *contigs* da região central de *Vml* com lectinas de leguminosas depositadas no GenBank. VML (lectina de sementes de *Vatairea macrocarpa*), CLLRP (proteína relacionada a lectina de *Cladrastis lutea*), SJAsg (lectina galactose ligante de semente de *Sophora japonica*), SJAbg (lectina galactose ligante da casca de *S. japonica*), RPAsI (cadeia alfa da lectina de semente de *Robinia pseudoacacia*), RPAs2 (lectina de semente de *R. pseudoacacia*), RPbAI (lectina de casca de *R. pseudoacacia*).

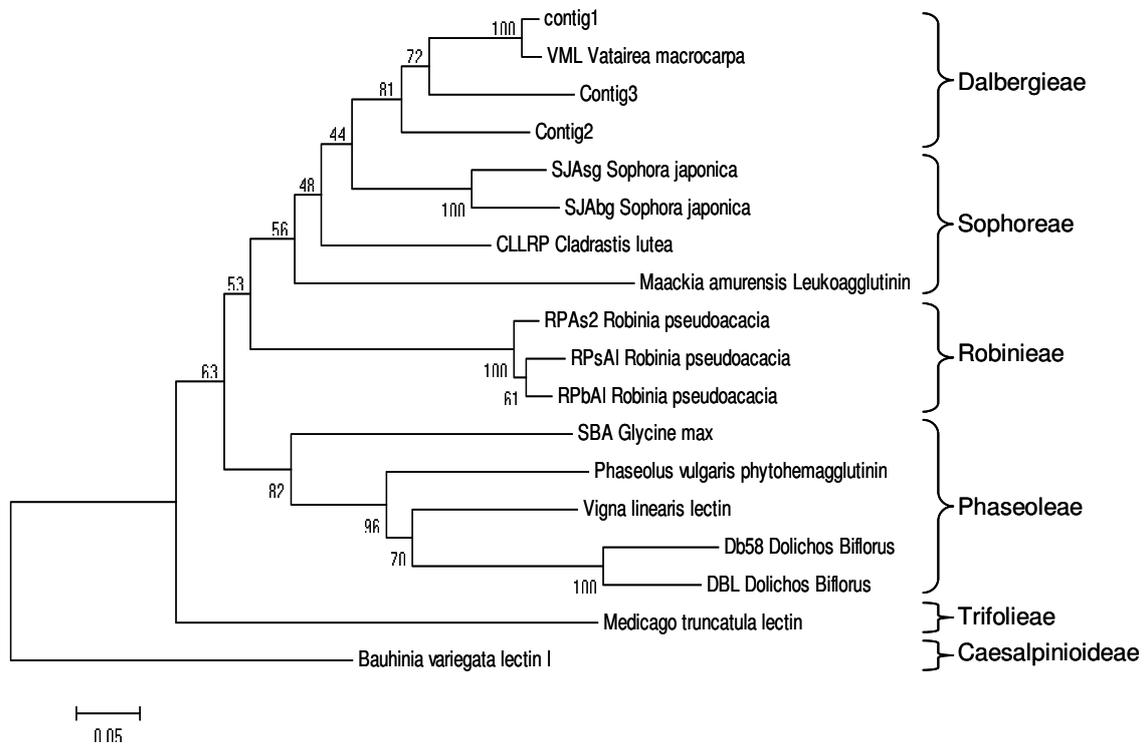


Figura 15 – Dendrograma obtido pelo método de distância do vizinho mais próximo (Neighbor-Joining), utilizando a distância-p como modelo de substituição a partir das sequências centrais traduzidas das três sequências consenso *contigs* e outras lectinas de leguminosas. Os números representam valores de *bootstrap*. A escala representa mudanças entre as sequências no comprimento do ramo.

### 5.7. Amplificação Rápida das Extremidades de cDNA (RACE)

O isolamento de RNA total de *Vatairea macrocarpa* utilizando o Plant RNA Reagent™ foi feita em quadruplicata. O resultado revelado pela eletroforese em gel de agarose a 1% mostra bandas contendo o tRNA, rRNA18S, rRNA28s, e um *smear* contendo mRNA típicas de extração de RNA total sem sinais de degradação (FIGURA 16). As concentrações variaram de 1512 a 2540 ng/μl (TABELA 8).

Após a obtenção de cDNA de qualidade, procedeu-se com a amplificação das extremidades de cDNA pela técnica de 3'RACE. Os *primers* 3RACEVML e 3RACEVML2 se mostraram eficientes na amplificação das extremidades do gene da VML (FIGURA 17).

O *primer* desenhado para o peptídeo sinal se mostrou eficiente na amplificação dos cDNAs de *V. macrocarpa* amplificando a extremidade 5' utilizando o *primer* GSP2 como reverso. O fragmento inteiro do peptídeo sinal até a cauda poli A também foi amplificada (FIGURA 18). O sequenciamento da região 5' mostrou sequências de baixa qualidade mesmo nas sequências clonadas. Após várias tentativas não foi possível clonar o fragmento inteiro de *Vml* (*primers*: SignalVML + Q0). Contudo, o sequenciamento do fragmento obtido diretamente por PCR mostrou sequências de qualidade, mas essas não serviram para a construção do fragmento inteiro devido as perdas ocorridas nas extremidades da sequência.

Tabela 8 – Quantificação das amostras de RNA total extraídas pelo Plant RNA Reagent. A concentração das amostras varia de 1512 ng/μl a 2540 ng/μl, a absorbância  $A_{260}$  varia entre 0,219 a 0,545 e a razão entre  $A_{260}/A_{280}$  varia entre 1,707 a 2,025.

Amostra	Conc.	$A_{260}$	$A_{260}/A_{280}$
VML1	2540 ng/ul	0,545	1,707
VML2	1828 ng/ul	0,219	1,97
VML3	1512 ng/ul	0,295	1,909
VML4	1976 ng/ul	0,226	2,025

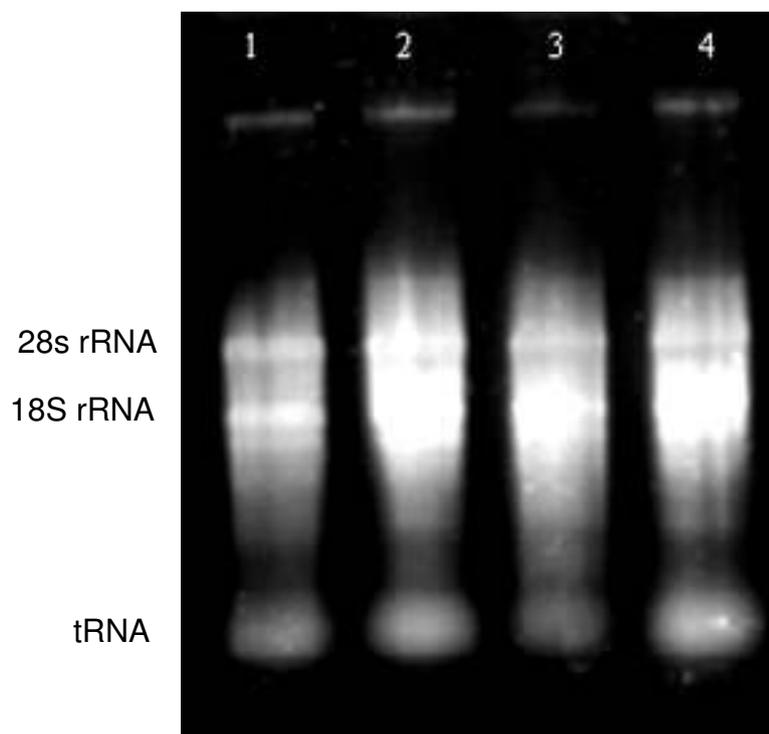


Figura 16 – Gel de agarose a 0,8% corado com brometo de etídio do resultado de extração de RNA total de *V. macrocarpa* pelo kit Plant RNA Reagent. A extração foi feita em quadruplicada conforme mostrado nos poços de 1 a 4.

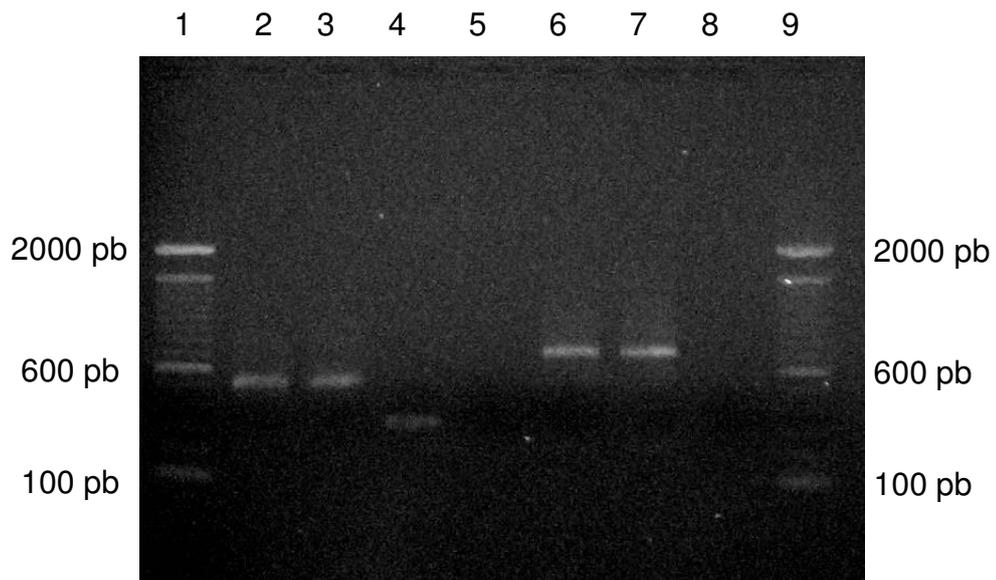


Figura 17 – Eletroforese em gel de agarose a 1% mostrando a 3'RACE da *Vml* utilizando os *primers* desenhado a partir da região central já sequenciada com os *primers* para o adaptador. Poços 1 e 9: marcador de peso molecular *Ladder* 100 pb; Poços 2 e 3: cDNA de *V. macrocarpa* amplificado com os *primers* 3RACEVML + Q1 com temperatura de anelamento de 55 e 58°C respectivamente; Poços 6 e 7: cDNA de *V. macrocarpa* amplificado com os *primers* 3RACEVML2 + Q1 com temperatura de anelamento de 55 e 58°C respectivamente; poços 5 e 8: controle negativo da reação; poço 4: controle positivo da reação.

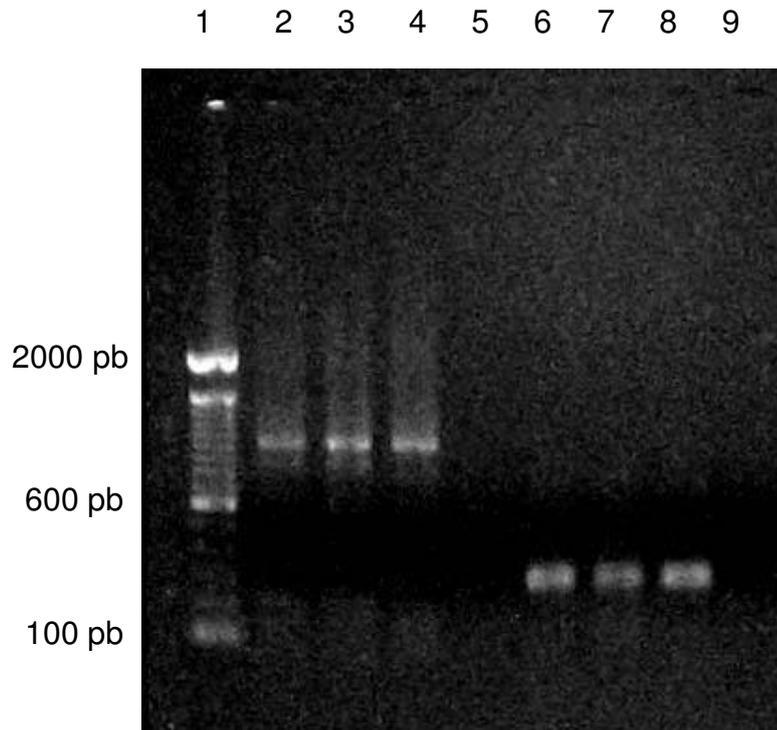


Figura 18 – Eletroforese em gel de agarose a 1% mostrando a amplificação da extremidade 5' da *Vml* utilizando os *primers* desenhado a partir do peptídeo sinal com os *primers* desenhados a partir da região interna já sequenciada e para o *primer* do adaptador da cauda poliA. Poço 1: marcador de peso molecular *Ladder* 100 pb; Poços 2, 3 e 4: cDNA de *V. macrocarpa* amplificado com os *primers* Signalvml + Q0; Poços 6, 7 e 8: cDNA de *V. macrocarpa* amplificado com os *primers* Signalvml + GSP2; poços 5 e 9: controle negativo da reação.

### 5.8. Alinhamento da região 3' de *Vml*

O *ContigA* formado pelo sequenciamento da região 3' a partir do cDNA de folhas mostrou-se maior (14 resíduos do final do *contig3* até o primeiro códon de parada) e diferente em tipos de aminoácidos (10 resíduos diferentes do final do *contig3* até o primeiro códon de parada) do que a lectina de sementes citada por CALVETE *et al* (1998). O alinhamento desta sequência com os três *contigs* obtidos a partir de DNA genômico mostrou que há complementaridade entre esta sequência e o *contig3* havendo sobreposição em 52 resíduos de aminoácidos (FIGURA 19).

A região 3' do *ContigA* codifica 49 resíduos adicionais ao final do *contig3* até o primeiro códon de parada e a sequência de nucleotídeos continua com 165 nucleotídeos adicionais até a cauda poli-A.

A sequência resultante da união das duas seqüências consenso mostrados na Figura 19 revela uma proteína com 19 inserções de aminoácidos do resíduo 55 da VML até o códon de parada e 40 substituições de aminoácidos indicando diferenças significativas entre essas seqüências.

O alinhamento da sequência híbrida *ContigA* e *contig3* juntamente com outras lectinas revela que há grande heterogeneidade de aminoácidos na região C-terminal das lectinas analisadas (FIGURA 20).

```

Contig3          CCG ATT CAG ATT TGG GAT AAT ACC ACA GGC AGG GTG GCA AGC TTT GCC 48
                  P  I  Q  I  W  D  N  T  T  G  R  V  A  S  F  A

Contig3 ACT TCC TTC TCC TTT GTT GTG GAA GCA CCT GAT GCT TCC CAA ATT GCA GAT GGT ATT GCC 108
          T  S  F  S  F  V  V  E  A  P  D  A  S  Q  I  A  D  G  I  A

Contig3 TTC TTT CTT GCA CCA CCA GAT ACT CAA ACC GAA TCA TGT GGG GGA TAC CTT GGA CTT TTC 168
          F  F  L  A  P  P  D  T  Q  T  E  S  C  G  G  Y  L  G  L  F

Contig3 CGC AAC AGT AGT GAC GAC AAT TAC AAT AAA TCC AAC CAA ACT GTT GCA GTG GAA TTT GAT 228
          R  N  S  S  D  D  N  Y  N  K  S  N  Q  T  V  A  V  E  F  D

Contig3 ACT TAC TCT AAC GAC TGG GAT CCA AGT ATT AAA CAT ATT GGA ATT GAT GTC AAC TCT ATT 288
          T  Y  S  N  D  W  D  P  S  I  K  H  I  G  I  D  V  N  S  I

Contig3 GTG TCA AAC AAT AAT GTG ATG TGG GTT TGG GAG AAT GGG GAA GTA ACC AAT GTT TAC ATA 348
ContigA          V  S  K  N  N  V  M  W  V  W  E  N  G  E  V  T  N  V  Y  I

Contig3 AGC TAT AAA GCT TCC ACC AAA ACT TTA ACT GCC TCT CTA ACT TAC CCT TCA AGT GCA ACA 408
ContigA AGC TAT AAA GCT TCC ACC AAA ACT TTA ACT GCC TCT CTA ACT TAC CCT TCA AGT GCA ACA 96
          S  Y  K  A  S  T  K  T  L  T  A  S  L  T  Y  P  S  S  A  T

Contig3 AGT TAT ATC GTC TCT GCA AAT GTG GAT TTG AAG AGT GCT CTT CCG GAA TGG GTA AGA GTT 468
ContigA AGT TAT ATC GTC TCT GCA AAT GTG GAT TTG AAG AGT GCT CTT CCG GAA TGG GTA AGA GTT 156
          S  Y  I  V  S  A  N  V  D  L  K  S  A  L  P  E  W  V  R  V

ContigA GGT TTC TCT GCT ACC ACA GGA TGT AAA GTA TCA AGA GAT AAA GCT GAA ACG CAT AAT GTT 216
          G  F  S  A  T  T  G  C  K  V  S  R  D  K  A  E  T  H  N  V

ContigA CTT AAC TGG TCT TTC ACT TCA TCC TTG CAA GCC ACC AGA GAT GGT GCT TAT CAT GAG AAC 276
          L  N  W  S  F  T  S  S  L  Q  A  T  R  D  G  A  Y  H  E  N

ContigA AAC GCT CAT CTT GCA AGT TAT GCT GCA TGA TTC CAT GCA CTA GTT ATC GGT GAC AAG TAG 336
          N  A  H  L  A  S  Y  A  A  *

ContigA AAA AAG CTA CAA TAA GAC CTG CAA CAA TGG TCA TCA TAT GTA GTA CTA CAG CAG CGG CCA 396

ContigA CGC TGA CCA GCT ATT GTA AAT TTG TAA TGG TTG TAA TAG TGT TGT TGT TAT AAT AAT GTT 456

ContigA AAT AAG TTT GTA CCC AAA AAA AAA AAA AAA  A 487

```

Figura 19 – Alinhamento do *ContigA* obtida pela técnica 3'RACE a partir de cDNA de folhas (até o primeiro códon de parada) e sua complementaridade com o *contig3* previamente identificado a partir de DNA genômico.

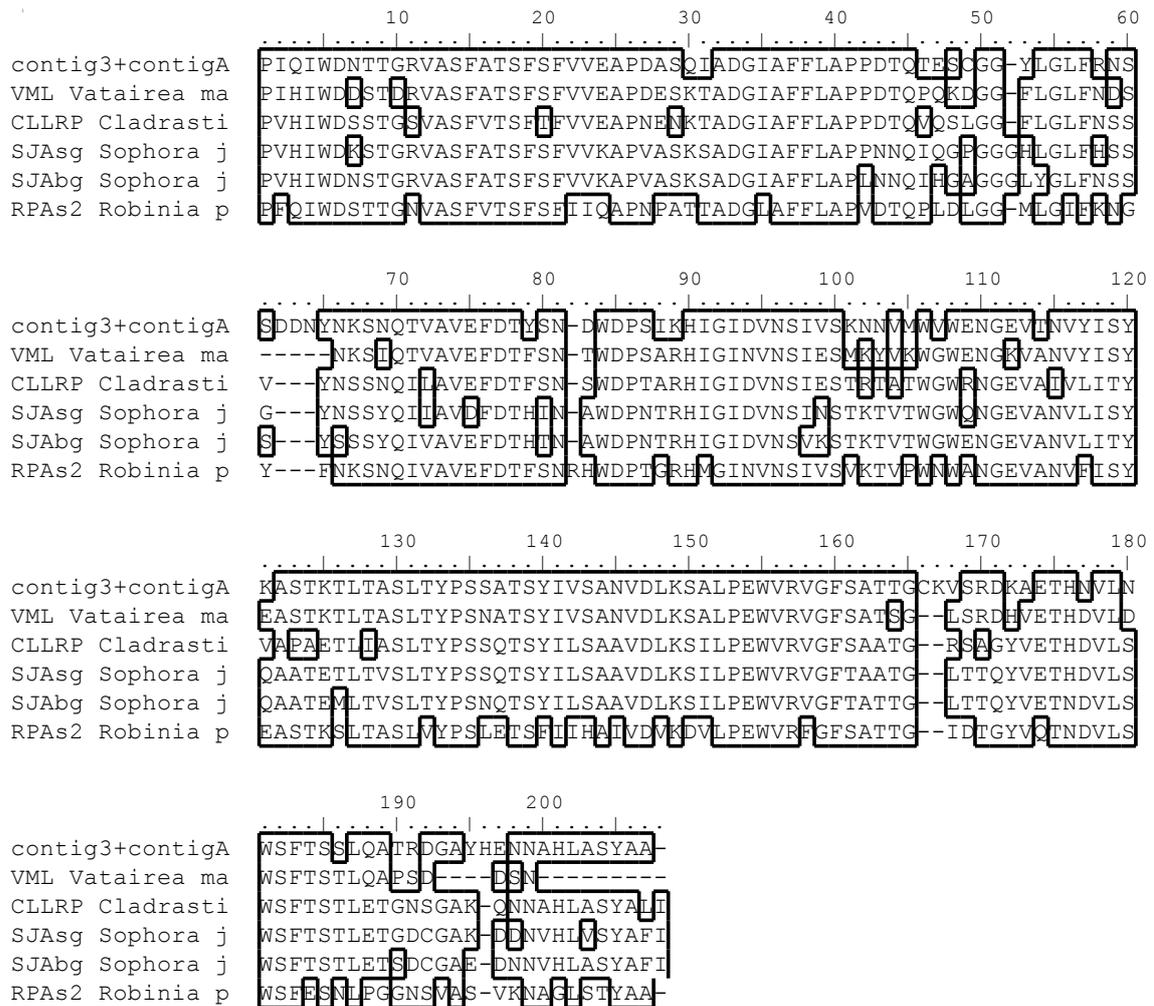


Figura 20 – Região C-terminal da sequência obtida pela união do *contig3* com o *contigA* até o primeiro códon de parada alinhada com sequências de precursores de outras lectinas de leguminosas: VML (lectina de sementes de *Vatairea macrocarpa*), CLLRP (proteína relacionada a lectina de *Cladrastis lutea*), SJAsg (lectina galactose ligante de semente de *Sophora japonica*), SJAbg (lectina galactose ligante da casca de *S. japonica*), RPAs2 (lectina de semente de *R. pseudoacacia*).

## 6. DISCUSSÃO

Segundo (SAMBROOK, FRITSCH e MANIATIS, 1989), a relação entre as absorvâncias a 260 nm e 280 nm deve ser maior que 1,75 para que uma amostra seja considerada de boa qualidade e livre de contaminação de proteínas. Contudo, outros pesquisadores defendem que amostras que apresentem relação entre 1,6 e 2,0 já possuem qualidade satisfatória sendo que valores maiores que 2,0 indicam contaminação por fenol ou com outros tipos de álcoois (ROMANO, 1998). No presente trabalho, obtivemos valores de concentração de 105 ng/ $\mu$ l e relação  $A_{260}/A_{280}$  de 1,5. Além disso, os fragmentos de DNA obtidos possuem tamanhos iguais ou superiores a 23.000 pb não apresentando degradação aparente, possuindo tamanho e qualidade necessários para a realização dos estudos propostos (CUNHA, 2002). Assim, a utilização do detergente CTAB e da metodologia descrita para extração de DNA genômico se mostrou eficiente para obtenção de material genético de boa qualidade.

Plantas de uma mesma espécie frequentemente contêm várias lectinas, tanto no mesmo tecido como em tecidos diferentes, e algumas dessas lectinas são codificadas por genes diferentes tanto que famílias de genes de lectinas são comuns na família das leguminosas (SHARON e LIS, 2003).

O feijão comum (*Phaseolus vulgaris*) contém uma abundante lectina que possui atividades mitogênica e hemaglutinante. Esta lectina tetramérica é constituída pelas subunidades PHA-E e PHA-L que são codificados por dois genes distintos *dlec1* e *dlec2*, respectivamente. Esses genes, que possuem 90% de similaridade de sequência a nível de nucleotídeos, codificam proteínas que tem 82% de identidade nas suas sequências de resíduos aminoácidos. (HOFFMAN e DONALDSON, 1987), antes de relatarem a sequência de aminoácidos da PHA-E e PHA-L, isolaram um cDNA que codifica uma proteína relacionada a lectina com 27 kDa. A presença de dois resíduos de metionina na sequência de aminoácidos derivada indica que este cDNA não codifica PHA porque todas as análises de aminoácidos da PHA purificada mostraram uma completa ausência de resíduos de metionina. Esse novo gene codifica um inibidor de alfa amilase –  $\alpha$ AI e passou a receber o nome de *ai* o qual tem 82% de similaridade de sequência com *dlec1* e *dlec2* (CHRISPEELS e RAIKHEL, 1991).

O sequenciamento dos fragmentos de DNA genômico de *V. macrocarpa* amplificados por PCR revelou a presença de pelo menos três genes distintos relacionado à VML. O contig1 corresponde ao gene que codifica a VML enquanto o contig2 e contig3 possuem 2 e 5 inserções de aminoácidos, respectivamente. Além disso, o contig3 possui 85% de similaridade de sequência a nível de nucleotídeos, e codifica uma proteína com 77% de identidade com a VML. O Contig2 é o mais distante em termos de similaridade de sequência apresentando 82% de similaridade e 74% de identidade.

Segundo (VAN DAMME, PEUMANS, *et al.*, 1998), para se discutir sobre as famílias gênicas de lectinas de leguminosas é necessário que se faça uma clara distinção entre isolectinas e lectinas diferentes. Neste caso, quando uma leguminosa contém múltiplas lectinas, aquelas com a mesma especificidade é geralmente considerada isolectina, enquanto que lectinas com especificidade diferente são consideradas como verdadeiramente diferentes. As lectinas devem ser consideradas como verdadeiras isoformas se elas são derivadas de protômeros com a mesma sequência de aminoácidos. Diferenças na glicosilação podem gerar glicofomas enquanto diferenças nas modificações pós-traducionais podem gerar isoformas. Sendo assim, todas as outras lectinas, devem ser consideradas proteínas diferentes. Um exemplo de família multigênica é *Robinia pseudoacacia* que contém duas lectinas de casca: a primeira chamada RPbAI é composta de cinco isolectinas que se associam de dois polipeptídeos em tetrâmeros, A segunda lectina é um homotetrâmero chamado RPbAII. O alinhamento dos cDNAs mostram que a sequência de DNA traduzida mostra 80% de similaridade de sequência para os dois polipeptídeos da RPbAI e são homólogos com a lectina de semente. Por outro lado, a sequência que codifica RPbAII mostra apenas 33% de identidade de sequência com as lectinas previamente descritas. Análises de *Southern blot* de DNA genômico isolado de *R. pseudoacacia* revelaram que as lectinas de casca são resultado da expressão de uma pequena família de genes (VAN DAMME, BARRE, *et al.*, 1995).

A lectina de *V. macrocarpa* é processada pós-traducionalmente. Inicialmente, ocorre uma deglicosilação na asparagina 111 seguida pela clivagem proteolítica da ligação entre os resíduos de asparagina 114 e lisina 115. O Alinhamento da tradução das três sequências consenso revelou que os *contigs* 2 e 3 não possuem resíduos de asparagina nos sítios de glicosilação, que supostamente seriam necessários para as etapas de clivagem proteolítica. Além disso, entre os resíduos

114 e 115 existem inserções de um resíduo de glicina e tirosina no *contig2*. No *contig3* ocorrem 5 inserções de aminoácidos (Ser, Asp, Asp, Asn e Tyr) sendo que três deles (Asp, Asp e Asn) não estão presentes nas lectinas de *Sophora japonica* e *Robinia pseudoacacia* (FIGURA 13).

A complexidade da organização dos genes de lectinas varia de acordo com as espécies analisadas. Em ervilha, as lectinas pertencem a uma pequena família de genes relacionados, enquanto em *Phaseolus*, as lectinas e genes relacionados às lectinas são codificadas por famílias multigênicas e suas sequências de nucleotídeos tem sido utilizadas com sucesso para estudos filogenéticos e evolucionários (LIOI, GALASSO, *et al.*, 2006).

Até o momento da realização deste trabalho, não se sabia a sequência de nucleotídeos que codificava a lectina de *V. macrocarpa*. O objetivo inicial do trabalho foi obter a sequência completa que codifica a VML e sua expressão em células de *E. coli*. Contudo, após dificuldades na amplificação e sequenciamento, foi constatada a presença de sequências que, mesmo apresentando regiões conservadas, não poderiam ser classificados como sendo provenientes do mesmo gene. Neste trabalho foi constatada pela primeira vez a sequência parcial dos genes que podem codificar novas lectinas ou proteínas relacionadas em *V. macrocarpa*.

Atualmente, estamos continuando os experimentos para o sequenciamento do gene inteiro de *Vml* com *primers* construídos a partir da informação do peptídeo sinal e um outro específico para a extremidade 3' construído em uma região conservada do gene *Vml* previamente determinada pelo sequenciamento do resultado do 3'RACE. Também estão sendo feitos experimentos de PCR em tempo real a partir do cDNA de folhas para se avaliar o nível de expressão relativo de cada um dos genes em condições normais.

Após a caracterização completa dos genes relacionados à lectina de *V. macrocarpa*, o passo seguinte será a expressão de cada uma delas em sistemas heterólogos seguidos pela atividade aglutinante das proteínas recombinantes. Depois disso, a estrutura tridimensional da VML e proteínas relacionadas serão determinadas por difração de raios-X das proteínas cristalizadas. Espera-se que as proteínas relacionadas aos contigs 2 e 3 cristalizem com mais facilidade, uma vez que apresentam indícios de não serem proteínas glicosiladas, facilitando sua cristalização.

Estudos de PCR em Tempo Real de lectinas em diferentes tecidos e condições de estresse juntamente com estudos de RNA de interferência darão importantes informações sobre as funções dos genes relacionados à lectina de *V. macrocarpa*.

Estamos apenas no início de um estudo que irá contribuir para o melhor entendimento da família de genes relacionada à VML, suas vias de biossíntese e seu potencial biotecnológico. Estudos filogenéticos mais aprofundados das sequências relacionadas à VML juntamente com a região do rDNA ITS irão dar boas contribuições para o entendimento da evolução de leguminosas utilizando como organismo experimental uma planta ainda pouco estudada e que aparentemente não sofreu alterações genéticas com a influência humana ao contrário das espécies cultivadas.

## 7. SÍNTESE DE RESULTADOS E CONCLUSÕES

- *Vatairea macrocarpa* possui pelo menos três genes homólogos, mas não idênticos, que codificam lectinas ou proteínas relacionadas.
- A construção de árvores filogenéticas das sequências obtidas revela que as três sequências formam um grupo monofilético e são evolutivamente próximas das lectinas de *Sophora japonica*, *Robinia pseudoacacia* e da proteína de armazenamento relacionada à lectina de *Cladrastis lutea*.

Com base nos resultados apresentados, pode-se concluir que existe uma família de genes homólogos que codificam lectinas e/ou proteínas relacionadas à lectina de *V. macrocarpa*. Contudo, é necessário um estudo mais aprofundado para uma melhor caracterização desses genes. Este constitui o primeiro relato da presença de famílias multigênicas codificando potenciais lectinas em *V. macrocarpa*.

## BIBLIOGRAFIA

ABE, Y. SHIRANE, K., YOKOSAKA, H., MATSUSHITA, H., MITTA, H., KATO, I. *et al.* Asparginyl endopeptidase of jack bean seeds. **J. Biol. Chem.**, v. 268, p. 3525-3529, 1993.

ALENCAR, N. M. N. **Estudo do mecanismo e mediadores químicos envolvidos na atividade pró-inflamatória das lectinas de sementes de *Vatairea macrocarpa* e *Dioclea violacea***. Fortaleza: Universidade Federal do Ceará , 2001.

ALENCAR, N. M. N. ASSREUY, A. M., ALENCAR, V. B., MELO, S. C., RAMOS, M. V., CAVADA, B. S., *et al.* The galactose-binding lectin from *Vatairea macrocarpa* seeds induces in vivo neutrophil migration by indirect mechanism. **The international journal of biochemistry and cell biology**, v. 35, p. 1674-1681, 2003.

ALENCAR, N. M. N. ASSREUY, A. M., CRIDDLE, D. N., SOUZA, E. P., SOARES, P. M., HAVT, A. *et al.* *Vatairea macrocarpa* lectin induces paw edema with leucocyte infiltration. **Protein and Peptide Letters**, v. 11, p. 195-200, 2004. ISSN 2.

ALENCAR, N. M. N. ASSREUY, A. M., HAVT, A., BENEVIDES, R. G., MOURA, T. R., SOUZA, R. B. *et al.* *Vatairea macrocarpa* (Leguminosae) lectin activates cultured macrophages to release chemotactic mediators. **Naunyn-Schmiedeberg's Arch Pharmacol** , v. 374, p. 275-282, 2007.

ALTSCHUL, S. F. GISH, W., MILLER, W., MYERS, E. W., & LIPMAN, D. J. Basic local alignment search tool. **J. Mol. Biol.**, v. 215 , p. 403–410, 1990.

AVICHEZER, D.; ARNON, R. Differential reactivities of the *Arachis hypogea* (peanut) and *Vicia villosa* B4 lectins with human ovarian carcinoma cells, grown either in vitro or in vivo xenograft model. **FEBS Letters**, v. 395, p. 103-108, 1996.

BANERJEE, R. DAS, K., RAVISHANKAR, R., SUGUNA, K., SUROLIA, A., & VIJIYAN, M. Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of the peanut lectin-lactose complex. **J. Mol. Biol.**, v. 259, p. 281-296, 1996.

BARONDES, S. H. Bifunctional properties of lectins: lectins redefined. **TIBBS**, v. 13, p. 480-482, 1988.

BECKER, J. W. REEKE, G. N., WANG, J. L., CUNNINGHAM, B. A., & EDELMAN, G. M. The covalent structure and three-dimensional structure of concanavalin A III. Structure of the monomer and its interactions with metals and saccarides. **J. Biomol. Chem.**, v. 250, p. 1513-1524, 1975.

BOLLINI, R. CERIOTTI, A., DIMINATI, M. G., & VITALE, A. Glycosilation is not needed for the intracellular transport of phytohemagglutinin in developing *Phaseolus vulgaris* cotyledons and for the maintenance of its biological activities. **Physiologia Plantarum**, v. 65, p. 15-22, 1985.

BOYD, W. C.; REGUERA, R. M. Studies on hemagglutinins present in seeds of some representatives of the family Leguminosae. **J. Immunol.**, v. 62, p. 333-339, 1949.

BROWN, S. M. **Bioinformatics: A Biologist's Guide to Biocomputing and the Internet**. Natick: Eaton Publishing, 2000. 188p p.

CALVETE, J. J. SANTOS, C. F., MANN, K., GRANGEIRO, T. B., NIMTZ, M., URBANKE, C., *et al.* Amino acid sequence, glican structure, and proteolytic processing of the lectin of *Vatairea macrocarpa* seeds. **FEMS Letters** , v. 425, p. 286-292, 1998.

CALVETE, J. J. THOLEM, H. H., RAIDIA, M., URBANKE, C., ROMERO, A., GRANGEIRO, T. B. *et al.* Molecular characterization and crystalization of Diocleinae lectins. **Biochimica and Biophysica Acta**, v. 1430, p. 367-375, 1999.

CARLINI, C. R.; GROSSI-DE-SA, M. F. Plant toxic proteins with inseticidal properties: A review on their potencialities as bioinsecticides. **Toxicon**, v. 40, p. 1515-1539, 2002.

CAVADA, B. S. MOREIRA, R. A., OLIVEIRA, J. T., & GRANGEIRO, T. B. Primary structure and functions of plant lectins. **R. Bras. Fisiol. Veg.**, v. 2, p. 193-201, 1993.

CAVADA, B. S. SANTOS, C. F., GRANGEIRO, T. B., NUNES, E. P., SALES, P. V., RAMOS, R. L. *et al.* Purification and characterization of a lectin from seeds of *Vatairea macrocarpa* duke. **Phytochemistry**, v. 49, p. 675-680, 1998.

CHRISPEELS, M. J.; HARTL, P. M.; FAYE, L. Characterization of the endoplasmic reticulou-associated precursor of concanavalin A. **J. Biol. Chem.**, v. 261, p. 10021-10024, 1986.

CHRISPEELS, M. J.; RAIKHEL, N. V. Lectins, lectins genes and their role in plant defense. **The Plant Cell**, v. 3, p. 1-9, 1991.

COOK, G. M. W. Cell surface carbohydrates: molecules in search of a function. **J. Cell Sci. Suppl.**, v. 4, p. 45-70, 1986.

CUNHA, R. M. S. **Filogenia molecular em Anacardium (Anacardiaceae):** utilização do gene da subunidade pequena do RNA ribossômico (SSU rRNA). Fortaleza: Universidade Federal do Ceará, 2002.

DAM, T. K. ., GERKEN, T. A., CAVADA, B. S., NASCIMENTO, K. S., MOURA, T. R., & BREWER, F. *et al.* Binding Studies of a-GalNAc-specific Lectins to the a-GalNAc (Tn-antigen) Form of Porcine Submaxillary Mucin and Its Smaller Fragments. **The Journal OF Biological Chemistry**, v. 282, p. 28256–28263, 2007.

DIAZ, C. L. MELCHERS, L. S., HOOYKAAST, P. J., LUGTENBERG, B. J., & KIJNE, J. W. *et al.* Root lectin as specificity in the Rhizobium-legume symbiosis. **Nature**, v. 338, p. 579-581, 1989.

ELTZER, M. E. The Dolichos biflorus lectins family: A model system for studying legume lectin structure and function. In: VAN DRIESSCHE, E., et al. **Lectins: biology, biochemistry, clinical biochemistry.** [S.l.]: [s.n.], v. 11, 1996. p. 3-9.

ETZLER, M. E. Distribution and function of plant lectins. In: LIENER, I. E.; SHARON, N.; GOLDSTEIN, I. J. **The lectins: properties, functions and applications in biology and medicine.** Orlando: Academic Press Inc., 1986. p. 371-435.

ETZLER, M. E. Plant lectins: molecular biology, synthesis, and function. In: ALLEN, H. J.; KISAILUS, E. C. **Glycoconjugates.** New York: Marcel Dekker, 1992. p. 521-531.

ETZLER, M. E. MURPHY, J. B., EWING, N. N., KALSI, G., STACEY, G., DAY, B. D., *et al.* A NOD factor-binding root lectin/nucleotide phosphatase: a possible receptor/signal transducer in the Rhizobium-legume symbiosis. **Glycoconjugate J.**, v. 14, p. 27, 1997.

ETZLER, M. E. KALSI, G., EWING, N. N., ROBERTS, N. J., DAY, R. B., & MURPHY, J. B. Nod factor binding lectin with apyrase activity from legume roots. **Proc. Natl. Acad. Sci.**, v. 96, p. 5856-5861, 1999.

EWING, B.; GREEN, P. Base calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v. 8, p. 186-194, 1998.

FAYE, L. STURM, A., BOLLINI, R., VITALE, A., & CHRISPPEELS, M. J. The position on the oligosaccharide side-chains of phytohemagglutinin and their accessibility to glycosidase determines their subsequent in the Golgi. **Eur. J. Biochem.**, v. 158, p. 655-661, 1986.

FAYE, L.; CHRISPPEELS, M. J. Transport and processing of the glycosylated precursor of Concanavalin A in jack-bean. **Planta**, v. 170, p. 217-224, 1987.

FROHMAN, M. A.; MARTIN, G. R. Rapid Amplification of cDNA ends using nested primers. **Techniques**, v. 1, p. 165-170, 1989.

GIBAS, C.; JAMBECK, P. **Developing Bioinformatics Computer Skills**. [S.l.]: O'Reilly & Associates, Inc., 2001. 375p p.

GOLDSTEIN, I. J. Lectin Structure-Activity: The Story Is Never Over. **Agric. Food Chem.**, v. 50, p. 6583-6585, 2002.

GOLDSTEIN, I. J. HUGHES, R. C., MONSIGNY, M., OSAWA, T., & SHARON, N. What should be called a lectin? **Nature**, v. 285, p. 66, 1980.

GOLDSTEIN, I. J.; PORETZ, R. D. Isolation and properties of lectins. In: LIENER, I. C.; SHARON, N.; GOLDSTEIN, I. J. **The lectins: properties, functions and applications in biology and medicine**. London: [s.n.], 1986.

GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: A graphical tool for sequence finishing. **Genome Research**, v. 8, p. 195-202, 1998.

GRANGEIRO, T. B. SCHRIEFER, A., CALVETE, J. J., RAIDA, M., URBANKE, C., BARRAL-NETO, M., *et al.* Molecular cloning and characterization of ConBr, the lectin of *Canavalia brasiliensis* seeds. **European Journal of Biotechnology**, v. 248, p. 43-48, 1997.

HAMELRYCK, T. W. LORIS, R., BOUCKAERT, J., DAO-THI, M. H., STRECKER, G., IMBERTY, A. *et al.* Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. **J. Mol. Biol.**, v. 286, p. 1161-1177, 1999.

HEMPERLY, J. J.; MOSTOV, K. E.; CUNNINGHAM, B. A. In vitro translation and processing of a precursor form of favin, a lectin from *Vicia faba*. **J. Biol. Chem.**, v. 257, p. 7903-7909, 1982.

HERMAN, E. M.; SHANNON, L. M.; CHRISPPEELS, M. J. Concanavalin A is synthesized as a glycoprotein precursor. **Planta**, v. 165, p. 23-29, 1985.

HIGGINS, T. J. V. CHANDLER, P. M., ZURAWSKI, G., BUTTON, S. C., & SPENCER, D. The biosynthesis and primary structure of pea seed lectin. **J. Biol. Chem.**, v. 258, p. 9544-9549, 1983.

HOFFMAN, L. M.; DONALDSON, D. D. Synthesis of mitogenic phytohemagglutinin-L in *Escherichia coli*. **Biotechnology**, v. 5, p. 157-160, 1987.

HUSSAIN, N.; JANI, P. U.; FLORENCE, A. T. Enhanced oral uptake of tomato lectin-conjugated nanoparticles in the rat. **Pharma. Res.**, v. 14, p. 613-618, 1997.

KILPATRICK, D. C.; GREEN, C. Lectins as blood typing reagents. **Adv. Lectin Res.**, v. 5, p. 51-94, 1992.

KOCOUREK, J.; HOREJSI, V. Defining a lectin. **Nature**, v. 290, p. 188, 1981.

LIENER, I. L.; SHARON, N.; GOLDSTEIN, I. J. **The Lectins: properties, functions and applications in biology and medicine**. New York: Academic Press, 1986. 600p p.

LIOI, L. GALASSO, I., SANTANTONIO, M., LANAVE, C., BOLLINI, R., & SPARVOLI, F. Lectin gene sequences and species relationships among cultivated legumes. **Gen. Res. Crop Evol.**, v. 53, p. 1615-1623, 2006.

LORENZI, H. **Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas do Brasil**. Nova Odessa: Instituto Plantarum, v. 2, 2002.

MAKELA, O. Studies in hemagglutinins of leguminoseae seeds. **Ann. Med. Exp. Biol. Fenn.**, v. 35, p. 1-156, 1957. ISSN 11.

MARTINEZ, C. R. CHANWAY, C. P., ALBERTINI, A. V., FIGUEIREDO, M. W., SAMPAIO, A. H., CASTELLON, R. R. *et al.* The interaction of *Vatairea macrocarpa* and *Rhizobium tropici*: net H<sup>+</sup> efflux stimulus and alteration of extracellular Na<sup>+</sup> concentration. **FEMS microbiology Letters**, v. 238, p. 17-22, 2004.

MARTINS, A. M. C. MONTEIRO, A. M., HAVT, A., BARBOSA, P. S., SOARES, T. F., EVANGELISTA, J. S. *et al.* Renal effects induced by the lectin from *Vatairea macrocarpa* seeds. **Journal of Pharmacy and Pharmacology**, v. 57, p. 1329-1334, 2005. ISSN 10.

MIN, W.; JONES, D. H. Stability and detection of recombinant pre-pro-concanavalin A after cytoplasmic expression in *Escherichia coli*. **FEBS Letters**, v. 301, p. 315-318, 1992.

MONTEIRO-MOREIRA, A. C. O. **Caracterização estrutural de três lectinas apresentando especificidade por açúcar distintas, isoladas de sementes de fruta-pão (*Artocarpus incisa* L.)**. Fortaleza: Universidade Federal do Ceará, 2002.

MOREIRA, R. A. CASTELO-BRANCO, C. C., MONTEIRO, A. C., TAVARES, R. O., & BELTRAMINI, L. M. Isolation and partial characterization of a lectin from *Artocarpus incisa* L. seeds. **Phytochemistry**, v. 46, p. 139-144, 1998.

MURDOCK, L. L.; SHADE, R. E. Lectins and protease inhibitors as plant defense againsts insects. **J. Agric. Food Chem.**, v. 50, p. 6605-6611, 2002.

MURRAY, M. G.; THOMPSON, W. F. Rapid isolation of high molecular weight plant DNA. **Nucleic Acids Research**, p. 4321-4325, 1980.

NEWMAN, D. LEHR, C. M., LENHOF, H. P., & KOHLBACHER, O. Computational modeling of the sugar-lectin interaction. **Adv. Drug Deliver. Rev.**, v. 56, p. 437-457, 2004.

NOGUEIRA, N. A. P. **Purificação e caracterização biológica da lectina recombinante de *Canavalia brasiliensis* (rConBr) produzida em *Escherichia coli***. Fortaleza: Universidade Federal do Ceará, 1999.

PEUMANS, W. J. ZHANG, W., BARRE, A., HOULES ASTOUL, C., BELINT-KURTI, P., ROVIRA, P. *et al.* Fruit-specificity lectins from banana and plantain. **Planta**, v. 211, p. 546-554, 2000.

PEUMANS, W. J.; VAN DAMME, E. J. M. The role of lectins in plant defense. **Histochem. Journal**, v. 27, p. 253-271, 1995.

PEUMANS, W. J.; VAN DAMME, E. J. M. Prevalence, biological activity and genetic manipulation of lectins in foods. **Trends in Food Sci. Tech.**, v. 7, p. 132-138, 1996.

RAMOS, M. V. BOMFIM, L. R., CAVADA, B. S., ALENCAR, N. M., GRANGEIRO, T. B., & DEBRAY, H. Further characterization of the glycan-binding specificity of the seed lectin from *Vatairea macrocarpa* and its dependence of pH. **Protein Pep. Lett.**, v. 7, p. 241-248, 2000.

RENKONEN, K. O. Studies on hemagglutinins present in seeds of some representatives of Leguminosae. **Ann. Med. Exp. Biol. Fenn.**, v. 26, p. 66-72, 1948.

ROMANO, E. Extração de DNA de tecidos vegetais. In: BRASILEIRO, A. C. M.; CARNEIRO, V. T. C. **Manual de transformação genética de plantas**. Brasília: Embrapa, 1998. p. 163-177.

ROUX, K. H. Single-Step PCR optimization using touchdown and stepdown PCR programming. In: CHEN, B. Y.; JANES, H. W. **PCR Cloning Protocols**. Totowa: Humana Press Inc., v. 192, 2002.

RÜDIGER, H.; GABIUS, H. J. Plant lectins: occurrence, biochemistry, functions and applications. **Glycoconjugate Journal**, v. 18, p. 589-613, 2001.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS. **Molecular cloning. A laboratory manual**. Second edition. ed. [S.l.]: Cold Spring Harbor Laboratory Press, 1989.

SANTANA, M. A. **Isolamento, propriedades bioquímicas e estudos biológicos da lectina de sementes da *Macrotyloma axillare* (E. Meyer)**. Ouro Preto: Universidade Federal de Ouro Preto, 2004.

SHARMA, V.; SUROLIA, A. Analysis of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificities. **J. Mol. Biol.**, v. 267, p. 433-445, 1997.

SHARON, N. Carbohydrates. **Scientific American**, v. 243, p. 80-97, 1980.

SHARON, N.; LIS, H. Legume lectins - a large family of homologous proteins. **FASEB J.**, v. 4, p. 3198-3207, 1990.

SHARON, N.; LIS, H. **Lectins**. 2<sup>o</sup> Edição. ed. [S.l.]: Kluwer Academic Publishers, 2003.

SHARON, N.; LIS, H. History of lectins: from hemagglutinins to biological recognition molecules. **Glycobiology**, v. 14, p. 53-62, 2004. ISSN 11.

SINGH, R. S.; TIWARY, A. K.; KENNEDY, J. F. Lectins: Sources, Activities and Applications. **Critic. Rev. Biotech.**, v. 19, p. 145-178, 1999. ISSN 2.

SUMMER, J. B.; HOWELL, S. F. The identification of the hemagglutinin of the Jack bean with concanavalin A. **Journal of Bacteriology**, v. 32, p. 227-237, 1936.

TAMURA, K. DUDLEY, J., NEI, M., & KUMAR, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Molecular Biology and Evolution**, v. 24, p. 1596-1599, 2007.

TRINDADE, M. B. **Purificação, caracterização e estudos estruturais de duas novas lectinas ligantes de quitina das sementes do gênero Artocarpus**. São Carlos: Universidade de São Paulo, 2005.

VAN DAMME, E. J. M. BARRE, A., SMEETS, K., TORREKENS, S., VAN LEUVEN, F., ROUGÉ, P., *et al.* The Bark of *Robinia pseudoacacia* Contains a Complex Mixture of Lectins. **Plant Physiol**, v. 107, p. 833-843, 1995.

VAN DAMME, E. J. M. PEUMANS, W. J., BARRE, A., & ROUGÉ, P. Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. **Crit. Rev. Plant Sci.**, v. 17, p. 575-692, 1998. ISSN 6.

VAN DAMME, E. J. M. BARRE, A., ROUGÉ, P., & PEUMANS, W. Cytoplasmic/ nuclear plant lectins: a new story. **Trends in Plant Science**, v. 9, p. 484-489, 2004.

VAN DAMME, E. J. M.; PEUMANS, W. J. Progress in structural analysis of plant lectins through molecular cloning. In: VAN DRIESSCHE, E., et al. **Lectins: biochemistry, clinical biochemistry**. Hellerup: Textop, 1996. p. 17-27.

VOELKER, T. A.; HERMAN, E. M.; CHRISPEELS, M. J. "*In vitro*" mutated phytohemagglutinin genes expressed in tobacco seeds of glycans in protein targeting and stability. **The Plant Cell**, v. 1, p. 95-104, 1989.

WU, A. M.; SUGII, S. Coding and classification of D-galactose, 2-N-acetyl-D-galactosamine, and beta-D-Galp-[1-3(4)]beta-D-GlcpNAc specificities of applied lectins. **Carbohydrate Research**, v. 213, p. 127-143, 1991.

YAMAUCHI, D. NAKAMURA, K., ASAHI, T., & MINAMIKAWA, T. Nucleotide sequence of cDNA for concanavalin A from *Canavalia gladiata* seeds. **Plant Cell Physiol.**, v. 30, p. 147-150, 1989.

YOUNG, N. M. WATSON, D. C., YAGUCHI, M., ADAR, R., ARANGO, R., RODRIGUEZ-ARANGO, E. *et al.* C-terminal post-translational proteolysis of plants and their recombinant forms expressed in *Escherichia coli*: characterization of "ragged ends" by mass spectrometry. **J. Biol. Chem.**, v. 270, p. 2563-2570, 1995.

## ANEXOS

9.1. Alinhamento dos três *contigs* obtidos a partir de DNA genômico, comparado com a sequência da VML citada por CALVETE *et al* (1998). Os resíduos de *Vml* que diferem do *contig1* são mostrados em vermelho. As regiões conservadas estão mostradas nos retângulos e mutações silenciosas são mostradas em negrito. # (sítio de ligação a metais), \* (sítio de ligação a carboidratos), *Glc* (presença de N-glicano), *Rem* (Remoção de resíduo durante processamento proteolítico)

<i>contig3</i>	CCG	<b>ATT</b>	CAG	ATT	TGG	GAT	AAT	ACC	ACA	GGC	<b>AGG</b>	<b>GTG</b>	GCA	AGC	TTT	GCC
<i>contig2</i>	CCG	<b>ATA</b>	CAG	ATT	TGG	GAT	GAT	AGC	ACA	GCA	<b>AGC</b>	<b>TTT</b>	GCA	AGC	TTT	GCC
<i>contig1</i>	CCG	<b>ATT</b>	CAG	ATT	TGG	GAT	GAT	AGC	ACA	GCA	<b>AGC</b>	<b>TTT</b>	GCA	AGC	TTT	GCC
<i>contig3</i>	P	I	Q	I	W	D	N	T	T	G	R	V	A	S	F	A
<i>contig2</i>	P	I	Q	I	W	D	D	S	T	G	R	V	A	S	F	A
<i>contig1</i>	P	I	Q	I	W	D	D	S	T	D	R	V	A	S	F	A
VML	P	I	<b>H</b>	I	W	D	D	S	T	D	R	V	A	S	F	A

70

ACT	TCC	TTC	TCC	TTT	GTT	GTG	GAA	GCA	CCT	<b>GAT</b>	GCT	TCC	CAA	ATT	GCA	GAT	<b>GGT</b>	ATT	<b>GCC</b>
ACT	TCC	TTC	TCC	TTT	GTT	GTG	GAA	GCA	CCT	<b>GAT</b>	GCT	TCC	CAA	ATT	GCA	GAT	<b>GGC</b>	ATT	<b>GCG</b>
ACT	TCC	TTC	TCC	TTT	GTT	GTG	GAA	GCA	CCT	<b>GAC</b>	GAA	TCC	AAA	ACT	GCA	GAT	<b>GGT</b>	ATT	<b>GCC</b>
T	S	F	S	F	V	V	E	A	P	D	A	S	Q	I	A	D	G	I	A
T	S	F	S	F	V	V	E	A	P	D	A	S	K	T	A	D	G	I	A
T	S	F	S	F	V	V	E	A	P	D	E	S	K	T	A	D	G	I	A
T	S	F	S	F	V	V	E	A	P	D	E	S	K	T	A	D	G	I	A

90

TTC	TTT	CTT	GCA	<b>CCA</b>	CCA	GAT	ACT	CAA	ACC	GAA	TCA	TGT	GGG	<b>GGA</b>	TAC	CTT	GGA	CTT	TTC
TTC	TTT	CTT	GCA	<b>CCA</b>	CCA	GCT	ACT	CAA	CCC	CAA	GGT	CCT	GGG	<b>GGT</b>	CTC	CTT	GGA	CTT	TTC
TTC	TTT	CTT	GCA	<b>CCT</b>	CCA	GAT	ACT	CAA	CCC	CAA	AAA	AAT	GGG	<b>GGT</b>	TTC	CTT	GGA	CTT	TTC
F	F	L	A	P	P	D	T	Q	T	E	S	C	G	G	Y	L	G	L	F
F	F	L	A	P	P	A	T	Q	P	Q	G	P	G	G	L	L	G	L	F
F	F	L	A	P	P	D	T	Q	P	Q	K	N	G	G	F	L	G	L	F
F	F	L	A	P	P	D	T	Q	P	Q	K	<b>D</b>	G	G	F	L	G	L	F

110

<i>Glc</i>	CGC	AAC	<b>AGT</b>	AGT	GAC	GAC	AAT	TAC	<b>AAT</b>	AAA	TCC	AAC	CAA	ACT	GTT	GCA	GTG	GAA	TTT	<b>GAT</b>
	CCT	GAC	<b>AGT</b>	GGT				TAC	<b>AAT</b>	GAC	TCC	AAC	CAA	ATT	GTT	GCA	GTG	GAA	TTT	<b>GAC</b>
	AAT	GAT	<b>AGT</b>					TAC	<b>AAT</b>	AAA	TCC	ATC	CAA	ACT	GTT	GCA	GTG	GAA	TTT	<b>GAT</b>
	R	N	S	S	D	D	N	Y	N	K	S	N	K	T	V	A	V	E	F	D
	P	D	S	G				Y	N	D	S	N	Q	I	V	A	V	E	F	D
	N	D	S					N	K	S	I	Q	T	V	A	V	E	F	D	
	N	D	S					N	K	S	I	Q	T	V	A	V	E	F	D	

125

	ACT	TAC	TCT	AAC	GAC	TGG	GAT	CCA	AGT	ATT	AAA	CAT	ATT	GGA	ATT	GAT	<b>GTC</b>	AAC	TCT	ATT
	ACT	TAC	CCT	GAC	GCA	TGG	GAT	CCA	AGT	TAT	ATA	CAT	ATT	GGA	ATT	GAT	<b>GTC</b>	AAC	TCT	ATT
	ACT	TTT	TCT	AAC	ACT	TGG	GAT	CCA	AGT	GCT	AGA	CAT	ATT	GGA	ATT	AAT	<b>GTT</b>	AAC	TCT	ATT
	T	Y	S	N	D	W	D	P	S	I	K	H	I	G	I	D	V	N	S	I
	T	Y	P	D	A	W	D	P	S	Y	I	H	I	G	I	D	V	N	S	I
	T	F	S	N	T	W	D	P	S	A	R	H	I	G	I	N	V	N	S	I
	T	F	S	N	T	W	D	P	S	A	R	H	I	G	I	N	V	N	S	I

145

	GTG	<b>TCA</b>	AAC	AAT	AAT	GTG	ATG	TGG	GTT	TGG	GAG	AAT	GGG	GAA	GTA	ACC	AAT	GTT	TAC	ATA
	GTG	<b>TCA</b>	ACA	AAA	ATT	GTG	AAA	TGG	GTT	TGG	GAG	AAT	GGG	AAA	GTA	GCC	AAT	GCT	CTC	ATA
	GAG	<b>TCT</b>	ATG	AAA	TAT	GTG	AAA	TGG	GGT	TGG	GAG	AAT	GGG	GAA	GTA	GCC	AAT	GTT	TAC	ATA
	V	S	K	N	N	V	M	W	V	W	E	N	G	E	V	T	N	V	Y	I
	V	S	T	K	N	V	K	W	V	W	E	N	G	K	V	A	N	A	L	I
	E	S	M	K	Y	V	K	W	G	W	E	N	G	E	V	A	N	V	Y	I
	E	S	M	K	Y	V	K	W	G	W	E	N	G	<b>K</b>	V	A	N	V	Y	I

165

AGC	TAT	AAA	GCT	TCC	ACC	AAA	<b>ACT</b>	TTA	ACT	GCC	TCT	<b>CTA</b>	ACT	TAC	CCT	<b>TCA</b>	AGT	GCA	ACA
AGC	TAT	GTA	GCT	TCC	ACC	AAA	<b>ACC</b>	TTA	ACT	GTC	TCT	<b>CTA</b>	ACT	TAG	CCT	<b>TCA</b>	AGT	GAC	ACA
AGC	TAT	CAA	GCT	TCC	ACC	AAA	<b>ACA</b>	TTA	ACT	GCC	TCT	<b>TTA</b>	ACT	TAC	CCT	<b>TCG</b>	AAT	GCA	ACA
S	Y	K	A	S	T	K	T	L	T	A	S	L	T	Y	P	S	S	A	T
S	Y	L	A	S	T	K	T	L	T	V	S	L	T	*	P	S	S	D	T
S	Y	Q	A	S	T	K	T	L	T	A	S	L	T	Y	P	S	N	A	T
S	Y	<b>E</b>	A	S	T	K	T	L	T	A	S	L	T	Y	P	S	N	A	T

185

AGT	TAT	ATC	GTC	TCT	GCA	AAT	GTG	GAT	TTG	AAG	<b>AGT</b>	<b>GCT</b>	CTT	<b>CCG</b>	GAA	TGG	<b>GTA</b>	AGA	GTT
AGT	TAT	ATC	CTC	TCT	GCA	AAT	GTG	GAT	TTG	AAG	<b>GAA</b>	<b>ATT</b>	CTT	<b>CCT</b>	GAA	TGG	<b>GTA</b>	AGG	GTT
AGT	TAT	ATC	GTC	TCT	GCA	AAT	GTG	GAT	TTG	AAG	<b>AGT</b>	<b>GCT</b>	CTT	<b>CCT</b>	GAA	TGG	<b>GTG</b>	AGA	GTT
S	Y	I	V	S	A	N	V	D	L	K	S	A	L	P	E	W	V	R	V
S	Y	I	L	S	A	Y	V	D	L	K	S	A	L	P	E	W	V	R	V
S	Y	I	V	S	A	N	V	D	L	K	S	A	L	P	E	W	V	R	V
S	Y	I	V	S	A	N	V	D	L	K	S	A	L	P	E	W	V	R	V

205

## 9.2. Alinhamento de proteínas permite ver conservação de aminoácidos no peptídeo sinal e nas sequências dos *primers*

```

          10          20          30          40          50
CLLRP Cladrasti  -ANSNSRPHLLQTQKPFVSVLVAISITFYLLLNKVNSEALSFTFTKFVS
RPAs2 Robinia p  ---KAMASYKFKTQNSFPLLSISF-FFLLLNKVNSTGSLSFSPKFAP
SJAbg Sophora j  -----ISITFFLLLNKVNSAEILSFSPKFVS
SJAsg Sophora j  MATSNSRPHLLQTHKPFVSVLVAISITFYLLLNKVNSAEILSFSPKFAS
VML Vatairea ma  -----SEVVSFSFTKFNP

          60          70          80          90          100
CLLRP Cladrasti  NQDELLLQGDALVSSKGELQLTRVENGQPIPHSVGRALYSDPVHIWDSST
RPAs2 Robinia p  NQPYLIFQRDALVTSTGVLQLTNVVNGVPPRRSIGRALYAAPFQIWDNTT
SJAbg Sophora j  NQDELLLQGDALVSSEGELQLTTVENGVPVWNSTGRALYAAPVHIWDNST
SJAsg Sophora j  NQDELLLQGDALVSSKGELQLTTVENGVPVWNSTGRALYAAPVHIWDKST
VML Vatairea ma  NPKDIILQGDALVTSKGLQLTKVKDGKPDHSLGRALYAAPVHIWDDST

          110         120         130         140         150
CLLRP Cladrasti  GSVASFVTSFTFVVEAPNENKTADGIAFFLAPPDTQVQSLGG-FLGLFNS
RPAs2 Robinia p  GNVASFVTSFSFIIQAPNPATTADGLAFFLAPVDTQPGDLGG-MLGIFKD
SJAbg Sophora j  GRVASFATSFSFVVKAPVASKSADGIAFFLAPLNNQIHGAGGGYGLFNS
SJAsg Sophora j  GRVASFATSFSFVVKAPVASKSADGIAFFLAPPNNQIQGGGGHGLFHS
VML Vatairea ma  DRVASFATSFSFVVEAPDESKTADGIAFFLAPPDTQPQKGG-FLGLFND

          160         170         180         190         200
CLLRP Cladrasti  SVYNSSNQILAVEFDTFNS-SWDPTARHIGIDVNSIESTRTATWGRNGE
RPAs2 Robinia p  GSYNKSQIVAVEFDTFNSIHFDPKGRHMGINVNSIVSVKTVPNWNTNGE
SJAbg Sophora j  SSSSSYQIVAVEFDTHN-AWDPNTRHIGIDVNSVSKTKTVTGWENGE
SJAsg Sophora j  SGYNSSYQIIAVDFDTHIN-AWDPNTRHIGIDVNSINSTKTVTGWQNGE
VML Vatairea ma  S--NKSIQTVAVEFDTFNS-TWDP SARHIGINVNSIESMKYVVKWGWENK

          210         220         230         240         250
CLLRP Cladrasti  VAIVLITYVAPAETLIASLTYPSSQTSYILSAAVDLKSIIP EWVRVGFSA
RPAs2 Robinia p  VANVFISYEASTKSLNASLVPSLETSEFIHAIVDVKDVIPEWVRVGFSA
SJAbg Sophora j  VANVLITYQAATEMLTVSLTYPSSQTSYILSAAVDLKSIIP EWVRVGFSA
SJAsg Sophora j  VANVLISYQAATETLTVSLTYPSSQTSYILSAAVDLKSIIP EWVRVGFSA
VML Vatairea ma  VANVYISYEASTKTLTASLTYPSSQTSYILSAAVDLKSIIP EWVRVGFSA

          260         270         280         290
CLLRP Cladrasti  ATGRSAGYVETHDVLVSWFTSTLETGNSGAKQNNHSLASYALI
RPAs2 Robinia p  TTGIDTGYVQTNVLSWFSFNLPGGNSVASVKNAGLSTYAA-
SJAbg Sophora j  TTGLTTQYVETNDVLSWFTSTLETSDCGAEDNNVHSLASYAFI
SJAsg Sophora j  ATGLTTQYVETHDVLVSWFTSTLETGDCGAKDDNVHSLVSYAFI
VML Vatairea ma  TSGLSRDHVETHDVLVSWFTSTLQAP---SDDSN-----

```

### 9.3. Alinhamento das sequências de DNA de lectinas homólogas à VML. As regiões conservadas ajudaram na construção de *primers* semi-degenerados.

```

          10          20          30          40          50
RPA2 Robinia p -----GCTATGGCTCCTACAAGTTCAAAACCCAAA
                      A M A S Y K F K T Q N
CLLRP Cladrasti -----GCTAACTCCAACCTCAAGACCACCTCTTGCAAACCCAGAA
                      A N S N S R P H L L Q T Q K
SJAbg Sophora j -----
SJAsg Sophora j AAAGCCATGGCTACCTCCAACCTCAAGACCACCTGTTGCAAACCCATAA
                      K A M A T S N S R P H L L Q T H K

          60          70          80          90          100
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|
CTCCTTCCCTCTTCTCCTATCCATATCCTT---TTTCTTCTCTTGTTAC
                      S F P L L L S I S F F F L L L
CLLRP Cladrasti ACCATTCTCTGTTGTTCTAGCCATCTCCATAACTTTCTACCTCTTGCTAC
                      P F S V V L A I S I T F Y L L L
SJAbg Sophora j -----TCCATAACTTTCTTCTCTTGCTAC
                      S I T F F L L L
SJAsg Sophora j GCCATTCTCTGTTGTTTGTAGCTATCTCCATAACTTTCTTCTCTTGCTAC
                      P F S V V L A I S I T F F L L L

          110          120          130          140          150
RPA2 Robinia p TCAACAAGGTGAATTCAACTGGATCTCTCTCCTTTTCTTTCCCAAGTTT
                      L N K V N S T G S L S F S F P K F
CLLRP Cladrasti TCAACAAGGTGAACCTCAGAAGAAGCCCTTTCTTTTACCTTCCAAAAGTTC
                      L N K V N S E E A L S F T F T K F
SJAbg Sophora j TCAACAAGGTGAACCTCAGCAGAAATCCTTTCTTTTAGCTTCCCAAGTTC
                      L N K V N S A E I L S F S F P K F
SJAsg Sophora j TCAACAAGGTGAACCTCAGCAGAAATCCTTTCTTTTAGCTTCCCAAGTTC
                      L N K V N S A E I L S F S F P K F

          160          170          180          190          200
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|
GCGCCTAACCAACCATATCTGATCTTCCAACGTGATGCCCTTGTGACATC
                      A P N Q P Y L I F Q R D A L V T S
CLLRP Cladrasti GTGTCAAACCAAGACGAACCTACTCCTCCAAGGCGATGCCCTGGTTTCATC
                      V S N Q D E L L L Q G D A L V S S
SJAbg Sophora j GTGTCAAACCAAGAGGATCTACTCCTCCAAGGTGATGCCTTGGTTTCATC
                      V S N Q E D L L L Q G D A L V S S
SJAsg Sophora j GCGTCAAACCAAGAGGATCTACTCCTCCAAGGTGATGCCTTGGTTTCATC
                      A S N Q E D L L L Q G D A L V S S

          210          220          230          240          250
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|
AACAGGGGTGTTACAACCTACCAACGTAGTTAATGGGGTACCACCGCGTC
                      T G V L Q L T N V V N G V P P R
CLLRP Cladrasti AAAAGGTGAGTTACAACCTACCAGGGTAGAAAATGGTCAACCAATCCCGC
                      K G E L Q L T R V E N G Q P I P
SJAbg Sophora j AGAAGGGGAGTTACAACCTACCACGGTAGAAAATGGTGTACCAGTCTGGA
                      E G E L Q L T T V E N G V P V W
SJAsg Sophora j AAAAGGGGAGTTACAACCTACCACGGTAGAAAATGGTGTACCAATCTGGA
                      K G E L Q L T T V E N G V P I W

          260          270          280          290          300
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|
GGTCAATGGTAGAGCTCTATATGCTGCCCTTTCCAAATCTGGGATAAC
                      R S I G R A L Y A A P F Q I W D N
CLLRP Cladrasti ACTCTGTTGGTCGAGCTCTATATCCGATCCTGTGCATATTTGGGATAGC
                      H S V G R A L Y S D P V H I W D S
SJAbg Sophora j ACTCTACTGGTCGAGCTCTATATATGCTCCTGTGCATATTTGGGATAAC
                      N S T G R A L Y Y A P V H I W D N
SJAsg Sophora j ACTCTACTGGTCGAGCTCTATATATGCTCCTGTGCATATTTGGGATAAG
                      N S T G R A L Y Y A P V H I W D K

          310          320          330          340          350
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|
ACCACAGGCAACGTGGCTAGCTTTGTCACTTCTTCTCTTATCATTC

```



```

        660          670          680          690          700
    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
    N A S L V Y P S L E T S F I I H
CLLRP Cladrasti AATTGCTTCTCTAACTTATCCTTCGAGTCAAACAAGTTATATCCTCTCTG
                I A S L T Y P S S Q T S Y I L S
SJAbg Sophora j AACTGTTTCTCTAACTTATCCTTCGAATCAAACAAGTTATATCCTCTCTG
                T V S L T Y P S N Q T S Y I L S
SJAsg Sophora j AACTGTTTCTCTAACTTATCCTTCGAGTCAAACAAGTTATATCCTCTCTG
                T V S L T Y P S S Q T S Y I L S

        710          720          730          740          750
    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
    CTATTGTGGATGTGAAGGATGTTCTTCCGGAGTGGGTAAAGATTGGTTTC
RPAs2 Robinia p
                A I V D V K D V L P E W V R F G F
CLLRP Cladrasti CAGCTGTTGATTTGAAGAGTATTCTTCCGGAGTGGGTAGAGTGGTTTC
                A A V D L K S I L P E W V R V G F
SJAbg Sophora j CAGCTGTTGATTTGAAGAGTATCCTAACAGAGTGGGTAGAGTGGTTTC
                A A V D L K S I L P E W V R V G F
SJAsg Sophora j CAGCTGTTGATTTGAAGAGTATTCTTCCGGAGTGGGTAGAGTGGTTTC
                A A V D L K S I L P E W V R V G F

        760          770          780          790          800
    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
    TCAGCTACCACAGGAATAGATACAGGCTACGTTCAAACAATGATGTTCT
RPAs2 Robinia p
                S A T T G I D T G Y V Q T N D V L
CLLRP Cladrasti TCTGCTGCCACAGGACGGTCTGCAGGCTACGTTGAAACACATGATGTTCT
                S A A T G R S A G Y V E T H D V L
SJAbg Sophora j ACTGCTACCACAGGACTGACTACTCAATACGTTGAAACAATGATGTTCT
                T A T T G L T T Q Y V E T N D V L
SJAsg Sophora j ACTGCTGCCACAGGACTGACTACTCAATACGTTGAAACACATGATGTTCT
                T A A T G L T T Q Y V E T H D V L

        810          820          830          840          850
    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
    CTCCTGGTCTTTCGAGTCAAACCTGCCAGGTGGTAACAGTGTGCTTCGG
RPAs2 Robinia p
                S W S F E S N L P G G N S V A S
CLLRP Cladrasti TTCCTGGTCTTTCACCTCAACCTTGAAACCGGCAACAGTGGTGCCTAAGC
                S W S F T S T L E T G N S G A K
SJAbg Sophora j TTCCTGGTCTTTCACCTCAACCTTGAAACCGGCAACAGTGGTGCCTAAGC
                S W S F T S T L E T S D C G A E
SJAsg Sophora j TTCCTGGTCTTTCACCTCAACCTTGAAACCGGCAACAGTGGTGCCTAAGC
                S W S F T S T L E T G D C G A K

        860          870          880          890          900
    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
    TGAAGAACGCGGGTCTTCAACCTATGCTGCATGATTCATGATGATGTTTCG
RPAs2 Robinia p
                V K N A G L S T Y A A * F H V V I
CLLRP Cladrasti AGAACACGCGCATCTTGCAAGTTATGCAT--TGATCTGATGGATTTTCG
                Q N N A H L A S Y A L I * W I F
SJAbg Sophora j ATAACAACGTGCATCTTGCAAGTTATGCAT--TCATCTGATGGACTTTCG
                D N N V H L A S Y A F I * W T F
SJAsg Sophora j ATGACAACGTGCATCTTGTAAGTTATGCAT--TCATCTGATGGACTTTCG
                D D N V H L V S Y A F I * W T F

        910          920          930          940          950
    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
    GGCG--ACAGAAGACAACGACAATAAGACCTGCAATAATGGTTGTTATCG
RPAs2 Robinia p
                G D R R Q R Q * D L Q * W L L S
CLLRP Cladrasti TACG-ACTAGTTATAAACTACAATAAAATTGAC--CTGTAATAATGGTCA
                V R L V I N Y N K I D L * * W S
SJAbg Sophora j TACGGTCTAGTTATAAACTACAATAAAATTGAC--CTGTGATAATGGTCA
                V R S S Y K L Q * N * P V I M V
SJAsg Sophora j TACGGACTAGTTATAAACTACAATAAAATTGAC--CTGTAATAATGGTCA
                V R T S Y K L Q * N * P V I M V

        960          970          980          990          1000
    .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
    TGTTGTATCAACTGTCCAGTCCAAAACATATAAAATCCTGTAATAGTTTT
RPAs2 Robinia p

```

```

          960          970          980          990          1000
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
TGTTGTATCAACTGTCACAGTCCAAAACATAAAATCCTGTAATAGTTTT
C C I N C H S P K L * N P V I V L
CLLRP Cladrasti TGTT--GGCAGCGGTCATACTGCTGAGCTGTAAAATATTGTAATGGT---
C W Q R S Y C * A V K Y C N G
SJAbg Sophora j CGTT--AGCAGCGGTCATATAGCCAGCTGTAAAATATTGTAATGGTTTT
T L A A V I * P S C K I L * W F
SJAsg Sophora j TGTT--AGCAGCGGTCAAAACGCTCATCTGTAAAATAGTGTAAATGGTTTT
M L A A V K L L I C K I V * W F

          1010          1020          1030          1040          1050
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
ATTAGTGACTGTAGCTAA--AATAAATAAGATCGGTCACCTTATGTTGTGG
L V T V A K I N K I G H L C C G
CLLRP Cladrasti -----ATGTAATAATATCGC---AATGGTTTCTCG
M * * Y R N G F S
SJAbg Sophora j ATCAGTTGTTGCTTCCAGAAATCATGTAATATCGCTTAAATAAATCTTG
Y Q L L L P E I M * Y R L N K F L
SJAsg Sophora j ATCAGTGGTTGCTTCAAGA-ATGATGTAATGTGCG---AATGGTTTCTCG
Y Q W L L Q E * C N V A M V S R

          1060          1070          1080          1090          1100
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
TTGCTGA-----
C *
CLLRP Cladrasti TA-----
SJAbg Sophora j TCGTAATAGAAAAGATTTAATAAACTAAAATAAAATGACCTGAAAAAT
S * * K R F N K L K I K L T * K M
SJAsg Sophora j TATGTGTATATATTGAT-----
M C I Y *

          1110          1120          1130          1140          1150
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
-----
CLLRP Cladrasti -----
SJAbg Sophora j GGTCATACTGCTAAGCTGTAAAATACTGTACGGTTTTATCAGTGGTTACT
V I L L S C K I L Y G F I S G Y
SJAsg Sophora j -----

          1160          1170          1180          1190          1200
RPA2 Robinia p .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
-----
CLLRP Cladrasti -----
SJAbg Sophora j TCCAGAATGATGTAATTTTCGCAATGGTTTCTCGTGTATATTGATATAA
F Q N D V I S Q W F L V C I L I *
SJAsg Sophora j -----

          1210          1220
RPA2 Robinia p .....|.....|.....|.....|.....|
-----
CLLRP Cladrasti -----
SJAbg Sophora j AAATAACGCTTTTAGATCAAAAA
K * R F * I K
SJAsg Sophora j -----

```