

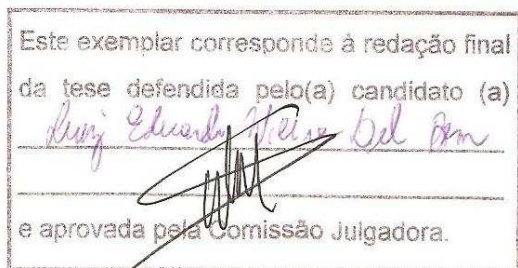
UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



**Luiz Eduardo Vieira Del Bem**

**“Análise filogenética das enzimas hidrolíticas de Xiloglucano no reino Viridiplantae e construção de bibliotecas de cDNA de Jatobá (*Hymenaea courbaril*)”**



Tese apresentada ao Instituto de Biologia para obtenção do Título de Mestre em Genética e Biologia Molecular, na área de Genética Vegetal e Melhoramento.

**Orientador: Prof. Dr. Michel Vincentz**

Campinas, 2008

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

D376a

Del Bem, Luiz Eduardo Vieira  
Análise filogenética das enzimas hidrolíticas de Xiloglucano no reino Viridiplantae e construção de bibliotecas de cDNA de Jatobá (*Hymenaea courbaril*) / Luiz Eduardo Vieira Del Bem. – Campinas, SP: [s.n.], 2008.

Orientador: Michel Vincentz.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Xiloglucano. 2. Glicosil hidrolases. 3. *Hymenaea courbaril*. 4. Evolução molecular. I. Vincentz, Michel. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

(rcdt/ib)

**Título em inglês:** Phylogenetic analysis of Xyloglucan's hydrolytic enzymes in the Viridiplantae kingdom and construction of cDNA libraries from Jatobá (*Hymenae courbaril*).

**Palavras-chave em inglês:** Xyloglucan; Glycosyl dydrolases; *Hymenaea courbaril*; Molecular evolution.

**Área de concentração:** Genética Vegetal e Melhoramento.

**Titulação:** Mestre em Genética e Biologia Molecular.

**Banca examinadora:** Michel Vincentz, Anete Pereira de Souza, Marcos Buckeridge.

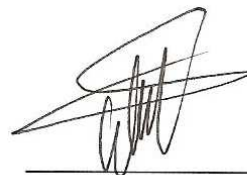
**Data da defesa:** 25/07/2008.

**Programa de Pós-Graduação:** Genética e Biologia Molecular.

**Campinas, 30 de julho de 2008**

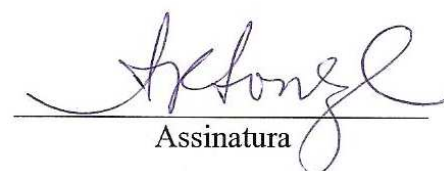
BANCA EXAMINADORA

Prof. Dr. Michel Vincentz (Orientador)



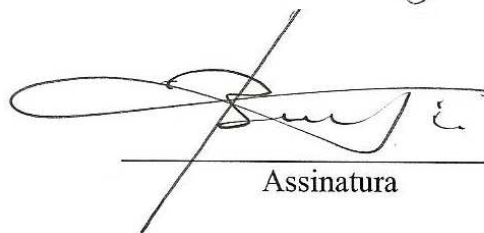
Assinatura

Profa. Dra. Anete Pereira de Souza



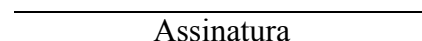
Assinatura

Prof. Dr. Marcos Buckeridge



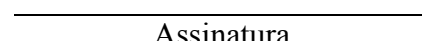
Assinatura

Prof. Dr. Jörg Kobarg



Assinatura

Dr. Vicente Eugênio de Rosa Júnior



Assinatura

**“A ignorância gera confiança com mais frequência do que o conhecimento: são aqueles que sabem pouco, e não aqueles que sabem muito, que tão positivamente afirmam que esse ou aquele problema jamais será resolvido pela ciência”.**

Charles Darwin (1809-1882)

**“O meu tempo chegará.”**

Gregor Mendel (1822-1884)

**Para meus pais, Luiz e Márcia, sem os quais nada disso seria possível.**

## **Agradecimentos**

Ao meu grande amigo e orientador, Michel Vincentz, por todas as conversas sobre evolução e genética (às vezes sobre camelos transcendentais ou outros temas menos ortodoxos), por tudo que me ensinou com grande dedicação e por ter me permitido realizar este e outros trabalhos científicos sob sua orientação.

Ao Prof. Dr. Marcos Buckeridge por ter me possibilitado trabalhar com Jatobá e pelos trabalhos prévios de seu grupo, que fazem parte da literatura básica sobre xiloglucano e o controle de sua degradação como reserva.

À Profa. Dra. Anete Pereira por aceitar fazer parte da banca e pelas incontáveis aulas de genética na graduação.

Aos amigos do Laboratório de Genética Molecular de Plantas/Arabidopsis: Diana, Amanda, Cleverson (Tutty), Gustavo (Johnny), Daniela, Mariane, Tammy e Sandra. E aos ex-membros Aline, Juarez, José Pedro e Luiz Gustavo (BH). Pela grade amizade e por todas as colaborações que desenvolvemos.

À Dra. Aline Brandão pela excelente colaboração neste e em outros trabalhos.

À Diana Branco pela colaboração que desenvolvemos em seu projeto.

Aos membros da pré-banca, Prof. Dr. Jörg Kobarg e Prof. Dr. Marcelo Dornellas pelas críticas e sugestões.

A todos os professores da graduação e pós do IB-Unicamp por terem me proporcionado uma excelente formação.

A todos os membros e ex-membros do CBMEG, por todos os anos de convívio e amizade.

À Profa. Dra. Marie-Anne Van Sluys, IB-USP, por ceder espaço em seu laboratório para que pudéssemos realizar os seqüenciamentos que aparecem neste trabalho. A todos os membros do Laboratório de Genética Molecular de Plantas da USP por terem me recebido muito bem ao longo de alguns meses de trabalho, em especial à Sílvia, por ter me ajudado muito nas minipreps e seqüenciamento.

Aos membros e ex-membros da República Beermania pela convivência e grande amizade, em todos esses anos vividos em Campinas: Júlio, Fábio, Guilherme, Maurício, Felipe, Ares, Zé, Rodolfo, César, Stênio e Rubens. E aos grandes amigos Gustavo Campos e Jairo Pakes, este último que cedeu sua república em minha estadia na USP, em São Paulo.

A todos os amigos (onças) da turma 02D de Biologia da Unicamp.

À FAPESP, pelo apoio financeiro.

À bióloga, amiga, namorada e companheira Maíra Nagai.

E em especial, à minha família. Meus pais, Luiz e Márcia, por terem me transmitido seus genes e sua cultura que fazem de mim o que sou. Por terem acreditado e apoiado minha carreira científica, desde que me conheço por gente, com toda a dedicação e amor que eu pude sonhar. E meu irmão Rodrigo, por ter me ensinado o significado da fraternidade, o que seguramente me fez uma pessoa melhor.

## ÍNDICE

LISTA DE FIGURAS .....	xi
LISTA DE TABELAS .....	xiv
ABREVIACÕES .....	xv
RESUMO .....	xvii
ABSTRACT .....	xix

### **1. Introdução**

1.1. O Jatobá .....	1
1.2. O Jatobá como modelo para estudos de impacto ambiental .....	3
1.3. O Jatobá como modelo para degradação de xiloglucano de reserva .....	6
1.4. Bibliotecas de cDNA .....	13
1.5. Grupos de ortólogos .....	13
1.6. Evolução molecular em famílias multigênicas .....	18
1.7. Filogenia .....	22
1.8. Objetivos .....	27
1.8.1. Para o modelo de impacto ambiental .....	27
1.8.2. Para o modelo de remobilização de xiloglucano .....	27

### **2. Material e Métodos**

2.1. Bibliotecas de cDNA	
2.1.1. Material biológico .....	28
2.1.2. Extração de RNA total .....	29
2.1.3. Purificação de RNA Poli A+ .....	30
2.1.4. Síntese de cDNA, LD-PCR e fracionamento de cDNA por tamanho .....	30
2.1.5. Clonagem via recombinase BP .....	32



2.1.6. Obtenção e transformação de células competentes .....	33
2.1.7. PCR de colônias .....	34
2.1.8. Construção das bibliotecas .....	35
2.1.9. Minipreparação plasmidial por lise alcalina .....	35
2.1.10. Seqüenciamento .....	36
2.1.11. Montagem das seqüências e anotação .....	36
2.2. Filogenias	
2.2.1. Material computacional .....	37
2.2.2. Obtenção das seqüências .....	37
2.2.3. Construção de bancos de seqüências .....	38
2.2.4. Estabelecimento da metodologia para obtenção de seqüências codificantes para as hidrolases que constituem o mecanismo de degradação de xiloglucano .....	39
2.2.5. Análise filogenética .....	41
2.2.6. Identificação de Possíveis Grupos de Ortólogos (PoGO) .....	43
<b>3. Resultados</b>	
3.1. Bibliotecas de cDNA de Jatobá .....	43
3.1.1. Otimização dos protocolos de extração de RNA Total e síntese de cDNA .....	43
3.1.2. Obtenção de RNA Poli A+, síntese de cDNA, LD-PCR e clonagem.....	45
3.1.3. Bibliotecas de cDNA .....	48
3.1.4. Seqüenciamento preliminar .....	50
3.1.5. Posição filogenética do Jatobá na família Leguminosae .....	56
3.2.1. Identificação de seqüências codificantes para as enzimas ativas sobre xiloglucano .....	58
3.2.2. Análise filogenética das seqüências de XTH e organização em possíveis grupos de ortólogos (PoGOs) .....	59
3.2.3. Análise filogenética das $\beta$ -Galactosidasas e organização em PoGOs.....	76
3.2.4. Análise filogenética das $\beta$ -Glucosidasas e organização em PoGOs .....	85

3.2.5. Análise filogenética das $\alpha$ -Xilosidases e $\alpha$ -Glucosidases e organização em PoGOs .....	91
3.2.6. Análise filogenética das $\alpha$ -Fucosidases ativas sobre xiloglucano e organização em PoGOs ...	95
<b>4. Conclusões e Perspectivas</b>	
4.1. Bibliotecas de cDNA .....	102
4.2. Análise filogenética das enzimas hidrolíticas de xiloglucano no reino Viridiplantae .....	103
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	106
Anexos: I – Blast to Fasta (código do programa, em JAVA®) .....	117
II – ESTs Translator Tool (código do programa, em JAVA®) .....	129
III – Árvore filogenética detalhada das XTHs .....	152

## LISTA DE FIGURAS

<b>Figura 1.</b> Representação esquemática de um xiloglucano hipotético. ....	8
<b>Figura 2.</b> Modelo para degradação do xiloglucano de reserva. ....	8
<b>Figura 3.</b> Vias possíveis na hidrólise completa de oligossacarídeos de xiloglucano de reserva. ....	9
<b>Figura 4.</b> Modelo para degradação do xiloglucano estrutural. ....	9
<b>Figura 5.</b> Modelo de extensão da parede celular. ....	11
<b>Figura 6.</b> Possíveis destinos de genes duplicados segundo uma trajetória temporal ....	15
<b>Figura 7.</b> Posição relativa da linhagem das plantas (Viridiplantae) dentro dos Eucariotos. ....	17
<b>Figura 8.</b> Representação das relações evolutivas de ortologia e paralogia entre genes. ....	17
<b>Figura 9.</b> Três modelos diferentes para evolução de famílias multigênicas. ....	20
<b>Figura 10.</b> Diagrama representando estrutura do rDNA no genoma de plantas vasculares. ....	20
<b>Figura 11.</b> Filogenia de seis seqüências com tamanho dos ramos conhecidos. ....	25
<b>Figura 12.</b> Mapa do vetor pDONR-222. ....	32
<b>Figura 13.</b> Esquema do protocolo bioinformático para construção de bancos não redundantes de seqüências de proteínas. ....	42
<b>Figura 14.</b> Análise da integridade dos RNAs totais de Cotilédone, Caule, Folhas e Raízes. ....	47

<b>Figura 15. 1</b> - Gel de agarose (1%) em condição desnaturante (6% de Formaldeído) mostrando a integridade dos Mix de RNA obtidos para cada tecido (ver Tabela 2). <b>2</b> – Gel de agarose (1%) mostrando o produto de amplificação do cDNA de Actina para os diferentes Mix de RNA. ....	47
<b>Figura 16.</b> cDNAs separado por tamanho em gel de agarose (2%) corado com Brometo de Etídio. ....	47
<b>Figura 17.</b> Estimativa do tamanho dos cDNA clonados na biblioteca de Caule. ....	49
<b>Figura 18.</b> Estimativa do tamanho dos cDNA clonados na biblioteca de Cotilédones. ....	49
<b>Figura 19.</b> Estimativa do tamanho dos cDNA clonados na biblioteca de Raízes. ....	49
<b>Figura 20.</b> Estimativa do tamanho dos cDNA clonados na biblioteca de Folha. ....	50
<b>Figura 21.</b> Distribuição relativa, em categorias funcionais, dos 82 ESTs obtidos da biblioteca de caule. ....	51
<b>Figura 22.</b> Distribuição relativa, em categorias funcionais, de 75 ESTs da biblioteca de raiz. ....	53
<b>Figura 23.</b> Cromatograma mostrando uma seqüência com <i>slippage</i> (repetição de C). ....	53
<b>Figura 24.</b> Distribuição relativa, em categorias funcionais, de 16 ESTs obtidos da biblioteca de folhas. ....	54
<b>Figura 25.</b> Distribuição da 198 seqüências obtidas ao longo de quatro categorias. ....	55
<b>Figura 26.</b> Distribuição relativa, por função, dos 150 ESTs codificantes obtidos. ....	55
<b>Figura 27.</b> Análise filogenética, por NJ baseada na seqüência de aminoácidos dos genes L10, S23 e Subunidade G do complexo NADH dehydrogenase. ....	57
<b>Figura 28.</b> Análise filogenética, por MP, baseada na seqüência de aminoácidos dos genes L10, S23 e Subunidade G do complexo NADH dehydrogenase. ....	57
<b>Figura 29.</b> Grupos de XTHs homólogos. ....	65

<b>Figura 30.</b> Árvore filogenética compactada mostrando os PoGOs das XTHs. ....	73
<b>Figura 31.</b> Alinhamento de seqüências do PoGO R com o conjunto completo de XTHs de Arabidopsis. ....	74
<b>Figura 32.</b> Alinhamento utilizado para gerar a árvore da Figura 33. ....	75
<b>Figura 33.</b> Árvore filogenética mostrando o agrupamento entre o Grupo VI (PoGO R) e as seqüências similares a XTHs em fungos. ....	75
<b>Figura 34.</b> Análise filogenética do Grupo II das $\beta$ -Galactosidases. ....	78
<b>Figura 35.</b> Árvore filogenética das $\beta$ -Galactosidases. ....	83
<b>Figura 36.</b> Árvore filogenética das $\beta$ -Galactosidases. ....	84
<b>Figura 37.</b> Árvore filogenética compactada das $\beta$ -Glucosidases específicas de plantas. ....	88
<b>Figura 38.</b> Árvore filogenética das $\beta$ -Glucosidases específicas de plantas. ....	89
<b>Figura 39.</b> Resultado do alinhamento por Blastp 2x2 entre as seqüências YP_001634696 ( <i>Chloroflexus aurantiacus</i> ) e CAA07070 ( <i>Tropaeolum majus</i> ). ....	90
<b>Figura 40.</b> Análise filogenética das $\alpha$ -Xilosidases e $\alpha$ -Glucosidases. ....	94
<b>Figura 41.</b> Análise filogenética por Máxima Parcimônia das seqüências de Bryophytas. ....	95
<b>Figura 42.</b> Possível grupo de ortólogos de ATFXG1. ....	96
<b>Figura 43.</b> Árvore filogenética dos 100 primeiros <i>hits</i> de Blastp e tBlastn de ATFXG1. ....	97

<b>Figura 44.</b> Árvore filogenética mostrando o grupo monofilético (PoGO B) de Embryophytas. ....	99
<b>Figura 45.</b> Resultado do alinhamento por Blastp 2x2 entre as seqüências NP_412708 ( <i>Bacillus halodurans</i> ) e BAF85832 ( <i>Lilium longiflorum</i> ). ....	100

## LISTA DE TABELAS

<b>Tabela 1.</b> Condições de crescimento das plântulas de Jatobá. ....	28
<b>Tabela 2.</b> Lista das espécies cujos genomas completos foram obtidos para análise. ....	37
<b>Tabela 3.</b> Bancos de ESTs utilizados. ....	39
<b>Tabela 4.</b> Número de seqüências, em cada organismo, para as diversas hidrolases. ....	59
<b>Tabela 5.</b> Classificação das XTHs de em grupos de genes homólogos e PoGOs. ....	66
<b>Tabela 6.</b> Classificação das $\beta$ -Galactosidases em Grupos de homólogos e PoGOs. ....	79
<b>Tabela 7.</b> Classificação das $\beta$ -Glucosidases em PoGOs. ....	87
<b>Tabela 8.</b> Agrupamento das $\alpha$ -Xilosidases e $\alpha$ -Glucosidases em PoGOs. ....	93
<b>Tabela 9.</b> Agrupamento das $\alpha$ -Fucosidases em PoGOs. ....	101

## LISTA DE ABREVIACOES

$\alpha$ -X	Alpha-Xilosidase
$\alpha$ -xyl	Alpha-Xilosidase
$\beta$ -G	Beta-Glucosidase
$\beta$ -Gal	Beta-Galactosidase
$\beta$ -gal	Beta-Galactosidase
$\beta$ -glc	Beta-Glucosidase
$\mu$ g	micrograma
$\mu$ L	microlitro
$\mu$ M	micromolar
At	<i>Arabidopsis thaliana</i>
BTF	<i>Blast To Fasta</i>
cDNA	<i>copied DNA</i>
CmR	<i>Chloramphenicol</i>
CNI	<i>Close neighbor-intercharge</i>
Cr	<i>Chlamydomonas reinhardtii</i>
DEPC	Dietil pirocarbonato
DNA	<i>Deoxyribonucleic acid</i>
dNTP	<i>Deoxynucleosides</i>
EST	<i>Expressed Sequence Tag</i>
ETTool	<i>ESTs Translation Tool</i>
Fuc	Fucose
Gal	Galactose
Glc	Glicose
Gm	<i>Glycine max</i>
kV	kilovolt
LB	<i>Lysogeny broth</i>
LD-PCR	<i>Long Distance-PCR</i>
M	Molar
min	minuto
mL	mililitro
mRNA	<i>Messenger RNA</i>
ME	<i>Minimum Evolution</i>
MHC	<i>Major histocompatibility complex</i>
MP	Mxima Parcimnia
MV	Mxima Verossimilhana
ng	nanograma
NJ	<i>Neighbor-Joining</i>
Os	<i>Oryza sativa</i>
PCR	<i>Polymerase Chain Reaction</i>
PGA	<i>Penta-galacturonic acids</i>
PoGO	<i>Possible Group of Orthologues</i>
Pp	<i>Physcomitrella patens</i>
Pt	<i>Populus trichocarpa</i>
RG I	<i>Rhamnogalacturonan I</i>
RNA	<i>Ribonucleic acid</i>

rRNA	<i>Ribosomal RNA</i>
RPM	Rotações por minuto
RT	<i>Reverse Transcriptase</i>
S	<i>Sum of tree</i>
Sb	<i>Sorghum bicolor</i>
Sm	<i>Selaginella moellendorffii</i>
UPGMA	<i>Unweighted pair-group method using arithmetic averages</i>
Vc	<i>Volvox carteri</i>
Vv	<i>Vitis vinifera</i>
XET	<i>Xyloglucan Endotransglycosylase</i>
XTH	<i>Xyloglucan Transglycosylase / Hydrolase</i>
Xyl	Xilose



## Resumo

**Introdução:** Os xiloglucanos são os polímeros de açúcar mais abundantes na hemicelulose da maioria das espécies de plantas terrestres, em especial nas eudicotiledôneas. Possuem papel estrutural na parede celular vegetal, interagindo com os filamentos de celulose, e podem ser utilizados como reserva em sementes de várias espécies de eudicotiledôneas, como o Jatobá (*Hymenaea courbaril*), onde correspondem a quase 50% do peso seco da semente. Este polímero é formado por uma cadeia central de  $\beta$ -glucano com ramificações que contêm xilose, galactose e fucose. O mecanismo de degradação deste polímero é realizado por cinco hidrolases: XTH,  $\beta$ -Galactosidase,  $\beta$ -Glucosidase,  $\alpha$ -Xilosidase e  $\alpha$ -Fucosidase. Estas enzimas são codificadas por genes que constituem famílias multigênicas nos genomas de plantas, e sua atividade na degradação seletiva de xiloglucano têm papel central na regulação do crescimento e morfogênese da célula vegetal.

O Jatobá (Leguminosae) é uma árvore tropical, nativa do Brasil, que vem sendo utilizada como modelo vegetal para estudos de impacto ambiental por efeito estufa e estresses abióticos oriundos do aquecimento global. Foi observado que mudas de Jatobá, crescidas numa atmosfera com 720 PPM de CO<sub>2</sub> (dobro da concentração atual), apresentam até 50% de aumento de biomassa aos 100 dias. O entendimento das respostas transcricionais desta planta, em resposta a estes estresses, pode levar a conclusões à cerca de como as florestas tropicais responderão ao aumento inexorável na concentração de CO<sub>2</sub>, num quadro de aquecimento global.

**Resultados:** Construimos bibliotecas de cDNA de folhas, caule, cotilédones e raízes de plântulas de 45 dias de Jatobá. Um seqüenciamento amostral dos ESTs levou à obtenção de 103 seqüências, parciais ou completas, de proteínas de Jatobá. São os primeiros dados de ESTs numa árvore tropical brasileira.

Análises filogenéticas das enzimas que constituem o mecanismo de degradação de xiloglucano foram conduzidas ao longo de 13 genomas completos e 27 bancos de ESTs de espécies dos mais diversos grupos no reino Viridiplantae. Isso nos permitiu organizar a diversidade destas cinco famílias multigênicas em possíveis grupos de ortólogos (PoGOs). As XTHs foram divididas em seis grupos de genes homólogos e 19 PoGOs. As  $\beta$ -Galactosidases foram divididas em dois grupos de genes homólogos e 10 PoGOs.

As  $\beta$ -Glucosidases foram divididas em dois grupos de genes homólogos e dois PoGOs. As  $\alpha$ -Xilosidase foram divididas em três PoGOs e as  $\alpha$ -Fucosidase em dois PoGOs não relacionados evolutivamente.

**Conclusões e Perspectivas:** As 103 seqüências peptídicas obtidas de Jatobá foram anotadas por comparação e serão disponibilizadas nos bancos de dados internacionais. A perspectiva de seqüenciar mais clones poderá levar à montagem do transcriptoma do Jatobá, algo inédito para uma árvore tropical.

Concluimos, com as análises filogenéticas, que as XTHs, que formam um grupo monofilético de genes em Streptophyta, surgiram antes da conquista do ambiente terrestre. Estes genes foram progressivamente amplificados ao longo da evolução das plantas terrestres, o que sugere um ganho progressivo de complexidade, que teve seu auge nas Angiospermas. Apresentamos evidências que podem unir evolutivamente as XTHs exclusivas de plantas a enzimas transglicosiladoras de cadeias de  $\beta$ -glucano em fungos, o que sugere uma origem comum do processo de transglicosilação de cadeias de  $\beta$ -glucano como mecanismo de controle do crescimento e formato celular em eucariotos com parede celular. As  $\beta$ -Galactosidases formam um grupo monofilético em Embryophytas com nove PoGOs, no entanto sua grande diversificação (seis PoGOs) ocorreu apenas em Angiospermas. As  $\beta$ -Glucosidases formam um grupo monofilético em Embryophytas, seqüências similares em bactérias fotossintetizantes podem sugerir uma origem no ancestral dos cloroplastos. As  $\alpha$ -Xilosidase, que são monofiléticas nas Spermatophytas, derivaram das  $\alpha$ -Glucosidases que se encontram dispersas entre todos os eucariotos, é um caso de neofuncionalização. Duas linhagens distintas evolutivamente de  $\alpha$ -Fucosidases foram encontradas, uma delas é monofilética em Embryophytas e a outra pertence a uma grande família multigênica (*GDSL-motif*) da qual pouco se sabe. Mostramos que o mecanismo completo (cinco hidrolases) de degradação de xiloglucano existia no ancestral comum das Spermatophytas (plantas com semente). Como perspectivas este trabalho permite a racionalização de estudos funcionais destas hidrolases o que, em longo prazo, pode contribuir com processos biotecnológicos que passem pela modificação seletiva da parede celular vegetal.

## Abstract

**Introduction:** Xyloglucans are the main hemicellulose in most of land plants, especially in eudicots. It is a structural compound of plant cell-wall that interacts with cellulose and can be used as seed's energy storage of many species, like Jatoba (*Hymenaea courbaril*). Xyloglucan structure is composed of a  $\beta$ -glucan backbone that it branched with xylose, galactose and fucose. Its degradation machinery is composed by five glycosyl hydrolases: XTH,  $\beta$ -Galactosidase,  $\beta$ -Glucosidase,  $\alpha$ -Xylosidase and  $\alpha$ -Fucosidase. These enzymes are codified by multigenic families in plant's genomes and it plays a central role in key processes like growth and morphogenesis of plant cells.

Jatoba (Leguminosae) is a tropical tree, native of Brazil. It's been used as a model tree in researches of plant's responses to stresses caused by global warming and high atmospheric CO<sub>2</sub> concentration. It was observed a 50% increase in biomass of a 100 days Jatoba seedling when grown in a 720 PPM of CO<sub>2</sub> atmosphere (two times bigger than today's atmospheric concentration). Understand the transcriptional responses to these stresses can lead to conclusions about how tropical forests will respond to high concentrations of CO<sub>2</sub> and global warming.

**Results:** We made cDNA libraries of leaves, stem, cotyledons and roots of 45 days seedlings of Jatoba. A preliminary sequencing of these libraries reveled 103 predict protein sequences (most partial sequences).

Phylogenetic analyses of xyloglucan hydrolytic enzymes were conducted using 13 completed genomes and 27 ESTs assemblies, from a wild range of taxonomic groups in the Viridiplantae kingdom. It allowed us to divide XTH's diversity of genes into six homology groups and 19 possible groups of orthologues (PoGOs).  $\beta$ -Galactosidases were divided into two groups of homologues and 10 PoGOs.  $\beta$ -Glucosidases were divided into two groups of homologues and two PoGOs.  $\alpha$ -Xylosidase were divided into three PoGOs and  $\alpha$ -Fucosidase into two PoGOs evolutionarily unrelated.

**Conclusions and Perspectives:** The 103 protein sequences of Jatoba were annotated by comparison to known proteins and will be deposited in international sequences assemblies. As a perspective, the sequencing of Jatoba ESTs will lead to the assembly of its transcriptome, something never done before in a tropical tree.

We concluded that XTHs are monophyletic group of genes in Streptophyta, what means they emerged before land conquest by plants. These genes were progressively amplified in land plants evolution, especially in Angiosperms, what suggests a progressive gain in complexity. We showed evidences of a possible evolutionary relation between plant's XTHs and fungus hydrolases/transglycosylases enzymes. It suggests a eukaryotic ancestral mechanism to control cell expansion and shape based in  $\beta$ -glucan transglycosylation and its interaction to cellulose (in plants) or chitin (in fungus). The  $\beta$ -Galactosidases are a monophyletic group in Embryophytas that were divided into nine PoGOs, six PoGOs only appeared in Angiosperms. The  $\beta$ -Glucosidases belongs to a monophyletic group in Embryophytas that has sequence similarity to bacterial proteins, especially ones from photosynthetic bacteria species. The  $\alpha$ -Xylosidases are a PoGO in Spermatophyta that probably emerged from  $\alpha$ -Glucosidases presents in all eukaryotes. It's probably a neofunctionalization process. Two evolutionary distinct lineages of  $\alpha$ -Fucosidases where found, one monophyletic in Embryophytas and another that belongs to the poorly understood multigenic family "GDSL-motif proteins". We showed that the complete machinery (all the five hydrolases) of Xyloglucan degradation already exists in Spermatophytas common ancestor. As a perspective, we expect to rationalize the functional characterization works among these multigenic families and to contribute in biotechnology processes that pass through cell-wall modification and selective control.

## 1. Introdução

### 1.1. O Jatobá (*Hymenaea courbaril*)

Jatobás são plantas da família Leguminosae pertencentes ao gênero *Hymenaea*, que é constituído por 17 espécies. O jatobá (*Hymenaea courbaril* L.) é uma Caesalpiniaceae (subfamília das leguminosas) encontrada desde o México, passando pela América Central, ocorrendo abundantemente na Amazônia, chegando até São Paulo. Aparece nas matas de terra firme, sobre solo argiloso e em certas várzeas altas; rara no campo e nas capoeiras (Loureiro *et al.*, 1979).

O gênero *Hymenaea*, de distribuição anfiatlântica, pertence à tribo Detarieae (Cynometreae). Apresentam árvores geralmente grandes (podendo superar os 25m), primariamente perenifólias, com troncos retos e cilíndricos de casca suave de cor cinza prateado, folhas alternas estipuladas, bifolioladas com pecíolo livre do lado interno, porém parcialmente coberto pela lâmina do lado externo, flores do tipo hypanthium com quatro sépalas, cinco pétalas, 10 estames e um pistilo, formando inflorescências do tipo panículo curto que se torna corimbosa quando madura (característica da seção *Hymenaea*). Frutos lenhosos do tipo vagem, indeiscentes, ovóides a oblongos, de cor que varia de marrom claro a escuro, quando maduros e verdes quando imaturos, com endocarpo farináceo cobrindo sementes ovóides a elipsóides de testa dura e lisa marrom a marrom escura. As células epidérmicas podem apresentar dois padrões em *Hymenaea*: células de paredes retas ou de paredes sinuosas. A germinação é epigea, sendo o primeiro par de folhas unifoliolado e oposto (eófilo), diferente das folhas subsequentes. É característica do gênero a presença do colar entre a base do hipocótilo e a raiz, junto ao solo. O número cromossômico diplóide é  $2n = 24$  para todas as espécies e variedades do gênero e não há evidência de hibridização ou poliploidia. Essa uniformidade no

número cromossômico é compatível com o padrão geral das leguminosas arbóreas tropicais e nem sempre fica restrita a espécies de um mesmo gênero, sendo também comum em nível de tribo (Atchison, 1951 e Heywood, 1971, apud Lee & Langenheim, 1975).

Economicamente, produtos valiosos das árvores do gênero *Hymenaea* são as resinas (já foi objeto de exportação) de cores amarelas e avermelhadas, usadas como cimento ou incenso, em esculturas e joalheria por índios pré-colombianos (Balser, 1960), e até, dissolvida em xilol, em preparo de lâminas permanentes para microscopia (Le Cointe, 1947). Conhecida como copal sul americano (Tschirch & Stock, 1936) é utilizada na fabricação de vernizes e com fins medicinais (Langenheim, 1966). A sua madeira, de excelente qualidade (madeira de lei), é usada em obras hidráulicas, carroçarias, postes, tonéis, dormentes, construções de todas as espécies, móveis, laminados, esteios, tacos, entre outras formas (Loureiro *et al.*, 1979). Índios amazônicos utilizam a casca, que pode chegar a quatro centímetros de espessura, na fabricação de canoas (Record & Mell, 1924; Lima, 1950; Sandwith, 1938). As folhas e pedaços de casca macerados em cachaça são utilizados com fins medicinais, e a polpa farinácea é comestível e utilizada no preparo de bolos.

As variedades de jatobá que ocorrem na Mata Atlântica formam árvores de até 20 metros de altura e podem durar centenas de anos. São espécies que desempenham suas funções ecológicas na fase final de formação da floresta (Lee & Langenheim, 1975). Por isso, caracteristicamente crescem lentamente no chão da floresta e à sombra de outras espécies. Este acúmulo lento e contínuo de biomassa por muitos anos faz com que o jatobá seja uma das espécies importantes no processo de seqüestro de carbono pelas florestas. Por apresentar ampla ocorrência e muitas espécies e variedades

diferentes em vários ambientes, o jatobá tem grande capacidade de adaptar-se a diferentes ecossistemas. Estudos genéticos de diferentes populações e genótipos adaptados a diferentes ecossistemas podem fornecer subsídios para entender as respostas das plantas às mudanças climáticas, representadas pelos diferentes ecossistemas onde ocorre.

## **1.2. O Jatobá como modelo para estudos de impacto ambiental**

O jatobá é uma espécie do estágio final de sucessão ecológica e passam grande parte da vida crescendo. Estudos fisiológicos e moleculares de plantas jovens de jatobá são de grande importância, não somente para compreender melhor como as florestas tropicais se formam, mas especificamente para avaliar como elas sequestram carbono.

Alterações na concentração atmosférica de dióxido de carbono (CO<sub>2</sub>) e metano têm elevado a temperatura global de forma muito relevante nos últimos 200 anos (Mann, *apud* Appel, 2005). Teorias mais recentes mostram o aumento desproporcional destes gases, contrariando as previsões de modelos estabelecidos, como tendo origem na mesma época em que o homem desenvolveu a agricultura e desflorestou áreas nativas para o plantio e a criação de animais. Sendo assim o homem teria iniciado o efeito de aquecimento do planeta há mais de oito mil anos atrás. No entanto todas as teorias concordam que o efeito de aquecimento global pelo efeito estufa provocado, principalmente, pelo CO<sub>2</sub> tem origem diretamente na atividade humana de modificação dos ambientes naturais (Ruddiman 2005).

É consenso que a concentração atmosférica de CO<sub>2</sub> vem aumentando, de forma extremamente rápida, desde a revolução industrial como consequência da queima de combustíveis fósseis e do desflorestamento. No início do século XX a concentração atmosférica de CO<sub>2</sub> aumentou de 280ppm para cerca de 370ppm em 2002. Entre 1980 e

2000, o incremento anual foi de 1,53PPM/ano, ou 0,45%, de modo bastante constante (Prentice *et al.*, 2001). O nível atual é o maior dos últimos 20 milhões de anos, e ainda neste século pode atingir níveis ainda maiores do que os estimados no início do Eoceno – 55 a 34 milhões de anos (Royer *et al.*, 2001), dependendo da magnitude da emissão de gases do efeito estufa devido às atividades humanas e aos mecanismos de retro-alimentação entre o sistema climático e a biosfera (Cox *et al.*, 2000).

Acredita-se que as mudanças climáticas que estão ocorrendo possam causar um impacto considerável sobre as florestas tropicais. Lewis *et al.* (2004) identificaram as possíveis alterações que podem afetar a dinâmica e a estrutura das florestas tropicais: (1) o aumento de temperatura afetaria diretamente a fotossíntese e a respiração, além de alterar a disponibilidade de nutrientes no solo por alterar a dinâmica da floresta; (2) o aumento da concentração de CO<sub>2</sub> atmosférica, relevante à permeabilidade da atmosfera à luz, poderia afetar tanto a eficiência no uso da luz como da água; (3) a deposição de nutrientes aumentaria, por exemplo, a entrada de nitrogênio; e (4) alterações no uso da terra poderiam levar à fragmentação de florestas e alterar as dinâmicas das populações vegetais e da própria floresta. Segundo Mann, (2005) já estamos no curso de 50 a 100 anos de aquecimento e vários séculos de elevação do nível do mar, simplesmente devido à quantidade de gases-estufa que já lançamos na atmosfera (Mann, *apud* Appel, 2005).

Nas plantas o carbono entra e sai por difusão, na forma de CO<sub>2</sub>, através dos estômatos presentes na epiderme das folhas. O CO<sub>2</sub> vai servir como matéria-prima de compostos orgânicos, durante a fotossíntese. Saindo, o CO<sub>2</sub> é um dos produtos finais da respiração. Já animais e fungos realizam apenas a respiração liberando o CO<sub>2</sub> na atmosfera, e obtêm o carbono de que precisam direto ou indiretamente através do processo de fixação de carbono realizado pelos organismos fotossintetizantes. O



carbono será devolvido indubitavelmente para a atmosfera através, além da respiração, da decomposição. Um dos principais componentes do ciclo global de carbono são as florestas tropicais que são responsáveis por 32-36% do seqüestro global de carbono e correspondem ao maior estoque de carbono orgânico na biosfera: entre 400 e 500 bilhões de toneladas de carbono, ou aproximadamente, dois terços da quantidade presente como dióxido de carbono na atmosfera (700 bilhões de toneladas). (Melillo *et al.* 1993; Field *et al.* 1998). Nosso conhecimento sobre a ciclagem de carbono nas florestas tropicais é muito limitado, sendo necessário, ainda mais estudos para a compreensão deste processo complexo e a determinação de como este ecossistema pode ser caracterizado como dreno ou fonte de carbono para a atmosfera, e como esta situação pode ser alterada nos próximos séculos (Clark 2004).

Foi observado que, em atmosfera enriquecida, mudas de Jatobá aumentaram a capacidade de assimilação de CO<sub>2</sub> (fotossíntese) em mais do que 60%. Medidas de crescimento ao longo dos primeiros 100 dias de vida da muda revelaram que havia alteração no crescimento da planta. Medições com folhas e caule das mudas de jatobá, revelaram que houve um aumento de cerca de 20% de celulose em resposta à atmosfera enriquecida com 720ppm de CO<sub>2</sub>. Além de uma duplicação na produção de carboidratos e aumento de até 50% na biomassa, sobretudo na área foliar e nas raízes, visto que, com essa idade, as plântulas ainda não produzem crescimento secundário (madeira). Todas essas alterações são apresentadas por plantas que cresceram em atmosfera enriquecida (720ppm de CO<sub>2</sub>). (Aidar *et al.* 2002).

A possibilidade de ensaios relativamente rápidos (de 45 a 100 dias) com plântulas de Jatobá, em câmaras de topo aberto onde é possível controlar umidade, temperatura e concentração de CO<sub>2</sub>, torna este sistema um bom modelo para obtenção

de dados, fisiológicos e moleculares, sobre como árvores tropicais de último estágio de sucessão ecológica respondem as mudanças climáticas.

### **1.3. O jatobá como modelo para remobilização de Xiloglucano de reserva**

O formato celular e a morfologia de todos os tecidos vegetais são conseqüências das expansões celulares e da divisão celular ao longo dos seus ciclos de vida. Modificações estruturais na parede celular são processos chave na regulação do crescimento e diferenciação celular. A parede celular primária exibe uma estrutura complexa e dinâmica com variabilidade espacial e temporal em sua composição e organização. A forma das células, seu tamanho e as adesões célula-célula são processos que passam pela ação coordenada de enzimas envolvidas na síntese, reorganização e degradação seletiva da parede celular. A habilidade de modificar seletivamente a arquitetura da parede é essencial para boa parte de processos do desenvolvimento como o crescimento celular, a abscisão de órgãos, a diferenciação vascular, o amadurecimento de frutos e respostas contra patógenos (Fry, 1986; Carpita and McCann 2000; Albert *et al.* 2004).

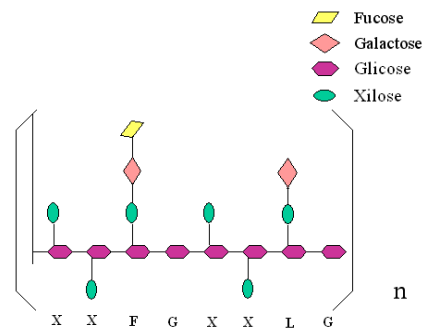
De forma geral a parede celular primária é composta basicamente de camadas de celulose imersas numa matriz de hemicelulose que totalizam 50% do peso seco da parede, pectinas que correspondem a 30%, ligninas e algumas proteínas estruturais com os 20% restantes do peso seco da parede celular (Carpita and Gibeaut 1993; Cosgrove 1997). Xiloglucanos são os principais componentes da hemicelulose da maioria das plantas terrestres, especialmente nas eudicotiledôneas (Kallas *et al* 2005). Eles se ligam à celulose por pontes-de-hidrogênio, revestindo e conectando microfibras adjacentes de celulose. (Hayashi 1989; McCann *et al.* 1990). Mudanças estruturais nestas redes são reguladas por modificações enzimáticas e acredita-se que enzimas de atividade

modificadora da parede tenham papel central na regulação da plasticidade adaptativa dos diversos tipos de paredes celulares. (Liu *et al.* 2007).

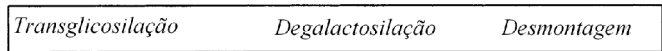
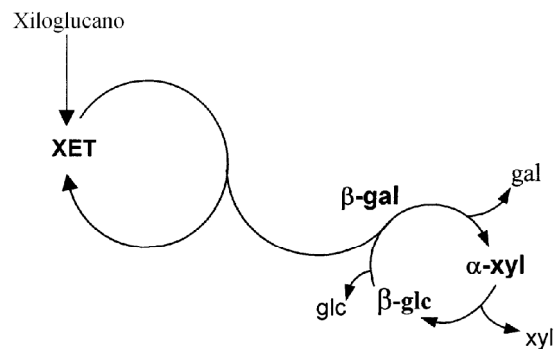
A estrutura básica dos xiloglucanos é composta por uma espinha dorsal de  $\beta$ -(1 $\rightarrow$ 4)-D-glucano que é regularmente substituída com resíduos  $\alpha$ -(1 $\rightarrow$ 6)-D-xilopiranosil. Dois padrões principais de xilosilação foram observados em paredes primárias estruturais e de estocagem, ambos baseados na repetição a cada quatro glicoses (Glc<sub>4</sub>) na espinha dorsal: XXXG e XXGG, onde X representa [ $\alpha$ -D-Xylp(1 $\rightarrow$ 6)]-Glc $\beta$ (1 $\rightarrow$ 4) e G representa Glc $\beta$ (1 $\rightarrow$ 4), de acordo com a nomenclatura de Fry *et al.* (1993). Repetições baseadas em cinco (Glc<sub>5</sub>) e seis (Glc<sub>6</sub>) glicoses são mais raras; XXXXG e XXXXXG respectivamente foram observadas em xiloglucano de reserva em sementes de Jatobá (Buckeridge *et al.*, 1997; Tiné *et al.*, 2006). O xiloglucano arquetípico de reserva da semente de *Tamarindus indica* é composto de oligossacarídeos dos tipos XXXG, XXLG, XLXG e XLLG, numa relação molecular de 1,4 : 3 : 1 : 5,4 (York *et al.* 1990) onde L é [ $\beta$ -D-Galp-(1 $\rightarrow$ 2)- $\alpha$ -D-Xylp(1 $\rightarrow$ 6)]-Glc $\beta$ (1 $\rightarrow$ 4). A maioria dos xiloglucanos estruturais apresenta também oligossacarídeos do tipo XXFG, onde F é L contendo uma substituição de [ $\alpha$ -L-Fucp(1 $\rightarrow$ 2)] ligada à galactose. (Baumann *et al.* 2007). Um exemplo esquemático pode ser visto na Figura 1.

Tiné *et al.* (2000), num trabalho com xiloglucano de semente de Jatobá, propuseram um modelo para a desmontagem ou degradação, *in vivo*, do xiloglucano de reserva (Figura 1) onde são necessárias quatro hidrolases para obtenção de monômeros de açúcar (galactose, glicose e xilose): XTH ou Xiloglucano Transglicosilase/Hidrolase,  $\beta$ -galactosidase,  $\alpha$ -xilosidase e  $\beta$ -glucosidase. Resumidamente, o polissacarídeo é hidrolisado por enzimas do tipo XTH e então subunidades do tipo XLLG e XLXG passam pela hidrolise da  $\beta$ -galactosidase, liberando monômeros de galactose. Os

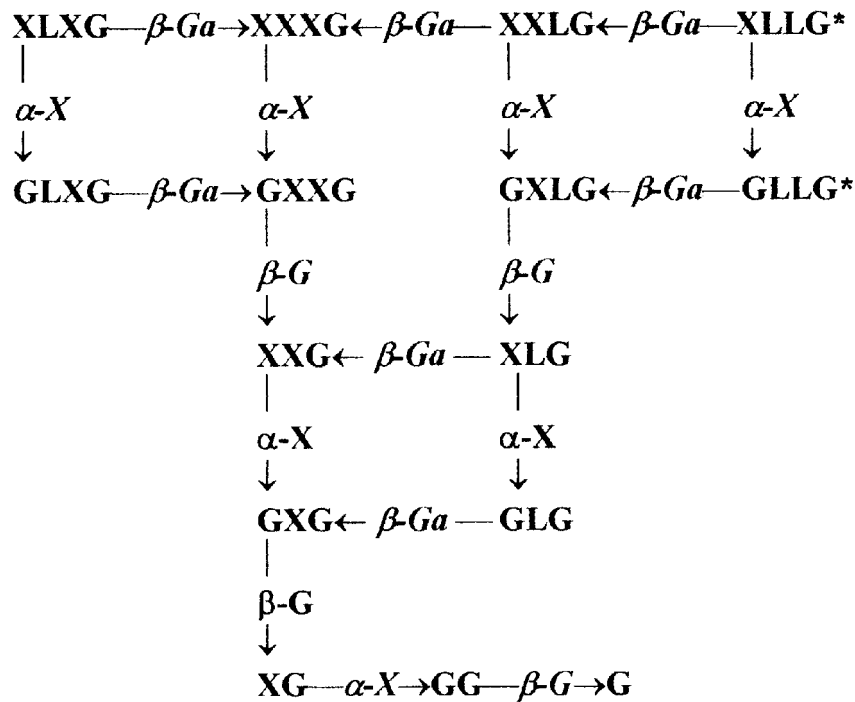
oligossacarídeos remanescentes são atacados por  $\alpha$ -xilosidase e  $\beta$ -glucosidase o que libera monômeros de xilose e glicose. Um modelo completo, contendo todas as possíveis etapas para a hidrólise completa de xiloglucano de reserva pode ser visto na Figura 2. Para a degradação do xiloglucano estrutural propusemos um modelo que tem a adição da atividade de uma  $\alpha$ -fucosidase, que seria necessária para a degradação de xiloglucanos contendo oligos do tipo XXFG (Figura 3).



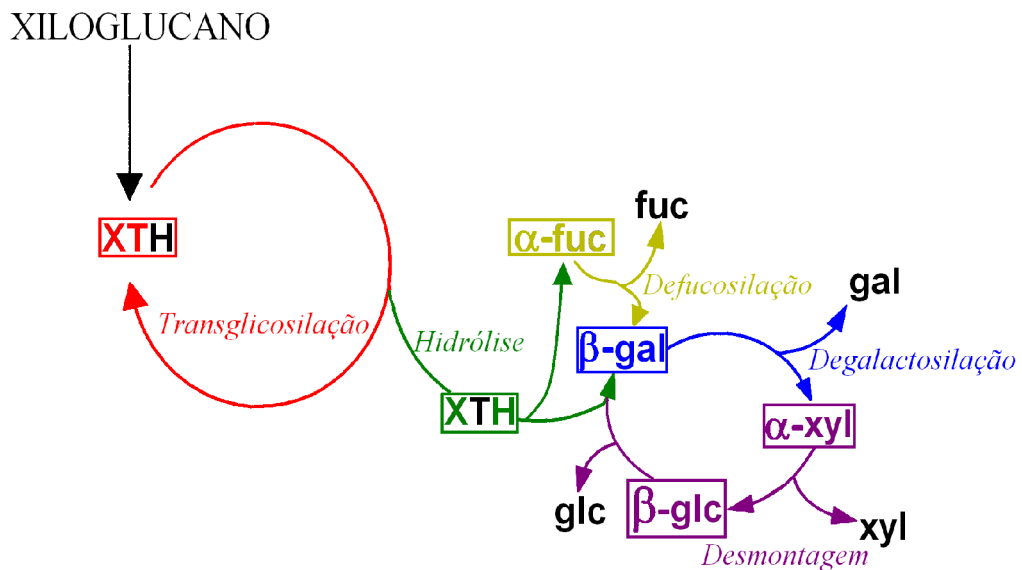
**Figura 1.** Representação esquemática de um xiloglucano hipotético. A figura mostra uma espinha dorsal de oito glicoses, formando dois oligos: XXFG e XXLG.



**Figura 2.** Modelo para degradação do xiloglucano de reserva. O modelo mostra a ação integrada das quatro hidrolases (XET – Xiloglucano endotransglicosilase;  $\beta$ -gal –  $\beta$ -galactosidase;  $\alpha$ -xyl –  $\alpha$ -xilosidase;  $\beta$ -glc –  $\beta$ -glucosidase) com a liberação de monômeros livres de glicose, galactose e xilose. Adaptado de Tiné *et al.* (2000).



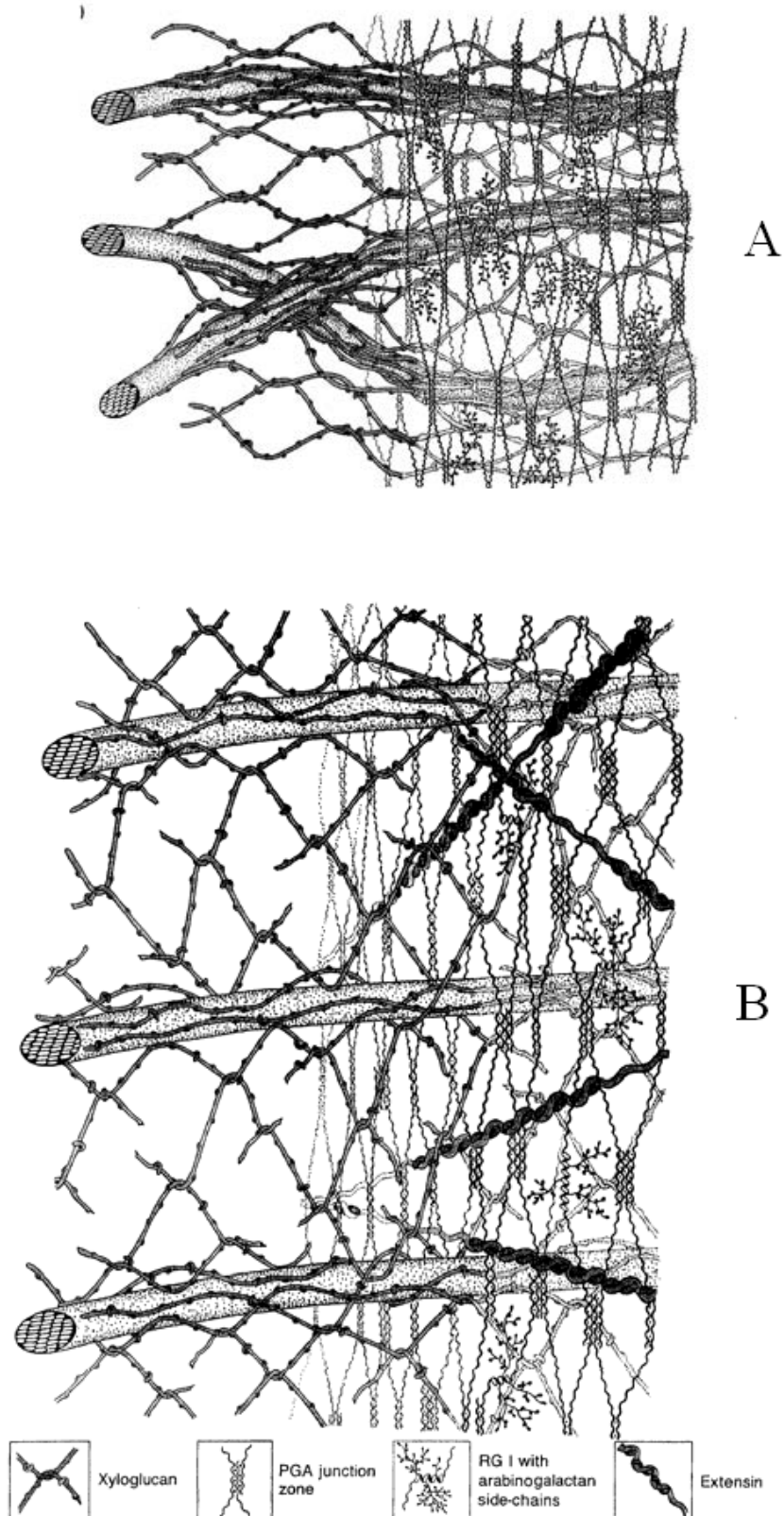
**Figura 3.** Vias possíveis na hidrólise completa de oligossacarídeos de xiloglucano de reserva.  $\beta\text{-Ga}$  =  $\beta$ -galactosidase,  $\alpha\text{-X}$  =  $\alpha$ -xilosidase e  $\beta\text{-G}$  =  $\beta$ -glucosidase. Os asteriscos (\*) indicam que as reações  $\text{XLLG} \xrightarrow{\beta\text{-Ga}} \text{XLXG}$  e  $\text{GLLG} \xrightarrow{\beta\text{-Ga}} \text{GLXG}$  não ocorrem. Figura reproduzida de Crombie *et al.* (1998).



**Figura 4.** Modelo para degradação do xiloglucano estrutural. Este tipo de polímero diferentemente do xiloglucano de reserva possui substituições  $[\alpha\text{-L-Fucp}(1\rightarrow2)\text{-}\beta\text{-D-Galp}(1\rightarrow2)\text{-}\alpha\text{-D-Xylp}(1\rightarrow6)]\text{-Glc}\beta(1\rightarrow4)$  que aparecem, normalmente, no padrão XXFG. A atividade hidrolítica da  $\alpha$ -fucosidase libera fucose e gera substrato para a  $\beta$ -galactosidase que, ao liberar galactose, gera substrato para a liberação de xilose e glicose por ação de  $\alpha$ -xilosidase e  $\beta$ -glucosidase respectivamente.

A matriz de hemicelulose e pectinas oferecem resistência dinâmica ao mesmo tempo em que permitem a expansão da parede celular. A dinâmica de alongação celular pode ser vista na Figura 5. A matriz de hemicelulose é rapidamente remodelada por XTHs realocando e conectando fragmentos de xiloglucano à porção redutora de outras cadeias já existentes de xiloglucano. Como resultado desta reação dinâmica a parede celular pode manter um metabolismo normal durante a expansão sem perda da resistência estrutural (Carpita, 1996).

Na privação de açúcares a parede celular vegetal pode ser usada na manutenção da homeostase energética por meio da degradação seletiva de polissacarídeos estruturais, como o xiloglucano, até a liberação de monômeros de açúcar. Lee *et al.* (2007) demonstrou, em *Arabidopsis thaliana*, que genes codificantes para uma  $\beta$ -galactosidase (At5g56870) e uma  $\beta$ -glucosidase (At3g60140), teoricamente ativas sobre xiloglucano, são induzidos pela privação de açúcares e reprimidos pelos açúcares que liberam (galactose e glicose, respectivamente) bem como por sacarose. A expressão desses genes e a acumulação das proteínas resultantes ocorrem na ausência de fotossíntese, o que reforça a idéia de que além de sustentação mecânica a parede celular vegetal também é usada como fonte de carboidratos na manutenção da homeostase energética.



**Figura 5.** Modelo de extensão da parede celular. **A.** Estágio de relaxamento da parede celular. **B.** Estágio de expansão da parede celular. PGA – *Penta-galacturonic acids*; RG I – *Rhamnogalacturonan I*. Reproduzido de Carpita & Gibeaut (1993).

O seqüenciamento completo do genoma de *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative, 2000) permitiu a identificação de 33 genes codificantes para XTHs (Rose *et al.* 2002). Em arroz o genoma completo possui 29 genes codificantes para XTHs (Yokoyama *et al.*, 2004). O genoma de *Arabidopsis* e o de Arroz codificam para, respectivamente, 17 e 15  $\beta$ -galactosidases (Ahn *et al.* 2007). Crombie *et al.* (1998) isolaram e caracterizaram uma  $\beta$ -D-glucosidase envolvida na degradação de xiloglucano de reserva em cotilédones de *Tropaeolum majus* (CAA07070 – GeneBank). Utilizamos esta seqüência para buscar genes evolutivamente relacionados em genomas vegetais. Apenas um gene codificante para uma  $\alpha$ -xilosidase, AtXYL1, foi caracterizado no genoma de *Arabidopsis* (Sampedro *et al.*, 2001) que é o provável ortólogo da  $\alpha$ -xilosidase isolada de cotilédones de *Tropaeolum majus* (CAA10382 – GeneBank), ambas atua sobre óligos de xiloglucano. Dois genes codificantes para proteínas com atividade  $\alpha$ -fucosidase foram identificados no genoma de *Arabidopsis*: AtFUC1 (Zeleny *et al.*, 2006) e AtFXG1 (de la Torre *et al.* 2002). Somente AtFXG1, uma  $\alpha$ -(1 $\rightarrow$ 2)-fucosidase, mostrou-se ativo sobre oligossacarídeos de xiloglucano clivando a ligação entre  $\alpha$ -L-Fuc(1 $\rightarrow$ 2)- $\beta$ -D-Gal enquanto AtFUC1 posteriormente foi caracterizado como sendo uma  $\alpha$ -L-3,4-fucosidase inativa sobre xiloglucano.



#### **1.4. Bibliotecas de cDNA**

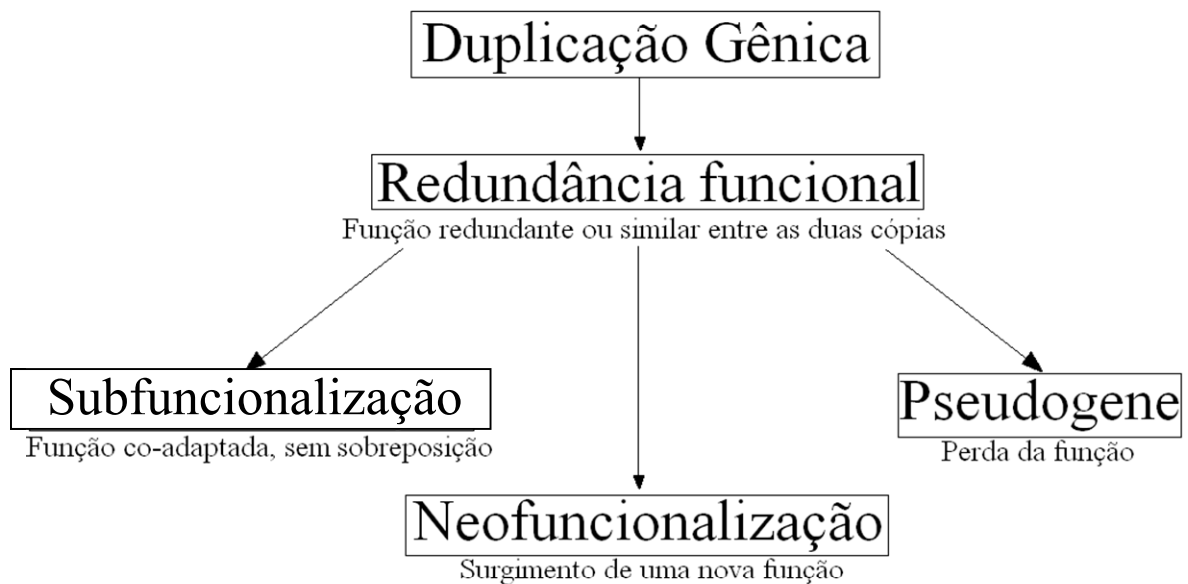
Seqüenciamentos simples de cDNAs clonados para gerar ESTs (“expressed sequence tags”) têm provado ser uma abordagem poderosa, rápida e econômica para identificar genes de uma determinada espécie e ainda ter idéia de quais genes são preferencialmente expressos em certos tecidos ou mesmo tipos celulares em organismos multicelulares. (Adams *et al.*, 1991; Hwang *et al.*, 1997; Liew *et al.*, 1994; Adams *et al.*, 1995). O uso de ESTs para anotar genomas completamente seqüenciados se mostrou uma ferramenta muito útil. Um único projeto EST fornece evidências biológicas de centenas de genes preditos computacionalmente, bem como permite a descoberta de novos genes ou de transcritos alternativos para genes conhecidos. Esta ferramenta permite avanços consideráveis na identificação de genes em organismos multicelulares (Andrews *et al.*, 2000). Atualmente (maio/2008) mais de 50 milhões de ESTs estão disponíveis através do dbEST (data base of “Expressed Sequence Tags”) no GeneBank ([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)). Mais de 11 milhões de ESTs foram obtidos em plantas e estão disponíveis (TIGR Plant Transcript Assemblies – Childs *et al.*, 2006), o que corresponde a aproximadamente 20% do total de ESTs gerados, totalizando 254 espécies ([http://plantta.tigr.org/cgi-bin/plantta\\_release.pl](http://plantta.tigr.org/cgi-bin/plantta_release.pl))

#### **1.5. Grupos de ortólogos**

Duplicação gênica é um processo essencial na diversificação e adaptação de organismos durante o processo de evolução, uma vez que genes duplicados representam adição de variabilidade sob a qual agem forças evolutivas, como mutação e seleção natural, que podem levar ao surgimento de novas funções (Alvarez-Buylla *et al.*, 2000;

Lawton-Rauh, 2003; Vandepoele *et al.*, 2003; Kellis *et al.*, 2004). A obtenção de novas cópias gênicas por recombinação não homóloga e/ou por duplicação de segmentos genômicos ou mesmo de genomas inteiros, são mecanismos que levam à formação de famílias multigênicas (Wendel, 2000; Bennetzen, 2002; Kellis *et al.*, 2004). Estes mecanismos vêm contribuindo evolutivamente para moldar o formato atual dos genomas eucarióticos conhecidos (Lespinet *et al.*, 2002; Lynch, 2002). Este fato pode ser ilustrado pelo genoma de *Arabidopsis thaliana*, onde as famílias multigênicas correspondem a 65% do seu genoma (Wendel, 2000).

Um gene duplicado apresenta, inicialmente, redundância funcional em relação ao gene do qual se originou. A partir deste ponto, quatro podem ser os destinos destas duas cópias (Figura 6). Primeiramente ambas as cópias podem manter a função original sendo completamente redundantes (redundância funcional). Segundo, as cópias podem evoluir para um estado onde a função que era desempenhada pelo gene ancestral passa a ser desempenhada pela ação conjunta de ambas as cópias (subfuncionalização). Terceiro, uma cópia pode manter a função ancestral enquanto a outra diverge funcionalmente a ponto de desempenhar uma nova função ou mesmo não participar mais do processo original. E, por último, o acúmulo de mutações pode tornar uma das cópias um pseudogene, que consiste em uma seqüência inativa de DNA genômico que apresenta similaridade e está evolutivamente relacionada a homólogos funcionais. (Lawton-Rauh, 2003).



**Figura 6.** Possíveis destinos de genes duplicados segundo uma trajetória temporal. Após a duplicação, as duas cópias geradas apresentam redundância funcional que pode se manter ao longo da evolução, ou pode haver subfuncionalização, neofuncionalização ou formação de um pseudogene. (Figura modificada de Lawton-Rauh, 2003).

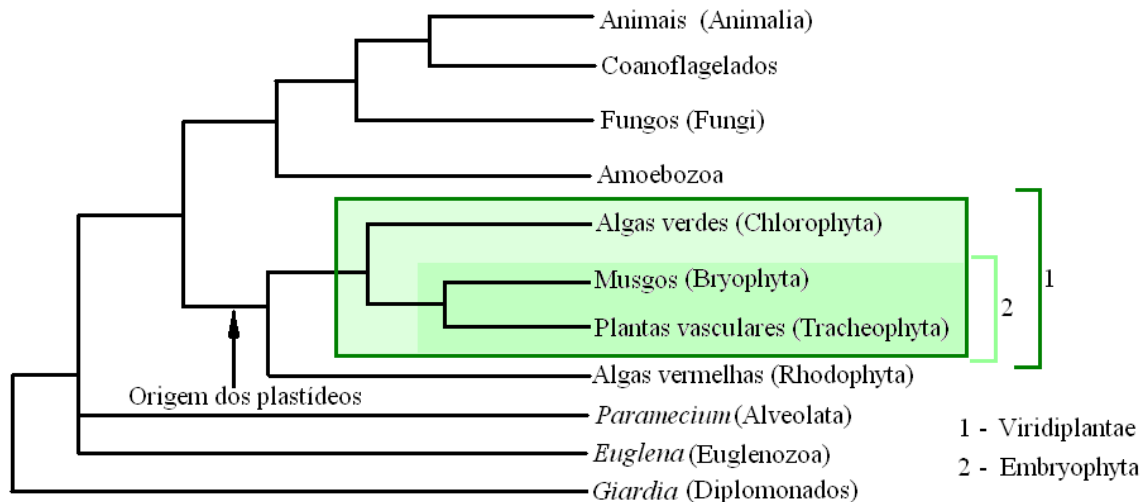
Através da comparação dos genomas de duas espécies de levedura, *Kluyveromyces waltii* e *Saccharomyces cerevisiae*, foi possível observar que para 95% dos 457 genes duplicados apenas uma das cópias sofreu evolução acelerada (gene derivado), enquanto a outra cópia mantém a informação original, permitindo haver uma distinção entre as funções ancestrais e derivadas (Kellis *et al.*, 2004). Este resultado sugere que a neofuncionalização foi o mecanismo mais privilegiado durante o processo evolutivo, pelo menos em leveduras.

Até o momento (junho de 2008) 86 espécies de Eucariotos tiveram seu genoma completamente seqüenciado (<http://www.genomesonline.org/>). No reino Viridiplantae (Embryophyta + Chlorophyta) 13 genomas completos foram seqüenciados: as angiospermas eudicotiledôneas *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), *Populus trichocarpa* (Álamo) (Tuskan *et al.* 2006), *Vitis vinifera* (uva vinifera) (Jaillon *et al.*, 2007) e *Glycine max* (soja) (2008, preliminar - <http://phyto3.phytozome.net/soybean.php>), as angiospermas monocotiledôneas *Oryza sativa* (arroz) (Goff *et al.*,

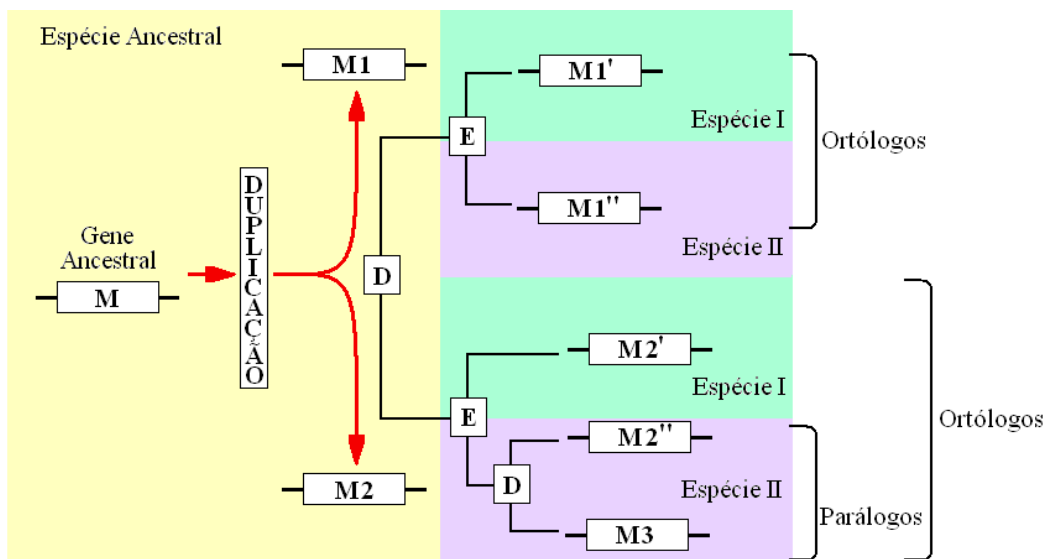
2002) e *Sorghum bicolor* (sorgo) (Paterson, 2008), a lycophyta *Selaginella moellendorffii* (2007, <http://selaginella.genomics.purdue.edu>), a bryophyta *Physcomitrella patens* (Rensing *et al.*, 2008) e as algas verdes *Ostreococcus lucimarinus* (Palenik *et al.*, 2007), *Ostreococcus tauri* (Derelle *et al.*, 2006), *Chlamydomonas reinhardtii* (Merchant *et al.*, 2007), *Micromonas pusilla* (2008, <http://genome.jgi-psf.org/MinpuC2/MicpuC2.info.html>) e *Volvox carteri* (2007, <http://genome.jgi-psf.org/Volca1/Volca1.info.html>). O acesso a esta grande quantidade de dados permite explorar comparativamente as bases genéticas da diversidade dentro da linhagem das plantas (Figura 7) (Hedges *et al.*, 2004; Chase & Fay, 2001). Um dos aspectos da abordagem comparativa envolve o estabelecimento de relações entre genes de diferentes genomas, em um sistema de genes homólogos, que inclui genes ortólogos e parálogos (Bennetzen, 2002; Pennacchio, 2003; Vincentz *et al.*, 2004; Figura 8). O termo ortólogo refere-se a genes homólogos que divergiram a partir de um evento de especiação, originados de uma cópia ancestral existente no ancestral comum entre as espécies que os contém, e parálogos são genes homólogos que resultaram de um evento de duplicação dentro de um genoma (Tatusov *et al.*, 1997; Thornton & DeSalle, 2000; Fitch, 2000; Meyerowitz, 2002).

As relações de ortologia entre genes de famílias multigênicas podem ser avaliadas através da definição de grupos de genes ortólogos. Um grupo de ortólogos é assumido como sendo o resultado da evolução de um gene ancestral através de eventos de especiação e duplicação. Considerando que todos os genes dentro de um mesmo grupo de ortólogos possuem uma mesma origem ancestral, a delimitação de um grupo deve facilitar a transferência de informações bioquímicas, estruturais e funcionais de uma proteína para outra dentro de um mesmo grupo (Tatusov *et al.*, 1997). Esta

abordagem otimiza a análise funcional de proteínas cujas funções ainda são desconhecidas. Grupos de ortólogos facilitam a interpretação de padrões evolutivos de genes homólogos, evidenciando taxas de evolução e de duplicação (Henikoff *et al.*, 1997).



**Figura 7.** Posição relativa da linhagem das plantas (Viridiplantae) dentro dos Eucariotos. As algas verdes formam um grupo monofilético com as Embryophytas que se separaram a aproximadamente 970 milhões de anos atrás. As Bryophytas se separaram das plantas vasculares (Tracheophyta) a aproximadamente 700 milhões de anos. A origem dos plastídios fotossintéticos remonta a 1,4 bilhões de anos atrás no ancestral comum entre Algas vermelhas e as plantas verdadeiras (Hedges *et al.*, 2004).



**Figura 8.** Representação das relações evolutivas de ortologia e paralogia entre genes. Um gene ancestral M sofre uma duplicação (D) originando os parálogos M1 e M2 no genoma da espécie ancestral. Um evento de especiação (E) gera dois grupos de ortólogos, M1 e M2 nas espécies I e II. Na espécie II, M2'' e M3 são parálogos gerados por uma duplicação que ocorreu depois da separação entre as espécies I e II.

## 1.6. Evolução molecular em famílias multigênicas

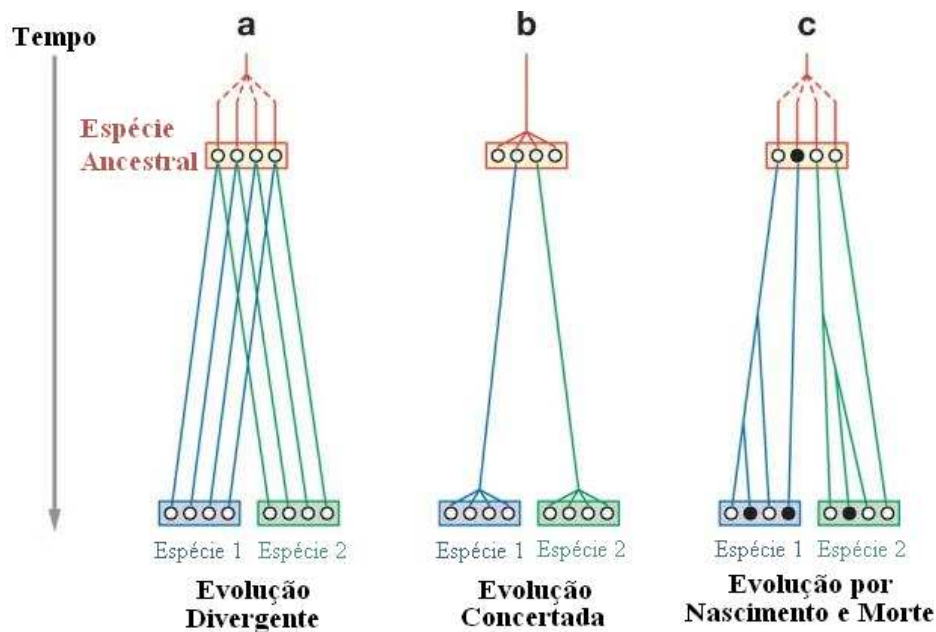
Uma família multigênica é um grupo de genes que descende de um mesmo gene ancestral e, portanto é esperado que tenham similaridades tanto em suas seqüências de DNA quanto em suas funções. (Nei & Rooney, 2005). Os processos evolutivos em famílias multigênicas tem sido objeto de controvérsia por muitos anos. O paradigma da evolução de famílias multigênicas antes de 1970 era o das Hemoglobinas  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  e mioglobina (Ingram, 1961). Os genes que codificam estas proteínas são filogeneticamente relacionados e divergiram gradualmente conforme os genes duplicados adquiriram novas funções. Este modelo de evolução é chamado de “evolução divergente” (“*divergente evolution*”) (Figura 9a).

Durante a década de 1970 pesquisas mostraram que RNAs ribossomais (rRNAs) em *Xenopus* são encodados por um grande número de genes repetidos *in tandem* e que as seqüências de nucleotídeos intergênica era mais similares dentro de uma espécie do que entre duas espécies relacionadas (*Xenopus laevis* e *X. mulleri*) (Brown, *et al.*, 1972). Em plantas os RNAs ribossomais também estão dispostos *in tandem* (Muir & Schlötterer, 1999; Figura 10) e apresentam o mesmo padrão evolutivo. Estas observações eram difíceis de explicar pelo modelo de “evolução divergente”, então um novo modelo chamado de “evolução orquestrada” (“*concerted evolution*”), onde era necessário um algum mecanismo que permita a homogeneização das diferentes cópias repetidas dentro de cada espécie ou população enquanto permite a divergência das repetições entre as espécies ou populações (Figura 9b). Neste modelo todos os membros de uma família de genes são assumidos como tendo evoluído de forma orquestrada, ou concatenada, ao invés de independentemente. Uma mutação ocorrida numa cópia se espalha através de todos os genes membros desta família através da ocorrência repetida

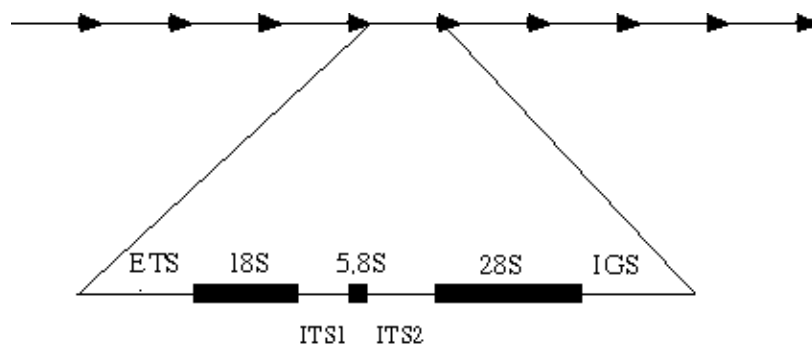
de recombinações (*crossover*) desiguais e/ou por conversão gênica (uma forma de recombinação não recíproca na qual segmento de DNA de um gene receptor é copiado de um gene doador). Este processo leva a uma homogeneização da seqüência gênicas e intergênicas e se mostrou capaz de explicar a evolução dos genes de rRNAs. No entanto, ao contrário do que se pensou inicialmente, não se pode excluir a ação de seleção purificadora nas seqüências transcritas dos rRNAs, que também leva à homogeneização, das cópias dentro de famílias multigênicas e o processo de evolução dos rRNAs parece ser um misto de seleção purificadora e “evolução orquestrada”. (Nei & Rooney, 2005).

Este sucesso aparente levou muitos autores a acreditar que a maioria das famílias multigênicas tinha evoluído seguindo o modelo de “evolução orquestrada” e impeliu investigações quanto aos aspectos evolucionários de várias famílias multigênicas (Hood *et al.*, 1975; Ohta, 1981; Zhang *et al.*, 2004). Mais tarde, no entanto, a aplicabilidade da “evolução orquestrada” para algumas famílias multigênicas foi questionada conforme um maior número de seqüências de DNA e proteínas se tornou disponível, e outro modelo de evolução chamado “nascimento e morte” (“*birth-and-death evolution*”) foi proposto (Nei *et al.*, 1992). Neste modelo novos genes são criados por duplicação gênica e alguns dos genes duplicados são mantidos no genoma por um longo período de tempo, enquanto outros são deletados ou se tornam não funcionais (pseudogenes) através de mutações deletérias (Figura 9c). Este modelo consegue explicar o surgimento de novas funções e se mostrou aplicável à maioria das famílias multigênicas relacionadas ao sistema imune como a das imunoglobulinas e MHC em animais (“*Major histocompatibility complex*”) (Hughes & Nei, 1990; Ota & Nei, 1994) e dos genes “*disease-resistance*” em primatas (Zhang *et al.*, 1998). Mesmo famílias

altamente conservadas como a das ubiquitinas e histonas parecem ter evoluído por “nascimento e morte” e seu alto grau de conservação se deve primariamente à seleção purificadora ao invés de “evolução orquestrada”. As controvérsias quanto à evolução de famílias multigênicas ainda se mantêm, em parte porque existem muitos tipos diferentes de famílias gênicas e em parte porque o mecanismo geral de conversão gênica ainda é pouco compreendido (Nei & Rooney, 2005).



**Figura 9.** Três modelos diferentes para evolução de famílias multigênicas. Círculos brancos representam genes funcionais e os pretos pseudogenes. Adaptado de Nei & Rooney, 2005.



**Figura 10.** Diagrama representando estrutura do rDNA no genoma de plantas vasculares. As repetições *in tandem* possuem quatro regiões intergênicas (ETS, ITS1, ITS2 e IGS) e três genes (18S, 5.8S e 28S) (Reproduzido de Muir & Schlötterer, 1999).



Em angiospermas o desenvolvimento dos órgãos florais é controlado por fatores de transcrição do tipo MADS-box e várias classes destes genes se mostraram essenciais para o desenvolvimento das flores (Ma & de Pamphilis, 2000; Theissen, 2001; Weigel & Meyerowitz, 1994). Análises filogenéticas dos genes MADS-box mostraram que estes genes, que controlam o desenvolvimento floral em angiospermas se originaram a aproximadamente 650 milhões de anos atrás (Nam *et al.*, 2003). Tanabe *et al.* (2005) encontrou genes MADS-box *like* em três espécies de algas verdes, que se acredita terem se originado a aproximadamente 700 milhões de anos atrás. Se pensarmos que os registros fósseis mais antigos de angiospermas e gimnospermas têm, respectivamente, 150 e 300 milhões de anos, aparentemente os genes ancestrais dos MADS-box florais existem por um grande período de tempo antes da evolução das flores. Especula-se que este grupo de genes originalmente controlava o desenvolvimento dos estágios haplóide e diplóide das algas-verdes (Tanabe *et al.*, 2005). MADS-box são genes ancestrais e existem em plantas, animais e fungos. Em animais eles controlam o desenvolvimento muscular. No processo de evolução das gimnospermas e angiospermas diferentes MADS-box parecem ter evoluído para formar os órgãos reprodutivos (Nam *et al.*, 2003), o que representa claramente um exemplo de evolução por “nascimento e morte”. Este parece ser o meio mais comum pelo qual evolui a maioria das famílias multigênicas, no entanto não se pode descartar o papel da de “evolução orquestrada” ou “evolução divergente” para algumas famílias, bem como o da seleção purificadora, além de processos mistos, como parece acontecer com proteínas *Heat-shock* e, em *Drosophila*, na família das amilases. Neste último caso ocorre um misto de por “nascimento e morte” e “evolução orquestrada”. (Nei & Rooney, 2005).

## 1.7. Filogenia

Para estabelecer as relações de ortologia/paralogia entre os membros homólogos que compõe uma família multigênica é necessário fazer inferências filogenéticas, na forma de árvore, o que permite reconstruir os grupos de ortólogos e estabelecer as relações evolutivas entre eles dentro da história evolutiva das espécies que compõe o *pool* de homólogos em análise.

Os métodos mais utilizados para construção de árvores filogenéticas incluem os métodos de distância, de parcimônia e de verossimilhança. Os métodos de reconstrução são baseados na disposição da topologia da árvore e no comprimento dos ramos de forma que a árvore resultante possa ser testada estatisticamente. (Nei & Kumar, 2000).

O método de Máxima Parcimônia foi originalmente desenvolvido para caracteres morfológicos (MP; Henning, 1966) sendo rapidamente utilizado para construção de árvores com dados de substituição de aminoácidos (Eck & Dayhoff, 1967). Neste método quatro ou mais seqüências de nucleotídeos ou aminoácidos alinhadas são utilizadas e então o modelo infere separadamente os nucleotídeos ou aminoácidos do taxa ancestral para cada posição do alinhamento para cada topologia, assumindo que as mudanças mutacionais acontecem em todas as direções entre os quatro nucleotídeos ou os 20 aminoácidos. O menor número de substituições que explica o processo evolutivo para cada topologia então é computado. Esta computação é feita para todas as árvores potencialmente corretas e a topologia que requer o menor número de substituições é então escolhida como a melhor árvore. A base teórica deste método é a idéia filosófica de William de Ockham, segundo a qual a melhor hipótese para explicar um processo é aquela que requer o menor número de eventos a serem assumidos.

Na ausência de substituições retrocedentes ou paralelas (ausência de homoplasia) em cada posição considerada e sendo o número de nucleotídeos ou aminoácidos ( $n$ ) muito grande o método de MP deverá produzir a topologia correta. Na prática, no entanto, as seqüências são alvo de mutações retrocedentes e paralelas e o número de nucleotídeos nos alinhamentos ( $n$ ) é em geral pequeno. Neste caso a MP tende a dar topologias incorretas. (Nei & Kumar, 2000). Felsenstein (1978) mostrou que quando a taxa de substituição de nucleotídeos varia muito entre as linhagens evolutivas os métodos de MP geram topologias incorretas mesmo quando um número infinito de nucleotídeos é analisado. Sob certas condições, mesmo quando a taxa de substituição for constante para todas as linhagens, os métodos de MP podem gerar resultados incorretos (Hendy & Penny 1989; Zharkikh & Li 1992; Takezaki & Nei, 1994; Kim, 1996). Nestes casos os ramos longos ou curtos da árvore correta tendem a se agrupar na árvore reconstruída por MP. Estes fenômenos são chamados de “*long-branch attraction*” (Hendy & Penny, 1989) e “*short-branch attraction*” (Nei, 1996). Outro aspecto negativo do método de MP é que o tempo de comparação necessário para analisar uma grande quantidade de seqüências é sempre muito longo. (Nei & Kumar, 2000).

O método de Máxima Verossimilhança (MV; Cavalli-Sforza & Edwards, 1967) originalmente foi desenvolvido para inferências filogenéticas baseadas em dados de frequência gênica. Felsenstein (1981) desenvolveu um algoritmo que utilizava seqüências de nucleotídeos para reconstruir árvores filogenéticas. Kishino *et al.* (1990) adaptou o algoritmo para utilizá-lo com seqüências de aminoácido usando a matriz experimental de substituição aminoácidos de Dayhoff *et al.* (1978).

No método de MV o parâmetro a ser considerado não é a topologia, mas sim o comprimento dos ramos para cada topologia e a verossimilhança é maximizada para

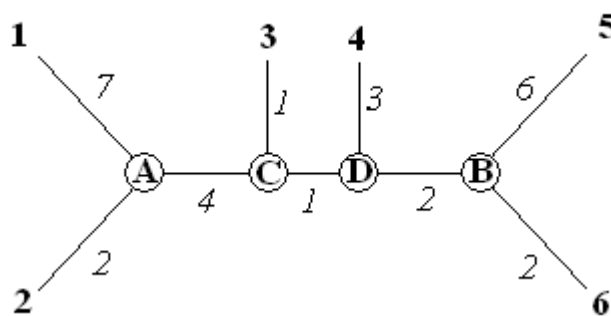
estimar o comprimento dos ramos em cada topologia testada. Este método apresenta menor variância do que os outros métodos e tende a ser robusto ao assumir o modelo evolutivo e tem uma boa base estatística para lidar com dados de seqüências, no entanto é um método quase proibitivo para análises que envolvem grande quantidade de seqüências devido ao seu grande consumo de memória (Nei & Kumar, 2000).

Métodos de distância computam as distâncias evolucionárias para todos os pares de taxa de forma que a árvore filogenética é construída considerando as relações ao longo destes valores de distância. (Nei & Kumar, 2000). Vários são os métodos que permitem construir árvores filogenéticas a partir de dados de distância. O mais simples é o UPGMA (*Unweighted pair-group method using arithmetic averages*) (Sokal & Michener, 1958; Sneath & Sokal, 1973). Originalmente desenvolvido para taxonomia numérica, este método funciona produz árvores razoavelmente boas quando as freqüências de substituição de nucleotídeos os aminoácidos mais ou menos constantes (Nei *et al.*, 1983; Takezaki & Nei, 1996). UPGMA tem como intenção reconstruir árvores de espécies (não árvores de genes), apesar de erros topológicos ocorrerem frequentemente quando a taxa de substituição não é constante ou quando o número de genes ou nucleotídeos é pequeno. (Nei & Kumar, 2000).

Outro método de distância bastante utilizado é o método de evolução mínima ou *Minimum Evolution* (ME; Edwards & Cavalli-Sforza, 1963; Kidd & Sgaramella-Zonta, 1971). Neste método o somatório total das distâncias (S) representadas pelo comprimento dos ramos de uma árvore é calculado para todas as topologias plausíveis e então a árvore com o menor valor de S é escolhida como a melhor árvore (Nei & Kumar, 2000). É um método com boa base estatística, mas consome muito tempo de processamento (Rzhetsky & Nei, 1992).

Para contornar o problema do grande tempo despendido para processar filogenias utilizando ME, Saitou & Nei (1987) desenvolveram um método de construção de árvores eficiente e que se baseia no mesmo princípio de ME. Este método não examina todas as topologias possíveis, mas a cada estágio de agrupamento das seqüências o princípio da evolução mínima é utilizado. Este método, que consiste numa simplificação de ME, é chamado *Neighbor-Joining* (NJ).

Um dos mais importantes conceitos no método de NJ é o de vizinhos, que são definidos como os dois taxa, ou seqüências, que estão conectadas por apenas um nó numa árvore sem raiz. Por exemplo, **1** e **2**, na árvore da Figura 11, são vizinhos pois estão conectados por somente um nó **A**. Da mesma forma **5** e **6** são vizinhos conectados pelo nó **B**. Todos os outros pares não são vizinhos. Entretanto, **1** e **2** podem ser combinados e considerados um único táxon, então **(1-2)** e **3** passam a serem vizinhos. É possível definir a topologia da árvore pelo agrupamento sucessivo dos vizinhos e pela produção de novos vizinhos. A topologia da árvore da Figura 11 pode ser descrita pelos seguintes pares de vizinhos: **(1, 2)**, **(5, 6)**, **(1-2, 3)**, e **(1-2-3, 4)** (Nei & Kumar, 2000).



**Figura 11.** Filogenia de seis seqüências com tamanho dos ramos conhecidos. Adaptado de Nei & Kumar, 2000.

O método de NJ é rápido, apropriado para grandes conjuntos de dados, permite linhagens com diferentes tamanhos de ramos e substituições múltiplas, no entanto ele mostra apenas uma topologia possível (Nei & Kumar, 2000).

A utilização de métodos que requerem uma menor capacidade computacional, como o NJ, pode levar a conclusões similares às aquelas por métodos mais dispendiosos de tempo, como MP e MV (Nei *et al.*, 1998). Principalmente quando o número de seqüências ( $m$ ) e a divergência são pequenos geralmente reconstruir a árvore é difícil por qualquer método (Nei & Kumar, 2000). Alguns autores utilizam métodos sofisticados baseados em MV, no entanto estes métodos são muito demorados ainda mais quando se deseja uma árvore que apresente testes estatísticos como o teste de *Bootstrap* (Felsenstein, 1985) ou *Bootstrap Interior Branch Test* (Dopazo, 1994). Construir árvores por MP com teste estatístico também é muito consumidor de tempo, principalmente quando  $m$  é grande. Em contraste, o método de NJ constrói, com facilidade, árvores com teste estatístico mesmo quando  $m$  é maior que 200. (Nei & Kumar, 2000).

A árvore consenso obtida pelo teste de *Bootstrap* pelos três métodos é usualmente muito similar, apesar de alguns ramos interiores com suporte fraco poderem diferir. Portanto, não existe a necessidade de gastar grandes períodos de tempo para construir árvores consenso baseadas em MV e MP. Neste caso o mesmo resultado pode ser obtido pelo método de NJ muito rapidamente. Além disso, para casos onde  $m > 200$ , árvores feitas com as distâncias calculadas por *p-distance* normalmente apresentam resultados melhores devidos às distâncias apresentarem menor variância (Nei & Kumar, 2000).

## **1.8. Objetivos**

### **1.8.1. Para o modelo de impacto ambiental:**

- Padronizar protocolos de extração de RNA total para todos os tecidos (raiz, caule, cotilédones e folhas) de plântulas de Jatobá;
- Construir bibliotecas de cDNA tecido-específicas de Jatobá, validá-las e analisar as primeiras seqüências EST;

### **1.8.2. Para o modelo de remobilização de Xiloglucano:**

- Identificar seqüências das cinco hidrolases (XTH,  $\beta$ -galactosidase,  $\alpha$ -xilosidase,  $\alpha$ -fucosidase e  $\beta$ -glucosidase), que constituem o mecanismo geral de degradação de xiloglucanos, em genomas completos e ESTs de plantas verdes (Reino Viridiplantae).
- Definição de grupos de genes homólogos e grupos de gene ortólogos das cinco famílias de hidrolases e a utilização destes dados para racionalizar estudos funcionais e buscar padrões evolutivos destas enzimas ao longo da evolução das plantas verdes.

## 2. Materiais e Métodos

### 2.1. Bibliotecas de cDNA

#### 2.1.1. Material biológico

As plântulas de Jatobá foram coletas com aproximadamente 45 dias após a germinação. As coletas foram feitas às 6h (início do dia) e às 18h (final do dia) com foto período de 12 horas luz / 12 horas escuro. Foram realizados quatro tipos de tratamento: 1- condição controle (12 horas luz e 12 horas escuro); 2- plântulas com folhas completamente cobertas (com papel alumínio) por 3 dias antes da coleta; 3- plântulas crescidas durante 30 dias em escuro contínuo e 4- plântulas machucadas com um bisturi (estresse mecânico) nas folhas (todas), caule (hipocótilo e epicótilo) e nos cotilédones 3 dias antes da coleta. Para cada tipo de tratamento foram feitas duas coletas (6h e 18 h) (Tabela 1).

**Tabela 1** - Condições de crescimento das plântulas de Jatobá.

	CC 6 horas	CC 18 horas	FC 6 horas	FC 18 horas	EC 6 horas	EC 18 horas	PM 6 horas	PM 18 horas
Caule	+	+	-	-	+	+	+	+
Cotilédones	+	+	-	-	+	+	+	+
Folhas	+	+	+	+	+	+	+	+
Raízes	+	+	-	-	+	+	-	-

+ indica a extração de RNA total e – indica que tal condição não foi contemplada. CC, Condição controle; FC, Folhas cobertas por 3 dias; EC, Escuro contínuo; PM, Plantas Machucadas (estresse mecânico)



### 2.1.2. Extração de RNA total

A extração do RNA total das folhas primárias e das raízes foi realizada segundo Gesteira et al. (2003). Para os cotilédones, utilizou-se a metodologia segundo Daohong et al. (2004) com modificações. O material foi macerado em N<sub>2</sub> líquido e homogeneizado com agitação durante 10 min em tampão de extração (Tris-HCl 0,5M, pH 8; LiCl 0,25M; EDTA 0,01M; SDS 2,5% e β-Mercaptoetanol 0,1%) e fenol:clorofórmio:álcool isoamílico (PCI) (25:24:1). Em seguida, o extrato foi centrifugado a 4°C durante 15min a 14000 RPM. Repetiu-se a etapa fenol:clorofórmio:álcool isoamílico (PCI) (25:24:1) mais 3 vezes. Para precipitar o RNA, 0,5 volumes de LiCl 12M foram adicionados e à mistura foi armazenada por 30 min a -80 °C. Após centrifugação a 4°C por 30min a 14000rpm o *pellet* foi lavado com 1mL de LiCl 6M, e centrifugado durante 15min nas mesmas condições. O precipitado foi diluído em 30µl de H<sub>2</sub>O tratada com DEPC (0,1%) (Diethyl pirocarbonato) e o RNA foi reprecipitado em 0,2 volumes de acetato de sódio 3M pH 5,2 e 1 volume de etanol absoluto durante 2 horas a -80 °C ou permanecendo *overnight* a -20°C. Após 30min de centrifugação a 4°C a 14000rpm, o *pellet* foi lavado com etanol 70% e o RNA foi ressuspendido em 30µl de H<sub>2</sub>O DEPC. Ambos os protocolos tiveram o acréscimo, de 2 etapas a mais de fenol:clorofórmio:álcool isoamílico o que melhorou a pureza do RNA.

Para extração de RNA Total de folhas secundárias e caule (hipocótilo e epicótilo) foi utilizado o reagente Concert – Invitrogen (de acordo com especificações do fabricante).

A quantificação da concentração foi feita através de absorbância a 260nm em espectrofotômetro. A integridade do RNA total foi verificada através de eletroforese em gel de agarose 1%, contendo 6% de formaldeído, em tampão MOPS 1X (MOPS 10x:

41,8g de MOPS, 16,6mL de acetato de sódio 3M, 20mL de EDTA 0,5M, pH 8; água q.s.p. 1000mL; pH 7). As amostras antes de aplicadas no gel são desnaturadas a 65°C por 15min em Tampão contendo 10% do volume total de MOPS 10x, 17,5% de Formaldeído 37%, 50% de Formamida, 4% de Brometo de Etídio (10mg/ml) e 17,5% de Tampão de Carga (2µl de EDTA 0,5M, 500µl de Glicerol 50%, 125µl de Xileno Cianol 10% e 125µl de Azul de Bromofenol 10%).

Os RNAs das condições de cultura descritas na Tabela 1 foram obtidos. Os RNAs totais obtidos foram misturados de forma equimolar para obtenção de um *mix* de RNA total para cada tecido (Ver Resultados).

### **2.1.3. Purificação de RNA Poli A+**

O RNA Total dos quatro mix (Folhas, Caule, Raízes e Cotilédone) foram purificados para RNA Poli A+ utilizando o kit Oligotex mRNA Purification System (Qiagen) segundo as instruções do fabricante. As quantidades de mRNA Poli A + obtidas estão descritas nos resultados.

### **2.1.4. Síntese de cDNA, LD-PCR e fracionamento dos cDNAs por tamanho**

Para gerar a primeira fita de cDNA foram combinados: 100ng de RNA Poli A+ em um volume de 2µl, 1µl de primer *attB2* Oligo d(T) (5' – GGC GGC CGC ACA ACT TTG TAC AAG AAA GTT GGG T(T)<sub>19</sub> – 3') (10µM) e 7µl de Água milli-Q RNase free, completando um volume final de 10ul. A mistura foi encubada por 2 minutos a 72°C e depois resfriada por 2 min em gelo. Adicionou-se à mistura 2,0µL de 5X First-Strand Buffer, 1,0µL DTT (20mM), 1,0µL dNTP Mix (10mM) e 2,0µL de SuperScript II RT (Invitrogen). A mistura foi encubada, para que ocorresse a transcrição

reversa, por duas horas à 42°C. Para inativação da enzima utilizamos uma incubação por 10 minutos a 75°C.

A enzima SuperScript II RT possui atividade de transferência de citosinas à parte 3' final da primeira fita do cDNA. Baseado nessa atividade, desenhamos um primer *attB1* Poly-G (5' – TCG TCG GGG ACA ACT TTG TAC AAA AAA GTT GGG – 3') que anela na parte 3' dos cDNAs e quando utilizado em conjunto com o primer *attB2* Oligo d(T) permite a realização do LD-PCR (Chen *et al* 2005) para que seja possível obter uma amplificação da quantidade de cDNA.

Para a amplificação por LD-PCR foram combinados em um tubo estéril 10ng de cDNA (aprox. 1/10 da reação de síntese), 2µl de dNTP Mix (10mM), 2µl de primer *attB1* Poly-G (10µM), 2µl de primer *attB2* Oligo d(T) (10µM), 10µl de 10X HF Buffer, 4µl de MgCl<sub>2</sub> (50mM), e 1µl de Platinum Taq Polymerase HiFi (5U/µL - Invitrogen). A reação foi completada com água milli-Q para o volume de 100µl e o volume total foi dividido em 4 tubos de PCR de 0,2ml. O PCR foi realizado em termociclador PTC-100 MJ Research Inc. Foram realizados 24 ciclos de amplificação (95 °C por 20 segundos, anelamento dos primers e extensão a 68°C por 6 min) com desnaturação inicial a 95°C por 5 min e uma fase de alongamento final de 10 min a 72 °C.

Após o LD-PCR 8µl da reação foi corrido em gel de agarose 1% corado com brometo de etídeo para verificar a amplificação.

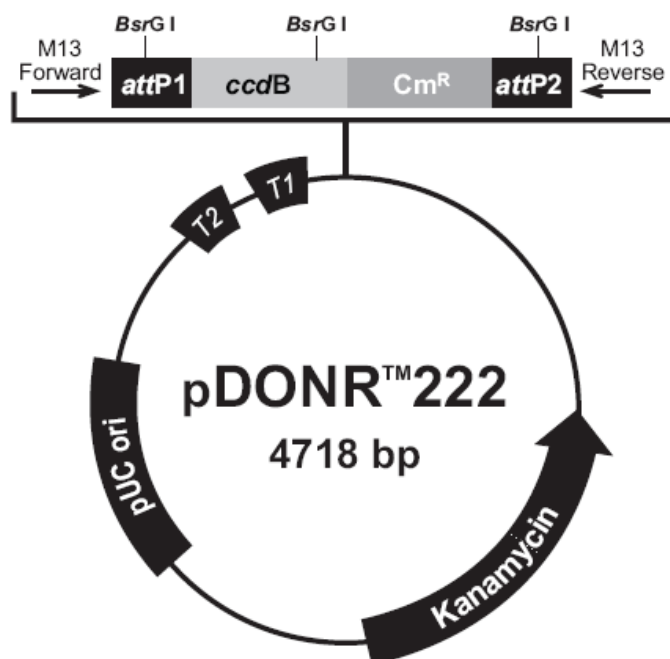
Uma vez verificada a amplificação o restante do produto de LD-PCR foi aquecido por 10 min a 75°C e foi corrido em gel de agarose 2,5% para separação dos cDNAs por tamanho (Kenzelmann & Mühlemann, 1999). A fração do gel correspondente aos fragmentos entre 500pb e 3000pb foi excisada do gel e purificada

utilizando o kit Wizard® SV Gel and PCR Clean-Up System (Promega) segundo as instruções do fabricante.

O cDNA purificado foi quantificado por espectrofotometria e toda a eluição foi seca utilizando um Speed Vac (Eppendorf). O cDNA foi ressuspensionado em TE 1x (Tris 10mM, EDTA 1mM, pH 8,0) na concentração de 50ng/μl.

### 2.1.5. Clonagem via recombinase BP

Após a amplificação por LD-PCR o cDNA é fita dupla e está flanqueado pelas seqüências de recombinação *attB1* e *attB2*, o que permitirá clonagem orientada, via recombinase sítio específica derivadas do bacteriófago lambda, no vetor pDONR-222 (Figura 12) entre os sítios *attP1* e *attP2*.



**Figura 12.** Mapa do vetor pDONR-222. O vetor possui gene de resistência à Kanamicina e origem de replicação de pUC. T1 e T2 são seqüências “transcription terminators”. Flanqueados pelas seqüências para os primers de seqüenciamento M13Forward e M13Reverse estão as seqüências alvo da ClonaseBP, *attP1* e *attP2*, e estas flanqueiam os gene de letalidade *ccdB* e o gene de resistência a Chloramphenicol.

Para a reação de recombinação foram combinados em um tubo estéril 2µl de cDNA (50ng/µl), 1µl de vetor pDONR-222 (250ng/µl), 2µl 10X BP Buffer e 2µl de BP Clonase Enzyme Mix (Invitrogen). A reação foi completada para 10µl com TE 1X (Tris 10mM, EDTA 1mM, pH 8,0) e encubada a 25°C por 20 horas. Depois da encubação 2µl de proteinase K (2µg/µl) foram adicionados à reação e esta encubada por 15 min a 37°C e depois 10 min a 75°C para a inativação da proteinase K.

Para a precipitação do produto da clonagem foi adicionado à reação 90µl de Água milli-Q, 100µl de NH<sub>4</sub>Ac. (7,5M) e 375µl de Etanol 100%. A mistura foi deixada overnight a -20°C e centrifugada a 14000rpm por 1 hora. O sobrenadante foi descartado e duas lavagens com 500µl de Etanol 70% a -20°C foram feitas com centrifugação de 5 min a 14000rpm. Após a ultima lavagem o sobrenadante foi descartado e o DNA foi seco à vácuo por 5 min e ressuscitado em 10µl de água milli-Q.

### **2.1.6 – Obtenção e transformação de células competentes**

Células de *E. coli* da linhagem DH10B foram plaqueadas em meio sólido LB (10g de Triptona, 5g de Extrato de levedura, 10g de NaCl, 15g de Agar em 1L de Água Milli-Q) contendo 100µg/ml de Streptomicina. Uma colônia isolada foi pega e crescida *overnight* a 37°C com agitação (250rpm) em erlenmeyer com 50ml de LB líquido contendo 100µg/ml de Streptomicina. Deste pré-inóculo 10ml foram misturados a 200ml de LB líquido (idêntico ao LB sólido exceto por não conter agar) sem antibiótico e deixado crescer até atingir OD<sub>600</sub> de 0,8. Então o meio foi transferido para quatro tubos falcon de 50ml e centrifugado por 10min a 4000 RPM a 4°C. O sobrenadante foi descartado e o *pellet* foi ressuscitado em 10ml de Glicerol 10% a 4°C e centrifugado por 6min a 4000rpm a 4°C. O sobrenadante foi novamente descartado e o *pellet* foi ressuscitado em 400ul de Glicerol 10% a 4°C e centrifugado novamente por 6min a

4000rpm a 4°C. Esta etapa foi repetida mais três vezes. Então as bactérias foram ressuspensas em 200µl de Glicerol 10% e aliqüotadas em tubos estéreis contendo 50µl de células que foram prontamente congeladas em nitrogênio líquido e estocadas a -80°C.

Para calcular a eficiência das células competentes uma aliqüota foi eletroporada (1,6kV) com 1ng de vetor PUC18, deixada crescer em 1ml de meio LB líquido por 1 hora e 50µl desse meio foi plaqueado em meio LB sólido contendo 100µg/ml de Ampicilina. A placa foi encubada overnight a 37°C e as colônias foram contadas e a eficiência das células foi calculada em colônias/µg de PUC18. Células com eficiência maior ou igual a  $10^7/\mu\text{g}$  de PUC18 foram consideradas adequadas.

Tendo sido obtidas células eficientes 1µl da recombinação (ressuspensa em 10µl, aproximadamente 35ng de DNA) foi eletroporado a 1,6kV numa aliqüota de 50µl de células utilizando-se cubetas com espaço de 1mm entre os eletrodos. Seguido à eletroporação 1ml de meio LB líquido foi adicionado às células e estas foram deixadas 1 hora a 37°C com agitação. Após este período 50µl de meio foi plaqueado em meio LB sólido contendo 50µg/ml de Kanamicina (resistência incorporada no vetor pDONR-222) e a placa foi encubada overnight a 37°C. As colônias são então contadas e o Título da biblioteca (quantidade de clones que podem ser seqüenciados) é calculado. O restante da transformação (1ml) foi misturado a 150µl de Glicerol 100% e congelado em nitrogênio líquido e armazenado a -80°C como estoque da biblioteca.

### **2.1.7 - PCR de colônias**

Para que pudéssemos estimar o percentual de clones positivos e a diversidade de tamanho dos insertos 29 colônias de cada biblioteca foram picadas individualmente numa reação de PCR com volume total de 10µl usando os primers de seqüenciamento

M13 (M13Forward 5'-GTAAAACGACGGCCAGT-3' e M13Reverse 5'-AACAGCTATGACCATG-3'). (200µM de dNTPs Mix, 200µM de primer M13Forward, 200µM de primer M13Reverse, 2mM de MgCl<sub>2</sub>, 1X Taq Polimerase Buffer e 0,05U/µl de Taq Polimerase Recombinante - Invitrogen). O PCR foi realizado em termociclador PTC-100 MJ Research Inc. Foram realizados 35 ciclos de amplificação (95 °C por 20s, anelamento dos primers a 56°C por 30s, extensão por 1min e 30s a 72°C) com desnaturação inicial a 95°C por 5min e uma fase de alongamento final de 10min a 72°C.

### **2.1.8 - Construção das Bibliotecas**

Uma vez que o título da biblioteca foi calculado e o tamanho dos insertos foi considerado satisfatório colônias isoladas eram picadas da placa de meio sólido para placas de cultura de 96 poços contendo 100µl de meio LB líquido (15 % Glicerol, 50µg/ml de Kanamicina) por poço. As placas de 96 poços foram deixadas crescer por 21 horas a 37°C e depois congeladas e armazenadas a -80°C. Foram obtidas 10 placas para cada biblioteca (raiz, folha, caule e cotilédones) totalizando 960 clones de cada tecido, 3840 clones no total.

### **2.1.9 – Minipreparação plasmidial por lise alcalina**

Cada placa da biblioteca teve 2µl da cultura retirados e estes foram inoculados em placas de 96 poços para crescimento num volume de 1ml de meio LB líquido contendo 50µg/ml de Kanamicina. Os inóculos foram crescidos por 21 horas a 37°C.

Após o crescimento as placas foram centrifugadas a 4000rpm por 10 minutos, o sobrenadante foi descartado e o *pellet* de bactérias foi ressuspenso em 240µl de GTE (Glicose 0,1%, Tris pH 7,4 0,025M, EDTA pH 8,0 0,01M) e novamente centrifugado a

4000rpm por 10 minutos e o sobrenadante descartado. O *pellet* foi ressuspenso em 80µl de GTE e 60µl do homogeneizado foram transferidos para placas de 96 poços fundo V contendo 5,0µl de RNase A (10µg/µl). Adicionou-se 60µl de NaOH (0,2M)/SDS (0,2%) e a placa, selada com adesivo, foi invertida para misturar e deixada 10 minutos à temperatura ambiente e então 60µl de KOAc (3M) foi adicionado e a placa foi novamente selada e invertida. O adesivo foi retirado e a placa incubada a 90°C por 30min. Centrifugou-se por 15min a 4000rpm e 120µl do sobrenadante foram transferidos para uma nova placa de fundo V contendo 100µl de Isopropanol 100% a -20°C, deixada por 5min a temperatura ambiente e centrifugada por 60min a temperatura ambiente. O sobrenadante foi descartado e adicionado 200µl de Etanol 70% a -20°C e novamente centrifugou-se por 5min. O sobrenadante foi novamente descartado e as placas foram secas sob vácuo por 5min. O DNA foi ressuspenso em 40µl de Água Milli-Q e 2µl de cada poço foram corridos em gel de agarose 0,8% corado com brometo de etídio para verificar a eficiência da extração plasmidial.

#### **2.1.10 - Seqüenciamento**

O seqüenciamento foi realizado utilizando o kit Big Dye terminator 3.1 (Applied Biosystems) seguindo as instruções do fabricante. As reações eram seqüenciadas em um Seqüenciador Automático de DNA ABI PRISM modelos 3730 (Applied Biosystems).

#### **2.1.11 - Montagem das seqüências e anotação**

As seqüências serão montadas utilizando o software CAP3 (Huang & Madan, 1999) e anotadas por comparação com bancos de seqüências de plantas que possuem genoma completo anotado (*Arabidopsis thaliana* e *Oryza sativa*) por Blastx local (Altschul *et al.*, 1997).



## 2.2. Filogenias

### 2.2.1. Material computacional

Todas as análises computacionais foram realizadas com um computador com processador Pentium® Dual Core 1,80 GHz, disco rígido de 250Gb e memória RAM de 2Gb. O sistema operacional utilizado foi Windows XP®. Para buscar as seqüências de interesse utilizamos o programa Blast 1.1.17 (Altschul *et al.*, 1990; [www.ncbi.nih.gov/blast](http://www.ncbi.nih.gov/blast)). Para gerar os alinhamentos utilizamos o software ClustalW (Thompson *et al.*, 1994) e para realizar as análises filogenéticas utilizamos o programa MEGA4 (Tamura *et al.*, 2007; [www.megasoftware.net](http://www.megasoftware.net)).

### 2.2.2. Obtenção das seqüências

Inicialmente obtivemos os conjuntos completo de seqüências genômicas, EST e seqüências protéicas de todos os genomas do reino Viridiplantae que haviam sido completamente seqüenciados até maio de 2008. No total, 13 genomas foram utilizados em nossas análises, totalizando um conjunto de 365.187 proteínas preditas. A lista dos genomas utilizados pode ser vista na Tabela 2.

**Tabela 2.** Lista das espécies cujos genomas completos foram obtidos para análise.

Espécie	Versão	Site	Referência
<i>Arabidopsis thaliana</i> (Eudicotiledônea)	7.0	<a href="http://www.arabidopsis.org">www.arabidopsis.org</a>	Arabidopsis Genome Initiative, 2000
<i>Populus trichocarpa</i> (Eudicotiledônea)	1.1	<a href="http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html">http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html</a>	Tuskan <i>et al.</i> 2006
<i>Glycine max</i> (Eudicotiledônea)	0.1	<a href="http://www.phytozome.net/soybean.php">www.phytozome.net/soybean.php</a>	Dados não publicados, 2008
<i>Vitis vinifera</i> (Eudicotiledônea)	1.0	<a href="http://www.gramene.org/Vitis_vinifera/index.html">www.gramene.org/Vitis_vinifera/index.html</a>	Jaillon <i>et al.</i> , 2007
<i>Oryza sativa</i> (Monocotiledônea)	5.0	<a href="http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml">www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml</a>	Goff <i>et al.</i> , 2002
<i>Sorghum bicolor</i> (Monocotiledônea)	1.4	<a href="http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html">http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html</a>	Paterson, 2008
<i>Selaginella moellendorffii</i> (Lycopphyta)	1.0	<a href="http://genome.jgi-psf.org/Selmo1/Selmo1.home.html">http://genome.jgi-psf.org/Selmo1/Selmo1.home.html</a>	Dados não publicados, 2007
<i>Physcomitrella patens</i> (Bryophyta)	1.1	<a href="http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html">http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html</a>	Rensing <i>et al.</i> , 2008
<i>Volvox carteri</i> (Chlorophyceae)	1.0	<a href="http://genome.jgi-psf.org/Volca1/Volca1.home.html">http://genome.jgi-psf.org/Volca1/Volca1.home.html</a>	Dados não publicados, 2007
<i>Chlamydomonas reinhardtii</i> (Chlorophyceae)	3.0	<a href="http://genome.jgi-psf.org/chlre3/chlre3.home.html">http://genome.jgi-psf.org/chlre3/chlre3.home.html</a>	Merchant <i>et al.</i> , 2007
<i>Ostreococcus lucimarinus</i> (Prasinophyceae)	2.0	<a href="http://genome.jgi-psf.org/Ost9901_3/Ost9901_3.html">http://genome.jgi-psf.org/Ost9901_3/Ost9901_3.html</a>	Palenik <i>et al.</i> , 2007
<i>Ostreococcus tauri</i> (Prasinophyceae)	2.0	<a href="http://genome.jgi-psf.org/Ostta4/Ostta4.home.html">http://genome.jgi-psf.org/Ostta4/Ostta4.home.html</a>	Derelle <i>et al.</i> , 2006
<i>Micromonas pusilla</i> (Prasinophyceae)	2.0	<a href="http://genome.jgi-psf.org/MicpuC2/MicpuC2.home.html">http://genome.jgi-psf.org/MicpuC2/MicpuC2.home.html</a>	Dados não publicados, 2008

Todos os ESTs utilizados foram obtidos do banco de dados “TIGR Plant Transcripts Assemblies” (Childs *et al.*, 2006; <http://plantta.tigr.org/>). As seqüências ESTs deste banco de dados já se encontram montada em clusters, utilizando os programas TGICL (Perteau *et al.*, 2003), Megablast (Zhang *et al.*, 2000) e CAP3 Assembler (Huang & Madan, 1999). No total, 27 bancos de ESTs foram utilizados, contendo ao todo 444402 seqüências. A lista dos bancos de ESTs utilizados por ser vista na Tabela 3.

### **2.2.3. Construção de bancos de seqüências**

Os conjuntos preditos de proteínas provenientes dos genomas completos descritos na Tabela 2 foram baixados via internet e utilizados para gerar um único banco chamado Viridiplantae 1.0 (Seqüências de aminoácidos). Além dos genomas completos bancos de ESTs clusterizados descritos na Tabela 3 foram baixados e agrupados em um banco específico para as diferentes linhagens (Seqüências de nucleotídeos): Gimnospermas (Pinophytas, Cycadophytas, Gnetophytas e Ginkgophytas), Tracheophytas sem semente (Pteridophytas e Lycopodiophytas), Musgos (Bryophytas e Marchantiophytas), Eudicotiledôneas, Monocotiledôneas, e um banco contendo as demais Angiospermas (Amborellales, Magnoliales, Nymphaeales, Laurales, Piperales). Os bancos foram formatados, utilizando o programa formatdb (componente do pacote Blast 2.2.17), para sua utilização em buscas via Blast (Altschul *et al.*, 1997).

**Tabela 3.** Bancos de ESTs utilizados.

Nome científico	Nome comum	Data de deposição das seqüências	Total de ESTs (Cluster + Singletons)	Classificação
<i>Adiantum capillus-veneris</i>	Avenca	28/9/06	7775	Pteridophyta
<i>Amborella trichopoda</i>	-	28/9/06	6649	Angiosperma - Amborellales (Ordem)
<i>Ceratodon purpureus</i>	-	28/9/06	1044	Bryophyta
<i>Ceratopteris richardii</i>	-	28/9/06	4492	Pteridophyta
<i>Chamaecyparis obtusa</i>	Cipreste	5/6/06	4061	Pinophyta
<i>Cryptomeria japonica</i>	Cedro japonês	5/6/06	9098	Pinophyta
<i>Cycas rumphii</i>	Cica	28/9/06	4335	Cycadophyta
<i>Ginkgo biloba</i>	Ginkgo	28/9/06	4178	Ginkgophyta
<i>Gossypium hirsutum</i>	Algodão	28/9/06	70667	Eudicotiledônea - Fam. Malvaceae
<i>Liriodendron tulipifera</i>	Poplar amarelo	5/6/06	7087	Angiosperma - Magnoliales (Ordem)
<i>Marchantia polymorpha</i>	Hepática	28/9/06	10721	Marchantiophyta
<i>Nuphar advena</i>	Lírio-d'água	5/6/06	8144	Angiosperma - Nymphaeales (Ordem)
<i>Persea americana</i>	Abacate	28/9/06	6700	Angiosperma - Laurales (Ordem)
<i>Physcomitrella patens</i>	-	28/9/06	45149	Bryophyta
<i>Picea abies</i>	Abeto norueguês	1/6/07	5204	Pinophyta
<i>Picea engelmannii x Picea glauca</i>	-	28/9/06	14201	Pinophyta
<i>Picea glauca</i>	Abeto branco	5/6/06	49412	Pinophyta
<i>Picea sitchensis</i>	Abeto Sitka	1/6/07	25425	Pinophyta
<i>Pinus pinaster</i>	Pinheiro-bravo	5/6/06	13067	Pinophyta
<i>Pinus taeda</i>	Pinheiro	5/6/06	78873	Pinophyta
<i>Populus trichocarpa</i>	Álamo	5/6/06	31082	Eudicotiledônea - Fam. Salicaceae
<i>Pseudotsuga menziesii</i>	Pinheiro	1/6/07	12074	Pinophyta
<i>Saruma henryi</i>	-	5/6/06	6956	Angiosperma - Piperales (Ordem)
<i>Selaginella lepidophylla</i>	Planta da ressurreição	1/6/07	2861	Lycopodiophyta
<i>Taiwania cryptomerioides</i>	Conífera	28/9/06	778	Pinophyta
<i>Tortula ruralis</i>	-	28/9/06	7689	Bryophyta
<i>Welwitschia mirabilis</i>	-	5/6/06	6680	Gnetophyta

#### 2.2.4. Estabelecimento da metodologia para obtenção de seqüências codificantes para as hidrolases que constituem o mecanismo de degradação de xiloglucano

A identificação das seqüências codificantes para as enzimas ativas na degradação de xiloglucano foi realizada utilizando o programa Blastp para os bancos de proteínas, tBlastn para as seqüências genômicas e bancos de EST.

Para a identificação das seqüências codificantes para XTH utilizamos as 33 seqüências presentes no genoma de *Arabidopsis thaliana* descritas em Rose *et al.* (2002) e as 29 seqüências de arroz descritas em Yokoyama *et al.* (2004) como *queries* das buscas usando Blast. Para  $\beta$ -galactosidase, utilizamos as 17 seqüências codificantes do genoma de *Arabidopsis*, descritas em XX e as 15 seqüências de arroz descritas em Ahn *et al.* (2007). A  $\beta$ -D-glucosidase de *Tropaeolum majus* (CAA07070), descrita em Crombie *et al.* (1998) foi utilizada para buscar seqüências semelhantes nos diversos bancos. Para buscar  $\alpha$ -xilosidases utilizamos a seqüência da  $\alpha$ -xilosidase de *Tropaeolum majus* (CAA10382) e a seqüência codificante do gene AtXYL1 (Sampedro *et al.*, 2001) de *Arabidopsis*. A seqüência da proteína codificada pelo gene AtFXG1, descrita em de la Torre *et al.* 2002 foi utilizada para buscar por  $\alpha$ -fucosidases.

As seqüências que apresentavam um *e-value* inferior a  $1e-4$  eram selecionadas, utilizando o programa BTF, (*Blast to Fasta*) (Del Bem & de Mello, 2008 – não publicado – Anexo 1) em um arquivo FASTA e alinhadas por ClustalW (Chenna *et al.*, 2003) junto com as seqüências referencia, e então falsos positivos eram eliminados por conferência visual dos alinhamentos. Para as seqüências de ESTs utilizamos o mesmo corte ( $1e-4$ ), e a tradução foi feita utilizando o programa ETToll, (*ESTs Translator Tool*) (Del Bem & de Mello, 2008 – não publicado – Anexo 2), que gera um arquivo FASTA, contendo as seqüências traduzidas, que foi alinhado por ClustalW junto com as seqüências referencia para remoção manual dos falsos positivos. Correções de *frameshifts* (adições ou deleções que causam mudança no frame de leitura de parte da proteína encodada por um EST) foram realizadas manualmente.

As seqüências obtidas de genomas completos ou de ESTs eram alinhadas juntamente com as seqüências peptídicas de referência e uma árvore filogenética era

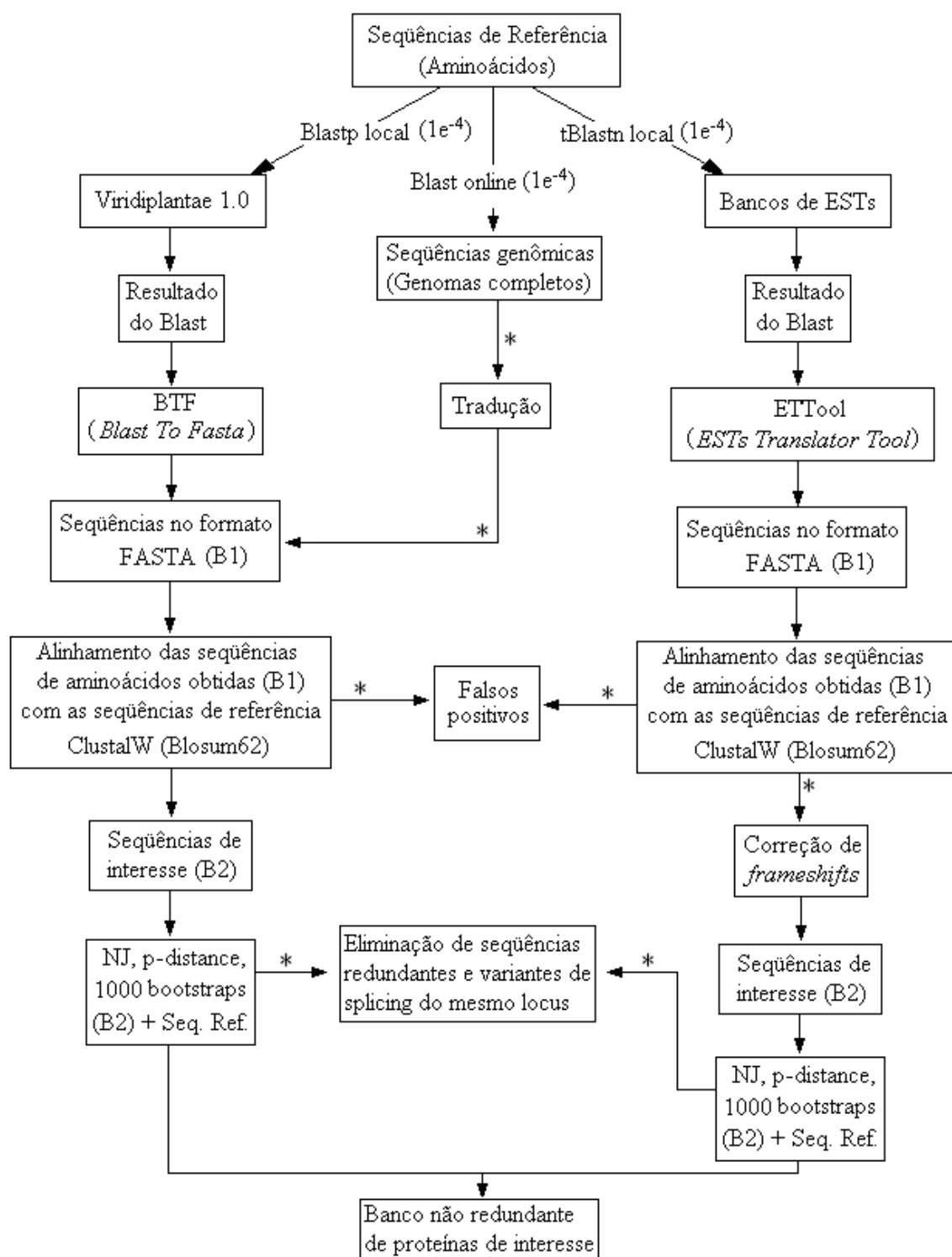
gerada utilizando Neighbor-Joining (NJ) para construção da topologia e p-distance como o matriz de distância, com 1000 réplicas de *bootstraps*. Os possíveis parálogos eram então conferidos manualmente para remoção de seqüências redundantes (similaridade maior ou igual a 97% - Vincentz *et al.*, 2003) e formas variantes de *splicing* do mesmo locus.

O protocolo para obtenção das seqüências aparece, em detalhe, na figura 13.

### **2.2.5. Análise filogenética**

As seqüências de aminoácidos foram alinhadas com ClustalW (Thompson *et al.*, 1994), utilizando as condições padrão do programa. As análises por distância foram realizadas por *Neighbor-Joining* (NJ) com o programa MEGA4 (Tamura *et al.*, 2007). As distâncias foram obtidas usando *p-distance* e a matriz de distâncias PAM (Dayhoff *et al.*, 1978). Foram realizadas 1000 replicatas de *bootstrap*.

As análises por Máxima Parcimônia (MP) foram realizadas utilizando o programa MEGA4 (Tamura *et al.*, 2007). As opções de busca foram: CNI (*Close neighbor-interchange*) (level=1) com árvore inicial gerada por adições randômicas (10 replicatas). Foram realizadas 1000 replicatas de *bootstrap*.



**Figura 13.** Esquema do protocolo bioinformático para construção de bancos não redundantes de seqüências de proteínas. As setas marcadas com asterisco (\*) representam processos que foram realizados manualmente. Os programas BTf e ETTTool (Del Bem & De Mello, 2008) foram desenvolvidos especificamente para este protocolo e os códigos-fonte encontram-se nos Anexos 1 e 2. Os bancos obtidos ao final do protocolo foram utilizados nas análises filogenéticas.

### **2.2.6. Identificação de Possíveis Grupos de Ortólogos (PoGO)**

A identificação de PoGOs, ao longo das famílias de glicosil hidrolases ativas sobre xiloglucano, foi realizada através da avaliação das relações evolutivas estabelecidas por NJ e MP a partir dos alinhamentos das seqüências de aminoácidos obtidas através do protocolo da Figura 13.

A identificação de um PoGO obedece a três premissas: (I) cada PoGO deve possuir um *bootstrap* maior que 50%; (II) o PoGO deve possuir representantes de pelo menos 2 genomas completos, assumindo que os conjuntos completos de gene das enzimas de interesse foram obtidos nestes genomas (Tabela 2); (III) a filogenia obtida deve refletir a filogenia das espécies (Vincentz *et al.*, 2003).

## **3. Resultados e Discussão**

### **3.1. Bibliotecas de cDNA de Jatobá**

#### **3.1.1. Otimização dos Protocolos de extração de RNA Total e síntese de cDNA**

Os protocolos que deram resultados satisfatórios na integridade e quantidade de RNA total foram os descritos por Gesteira et al. (2003) com modificações para folha primária e raiz e o descrito por Daohong et al. (2004) com modificações para os cotilédones. A mudanças nos protocolos consistiram na adição de duas etapas a mais de fenol:clorofórmio:álcool isoamílico para aumentar a pureza do RNA e limitar significativamente as degradações freqüentes do RNA que ocorrem durante a extração. A plântula de Jatobá apresenta muitos metabólitos secundários, como compostos

fenólicos, terpenóides e carboidratos complexos de reserva como xiloglucano, o que torna difícil extrair RNA com boa integridade.

Para a extração de RNA total de folhas secundárias e caule (epicótilo e hipocótilo) o reagente Concert (Invitrogen), seguindo as instruções do fabricante, se mostrou eficiente.

Elaboramos condições de crescimento para as plântulas de Jatobá visando aumentar a representatividade de grupos de genes menos expressos e diminuir a representatividade de grupos de genes muito expressos. A primeira condição foi chamada de Condição Controle (CC) onde as plântulas cresceram com foto período de 12 horas sem qualquer intervenção. Na segunda condição cobrimos todas as folhas (FC) com papel alumínio por três dias antes da coleta, com a intenção de minimizar a expressão de mRNAs relacionados à fotossíntese. Na terceira condição deixamos, por 30 dias, as plântulas crescendo em escuro contínuo (EC), buscando a maximização da expressão de mRNAs relacionados à degradação das reservas (cotilédones) devido à ausência das reservas de amido nas folhas, decorrente da ausência completa de fotossíntese. Na quarta condição as plântulas sofreram estresse mecânico (PM) nas folhas, caule (hipocótilo e epicótilo) e nos cotilédones. O estresse mecânico foi feito com um bisturi, produzindo cortes por toda a extensão destes tecidos. O tratamento visou maximizar a expressão de mRNAs relacionados a injúrias e reparo dos tecidos. Para todos os tratamentos foram realizadas duas coletas, uma às 6 horas e outra às 18 horas. Estes horários foram escolhidos baseados no fato de 30% a 50% dos genes em *Arabidopsis thaliana* apresentam variação circadiana pelo equilíbrio entre acumulação e degradação das reservas de amido presentes nas folhas. Segundo Oliver Bläsing *et al.* (2005) nos extremos do ciclo circadiano (6 horas e 18 horas) é onde ocorre o pico de



expressão para a maioria dos genes. Desta forma cada tratamento ainda engloba os dois pontos de maior variação de transcritos. Toda a abordagem tem como intenção maximizar a variabilidade de mRNAs nas bibliotecas, aumentando assim sua qualidade e representatividade de transcritos.

### **3.1.2. Obtenção de RNA Poli A+, síntese de cDNA, LD-PCR e clonagem**

Uma vez padronizados os protocolos de extração de RNA total pudemos obter quantidade satisfatória de RNA total íntegro, das diferentes condições propostas na tabela 1 (ver figura 14).

Uma vez obtidos os RNAs totais, quantidades equimolares foram misturadas obtendo um mix de RNA total para cada tecido (Figura 15). Totalizando, em miligramas:

**Folhas** (eófilo e folhas secundárias): 0,76mg de RNA total

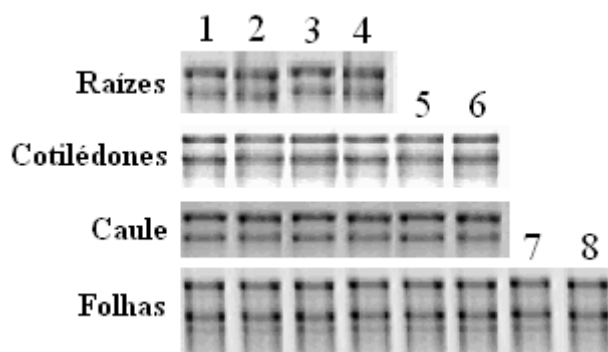
**Caule** (hipocótilo e epicótilo): 0,66mg de RNA total

**Cotilédone**: 0,71mg de RNA total

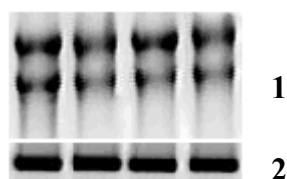
**Raízes**: 0,54mg de RNA total

Para a obtenção de RNA Poli A+ a partir dos mix utilizamos o kit Oligotex mRNA Purification System (Qiagen). Cada coluna tem a capacidade máxima de 650µl contendo até 3mg de RNA total. Foram recuperados (RNA Poli A+), aproximadamente, 8,0µg de folhas, 5,6µg de Raízes, 7,8µg de Cotilédones e 6,5µg de Caule de RNA Poli A+. O que corresponde, respectivamente, a 1,05%, 1,03%, 1,10% e 0,98% do RNA total que entrou na coluna.

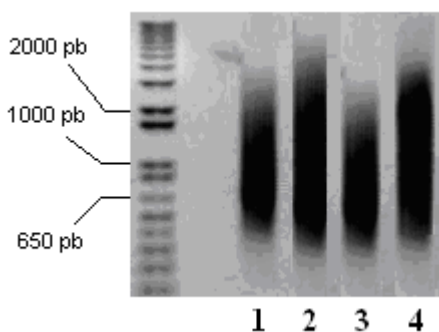
Para sintetizar a primeira fita de cDNA utilizamos a enzima SuperScriptII RT (Invitrogen) que possui atividade de transferência de três citosinas na parte 3' da primeira fita. O primer iniciador da transcrição reversa foi desenhado de forma que anelasse na cauda Poli A+ dos transcritos e ainda contivesse a seqüência *attB2* que é alvo da enzima Clonase BP (Invitrogen). Utilizamos 10ng de cDNA fita simples numa reação de PCR de 24 ciclos onde o primer reverso era o *attB2* Oligo (d)T e o direto *attB1* Poly G, este último anelando na cauda Poli C deixada pela SuperScriptII RT (Material e Métodos). Como resultado do LD-PCR obtivemos cerca de 300ng de cDNA flanqueado pelas seqüências de recombinação homóloga do vetor pDONR-222. Os cDNAs eram separados por tamanho em gel de agarose (Figura 16) e então a região entre 500pb e 3000pb era excisada. Estas seqüências *attB1* e *attB2* são alvos da enzima Clonase BP (Invitrogen) e permitem uma clonagem orientada via recombinação homóloga. Utilizando 100ng de cDNA dupla fita e 250ng de vetor pDONR-222 foi possível gerar bibliotecas com os títulos de  $\sim 1,0 \times 10^6$  clones para a biblioteca de Caule,  $\sim 1,2 \times 10^6$  clones para a biblioteca de Cotilédones,  $\sim 1,4 \times 10^6$  clones para a biblioteca de Raiz e  $\sim 1,7 \times 10^6$  clones para a biblioteca de Folhas. Considerando as dificuldades de obtenção de RNA integro de Jatobá, o sistema para gerar as bibliotecas que padronizamos se mostrou eficiente com quantidades muito pequenas de RNA Poli A+ o que deixa como perspectiva futura a possibilidade de construção de bibliotecas de outros tecidos da planta de Jatobá adulta (flores, por exemplo) ou outras condições de crescimento, mesmo que a eficiência de extração de RNA total não seja boa.



**Figura 14.** Análise da integridade dos RNAs totais de Cotilédone, Caule, Folhas e Raízes. Géis de Agarose (1%) em condição desnaturante (6% de Formaldeído) de RNA Total (desnaturado em tampão contendo Brometo de Etídio) de todos os tecidos e condições descritas na Tabela 1 (1 – Condição controle 6hrs, 2 – Condição controle 18hrs, 3 – Escuro contínuo 6hrs, 4 – Escuro contínuo 18hrs, 5 – Plantas machucadas 6hrs, 6 – Plantas machucadas 18hrs, 7 – Folhas cobertas 6hrs e 8 – Folhas cobertas 18hrs).



**Figura 15.** 1 - Gel de agarose (1%) em condição desnaturante (6% de Formaldeído) mostrando a integridade dos Mix de RNA obtidos para cada tecido (ver Tabela 2). 2 – Gel de agarose (1%) mostrando o produto de amplificação do cDNA de Actina para os diferentes Mix de RNA.

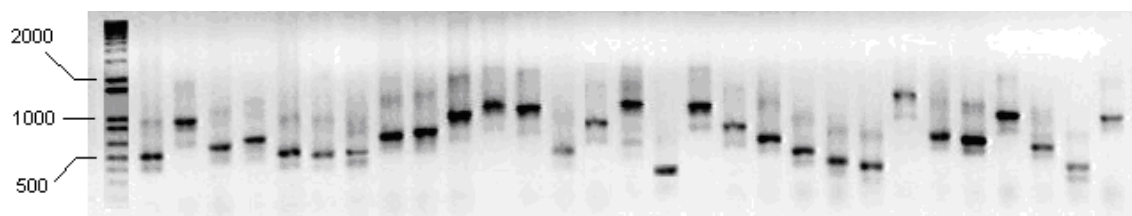


**Figura 16.** cDNAs separado por tamanho em gel de agarose (2%) corado com Brometo de Etídio. 1- Cotilédones, 2-Folhas, 3-Raízes, 4-Caule.

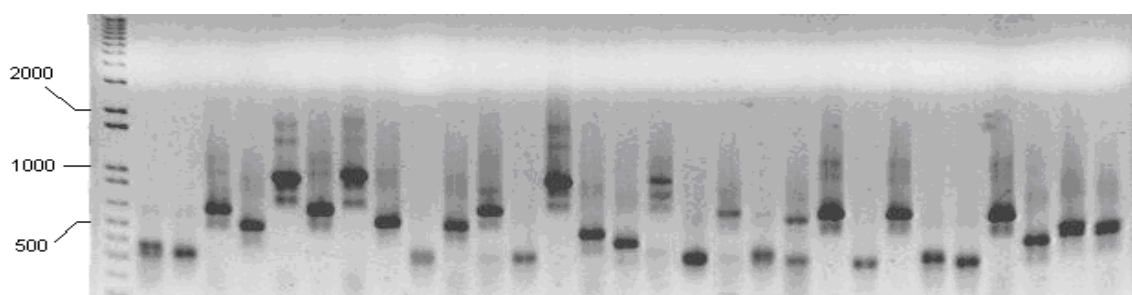
### 3.1.3. Bibliotecas de cDNA

As quatro bibliotecas obtidas foram conservadas em alíquotas de 50µl de um total de 1050ul. As células foram plaqueadas em meio de cultura seletivo, com Kanamicina, e colônias isoladas foram picadas em placas de 96 poços contendo meio líquido apropriado, crescidas e armazenadas em biofreezer. Fizemos, também, PCR de colônias isoladas de cada biblioteca para avaliação do percentual de clones positivos bem como a diversidade de tamanho dos insertos. A biblioteca de caule apresentou tamanho médio dos fragmentos, considerando 29 clones, de aproximadamente 780 pares de base, com fragmentos variando entre 400 e 1650 pares de base (Figura 17). A biblioteca de cotilédone apresentou um tamanho médio de fragmentos, para 29 clones, de 520 pares de base, com fragmentos variando entre 300 e 920 pares de base. (Figura 18). A biblioteca de raiz apresentou, para 29 clones, tamanho médio de fragmentos de 1000 pares de base com fragmentos variando entre 400 e 2000 pares de base (Figura 19) e a biblioteca de folha apresentou tamanho médio de 620 pares de base, com fragmentos entre 350 e 1200 pares de base (Figura 20). Todos os clones avaliados foram positivos, o que se deve à seleção dupla que está presente no vetor pDONR-222: o gene *ccdB*, inserido no vetor entre as seqüências de recombinação *attP1* e *attP2*, codifica uma proteína que interage com os complexos DNA-topoisomerase II causando um efeito letal à célula (em *E. coli* a mutação *gyrA462* confere total resistência a *ccdB* - Bernard & Couturier, 1992) que mata a célula de *E. coli* DH10B transformada com o vetor não recombinado. Quando a recombinação acontece o cassete contendo o gene *ccdB* e o gene de resistência a Chloramphenicol (CmR) são substituídos pelo cDNA, o que “desarma” a letalidade do vetor, e a célula transformada torna-se resistente à Kanamicina o que permite ser selecionada em meio com este antibiótico. As células que

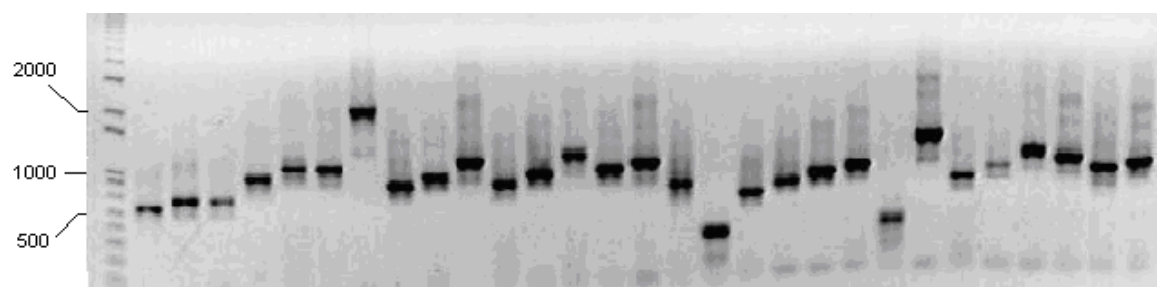
crecem nas placas com Kanamicina depois da eletroporação, pelo menos teoricamente, são clones positivos. Nossas bibliotecas mostraram 100% de clones positivos, o que está de acordo com o esperado dentro do sistema utilizado.



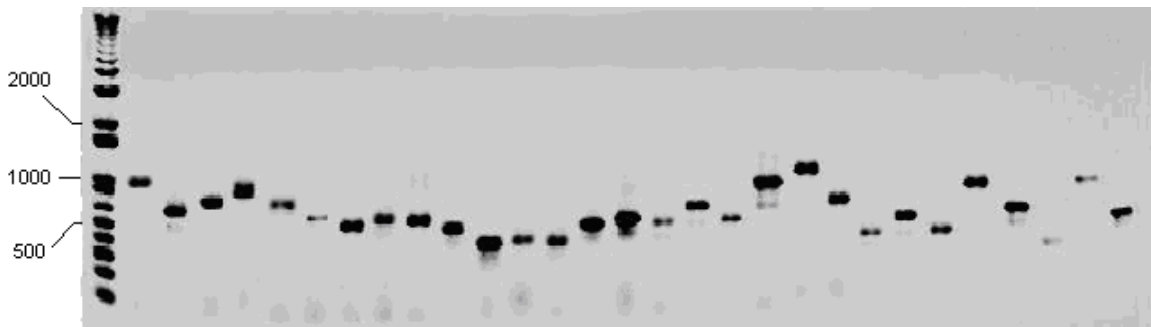
**Figura 17.** Estimativa do tamanho dos cDNA clonados na biblioteca de Caule. Gel de agarose (1%) corado com Brometo de Etídio. Fragmentos de PCR, de 29 colônias, correspondentes aos insertos clonados na biblioteca.



**Figura 18.** Estimativa do tamanho dos cDNA clonados na biblioteca de Cotilédones. Gel de agarose (1%) corado com Brometo de Etídio. Fragmentos de PCR, de 29 colônias, correspondentes aos insertos clonados na biblioteca.



**Figura 19.** Estimativa do tamanho dos cDNA clonados na biblioteca de Raízes. Gel de agarose (1%) corado com Brometo de Etídio. Fragmentos de PCR, de 29 colônias, correspondentes aos insertos clonados na biblioteca.



**Figura 20.** Estimativa do tamanho dos cDNA clonados na biblioteca de Folha. Gel de agarose (1%) corado com Brometo de Etídio. Fragmentos de PCR, de 29 colônias, correspondentes aos insertos clonados na biblioteca.

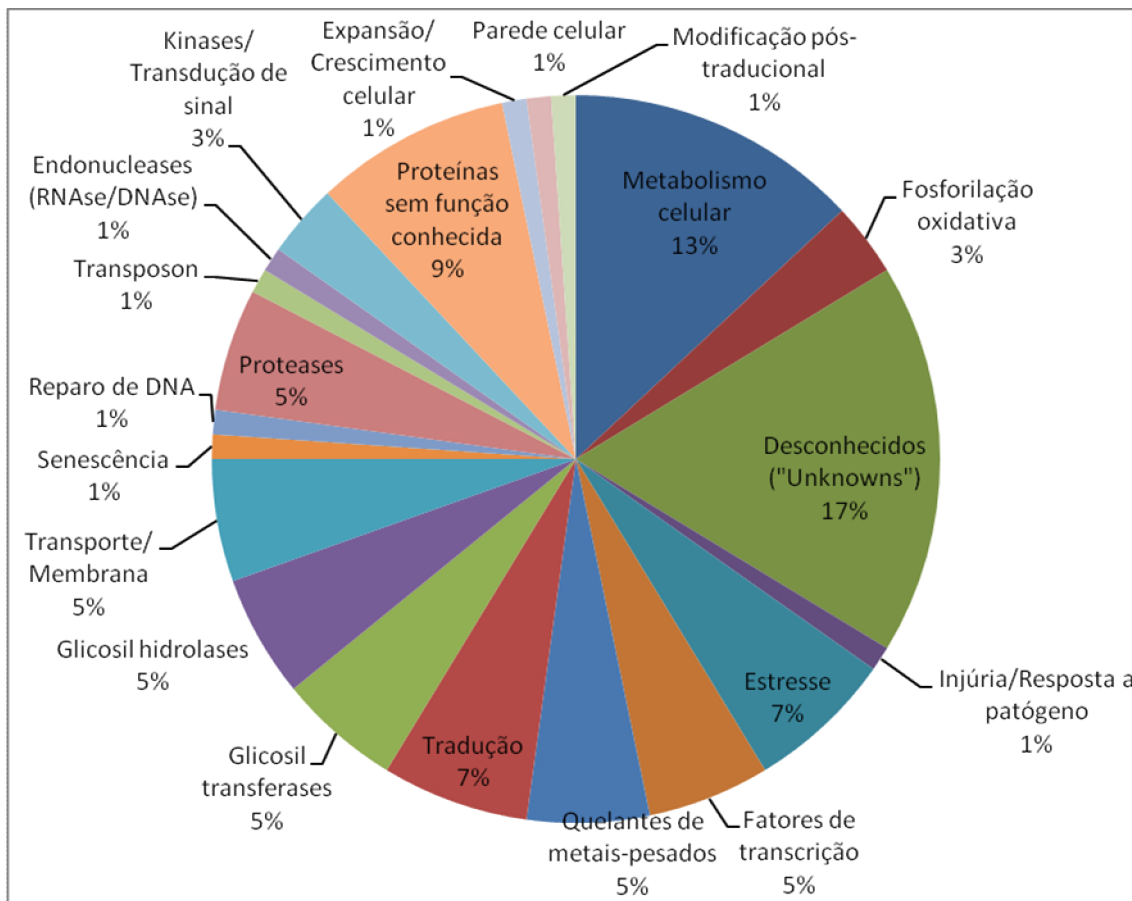
### 3.1.4. Seqüenciamento preliminar

Com objetivo de fazer uma avaliação preliminar das bibliotecas, realizamos o seqüenciamento de 82 clones da biblioteca de caule e 92 da biblioteca de raiz. Seqüenciamos ainda oito clones da biblioteca de cotilédones e 16 clones da biblioteca de folhas.

Com 82 *reads* da biblioteca de caule obtivemos 51 seqüências protéicas (ou fragmentos), sendo 45 *singletons* e seis *contigs*, num total de 66 seqüências codificantes. Quatro reações de seqüenciamento falharam do total de 86 reações. Um total de 16 seqüências não teve *hit* com qualquer transcrito (Figura 21). O índice de novidade pode ser estimado, em porcentagem, por:

$$[(\text{Contigs} + \text{Singletons}) / \text{Reads}] \times 100$$

Considerando o *pool* de seqüências obtidas para a biblioteca de caule, estimamos o índice de novidade em 62,2%. Vale lembrar que este valor é apenas preliminar, seria necessário seqüenciar mais clones, para que se possa ter uma avaliação mais precisa.



**Figura 21.** Distribuição relativa, em categorias funcionais, dos 82 ESTs obtidos da biblioteca de caule.

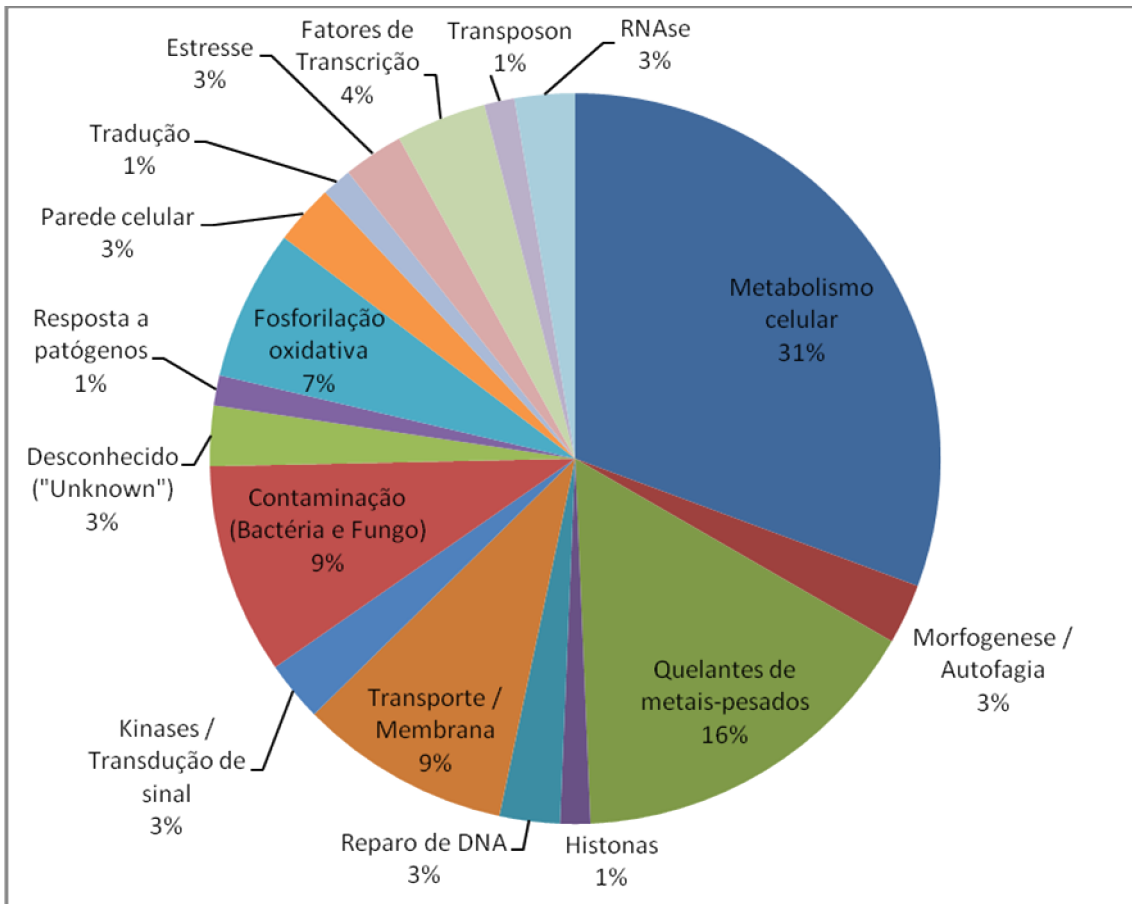
Os 92 reads da biblioteca de raiz permitiram obter 67 seqüências codificantes que correspondem a 40 diferentes fragmentos de proteínas (5 contigs e 35 singlets), o que resultou num índice de novidade de 43,5%. A distribuição relativa, por participação em diversos processos celulares, dos ESTs de raiz, pode ser vista na Figura 22. Das 96 reações de seqüenciamento quatro não tiveram incorporação. Seis seqüências tiveram *hit* com seqüências de cDNA de bactérias e uma com fungo, o que provavelmente se deve ao fato de bactérias e fungos estarem presente no substrato no qual as raízes estavam enterradas. Uma seqüência teve 96% de identidade ao longo de 238pb com um clone genômico (mth2-12p22) de *Medicago truncatula*, uma leguminosa. Este fragmento não tem similaridade com qualquer seqüência de proteína disponível no

GeneBank, o que sugere que se trata de uma seqüência genômica de Jatobá, que foi amplificada por anelamento parcial dos primers usados na amplificação do *pool* de cDNAs que precede a clonagem via recombinação homóloga (ver Material e Métodos). Será necessário o seqüenciamento de um número maior de clones para avaliar a freqüência de fragmentos genômicos nesta e nas outras bibliotecas. Um total de 16 seqüências foi descartado por “*slippage*” ou “derrapagem”. O fenômeno de “*slippage*” consiste na adição de uma seqüência mononucleotídica (repetição de uma única base) num resultado de seqüenciamento. Isto provavelmente ocorre pela presença de regiões mononucleotídicas, o que faz a DNA polimerase “derrapar” sobre a fita molde. A “derrapagem” ocorre devido à desconexão das fitas molde e o produto de extensão da DNA polimerase ao longo de uma região mononucleotídica e subsequente re-hibridização, o que resulta na formação de produtos seqüenciáveis de vários tamanhos. Esses produtos aparecem como sinais misturados formando um rastro mononucleotídico na corrida de seqüenciamento (Figura 23).

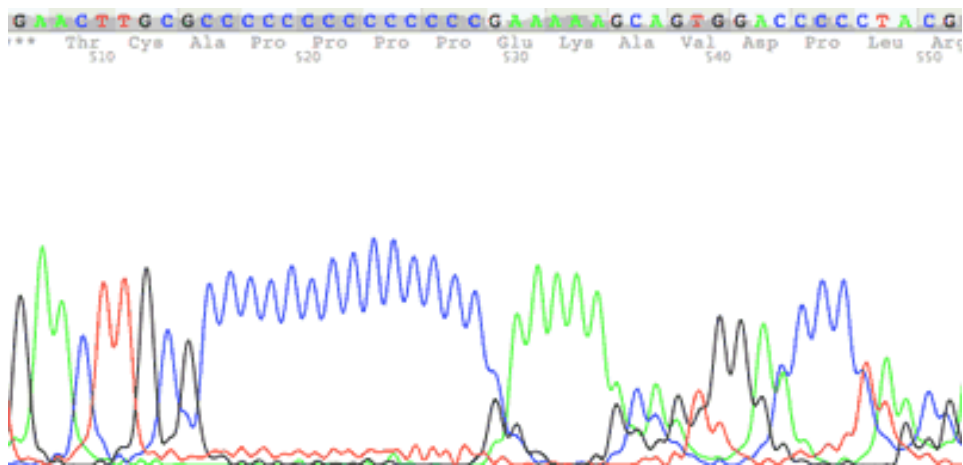
A biblioteca de cotilédones, por hora, teve somente oito clones seqüenciados. Obtivemos três seqüências codificantes, sendo dois *singletons* e um *contig* formado por duas seqüências. Uma seqüência não apresentou paridade com qualquer proteína ou EST e três foram descartadas por “*slippage*”. As três seqüências codificantes obtidas correspondem a uma Ribonuclease H, uma Acetato-CoA ligase e uma *Glycine-Rich protein* (proteína estrutura de parede celular).

Apenas 16 clones foram seqüenciados inicialmente na biblioteca de folhas. Com estas seqüências obtivemos 14 seqüências peptídicas, sendo todos *singletons*, com novidade estimada em 87,5% (Figura 24). Duas seqüências não apresentaram semelhança com qualquer proteína ou seqüência genômica disponível.

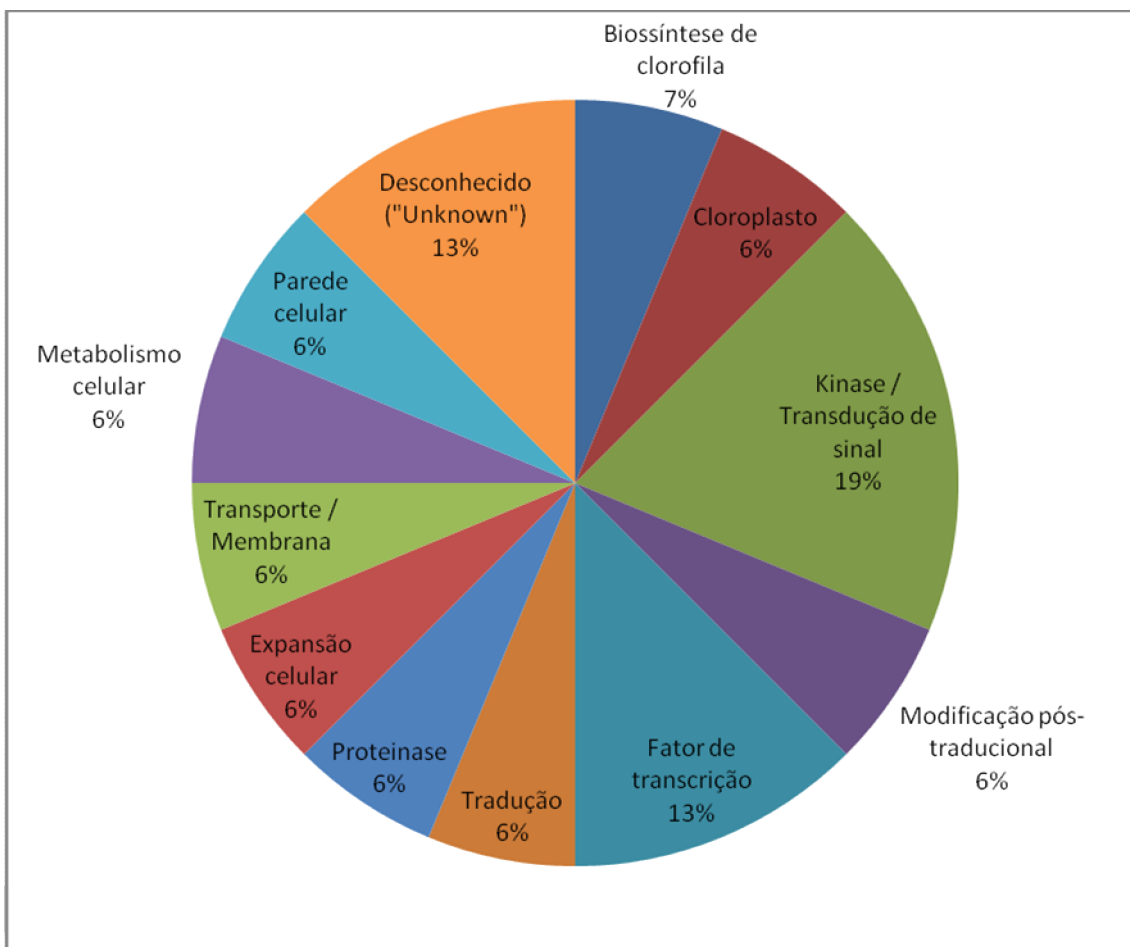




**Figura 22.** Distribuição relativa, em categorias funcionais, de 75 ESTs da biblioteca de raiz. As 16 sequências que sofreram “*slippage*” foram retiradas da análise.

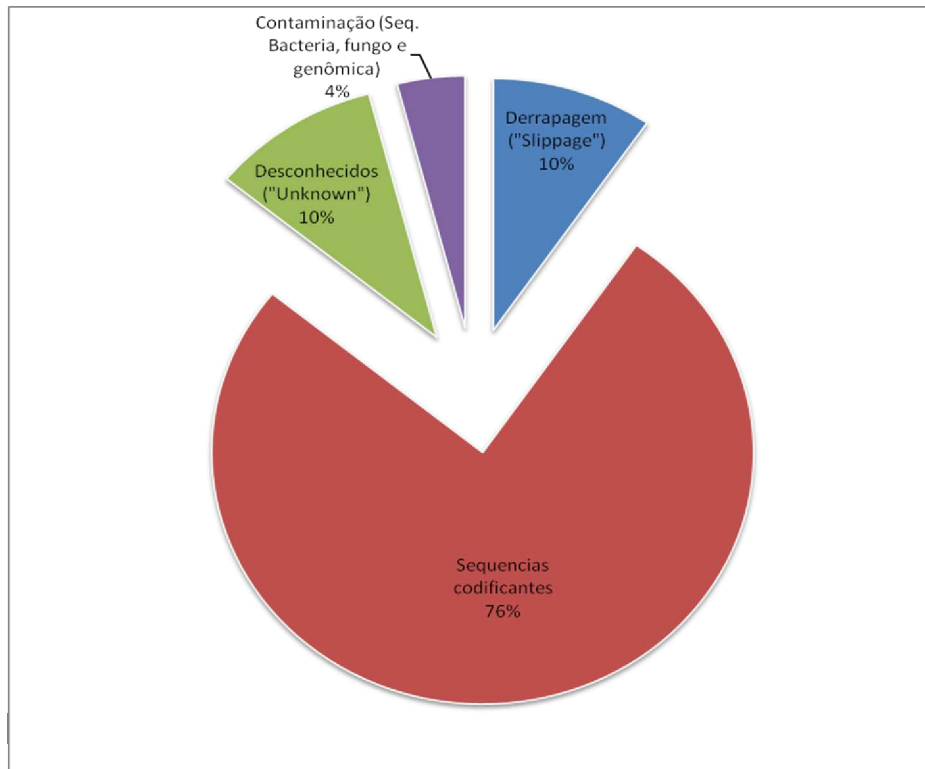


**Figura 23.** Cromatograma mostrando uma sequência com *slippage* (repetição de C).

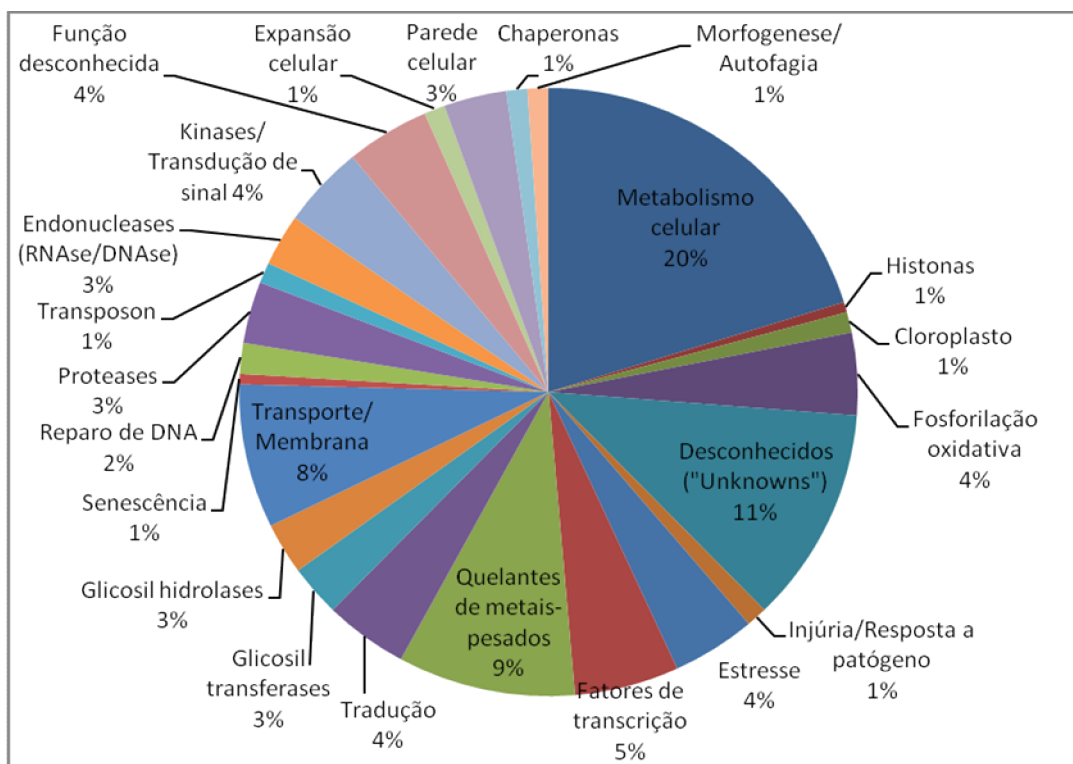


**Figura 24.** Distribuição relativa, em categorias funcionais, de 16 ESTs obtidos da biblioteca de folhas.

Concluindo, obtivemos 198 seqüências, das quais 19 foram descartadas por “*slippage*”. Dos 179 ESTs restantes, 151 apresentaram similaridade quando comparados com bancos de proteínas de genomas vegetais sendo agrupados em 103 *clusters* de proteínas (91 *singletons* e 12 *contigs*). Um total de 20 ESTs não pode ser identificado por comparações com seqüências genômicas e bancos de proteínas, e oito ESTs foram identificados como contaminação (seis cDNAs de bactéria, um de fungo e uma seqüência genômica de Jatobá) (Figura 25). O índice de novidade, considerando todas as seqüências, é de 52,0%. A distribuição, por função molecular, das 103 seqüências peptídicas obtidas pode ser vista na Figura 26.



**Figura 25.** Distribuição da 198 sequências obtidas ao longo de quatro categorias. Aproximadamente 76% das sequências obtidas correspondiam a transcritos codificantes.



**Figura 26.** Distribuição relativa, por função, dos 150 ESTs codificantes obtidos.

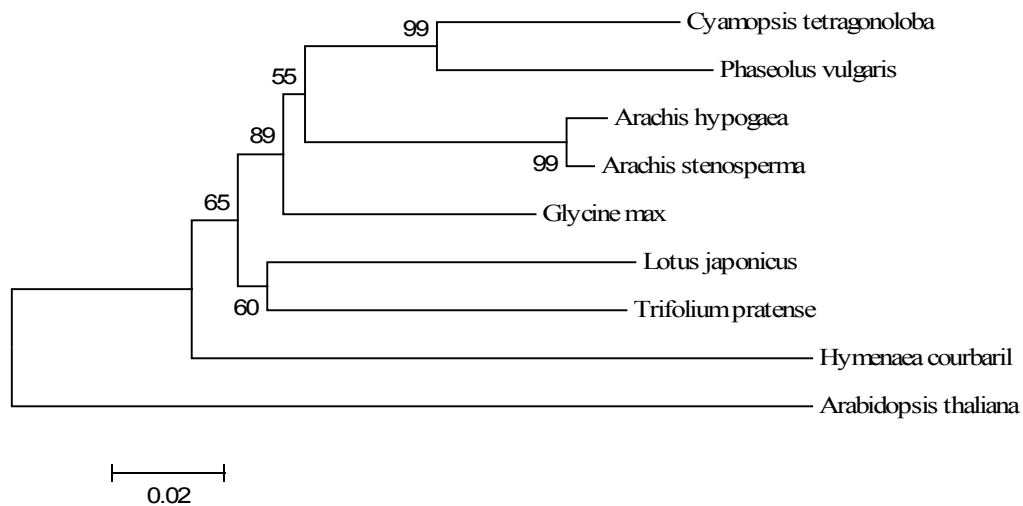
### 3.1.5. Posição filogenética do Jatobá na família Leguminosae

A família Leguminosae, também chamada de Fabaceae, engloba, atualmente sete subfamílias: Cercideae, Detarieae sensu lato, Cassieae sensu lato (pro parte), Caesalpinieae sensu lato e sensu stricto, Cassieae sensu stricto, Mimosoideae e Papilionoideae.

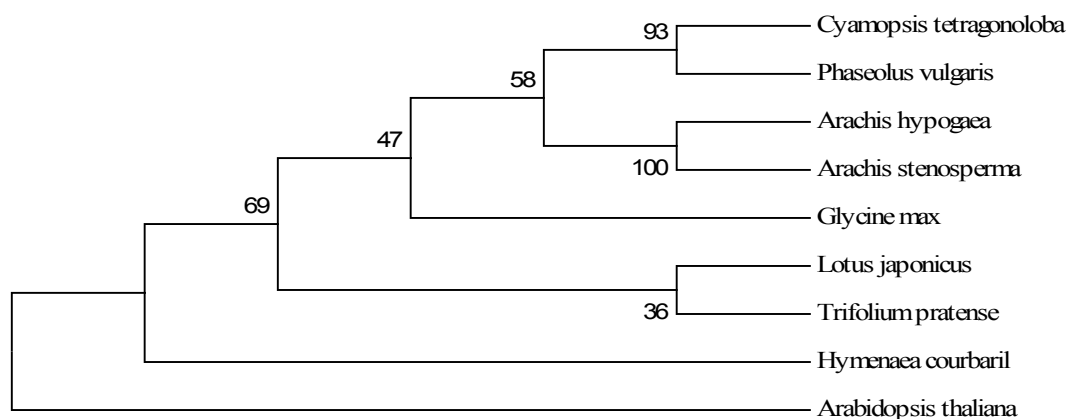
O gênero *Hymenaea*, tradicionalmente pertence à subfamília Caesalpinioideae e à tribo Detarieae. No entanto as relações entre as subfamílias de Leguminosae não são muito claras (Simpson *et al* 2003) e mais recentemente foi proposto que a tribo Detarieae saia da subfamília Ceasalpinioideae e passe a constituir a subfamília Detarieae sensu lato (Mackinder 2005). Dentro de Detarieae sensu lato o gênero *Hymenaea* está incluído no clado Detarieae sensu stricto. O clado Detarieae sensu stricto apresenta forte suporte filogenético, mas sua relação com os outros clados de Detarieae sensu lato e desde último com as outras subfamílias de Leguminosae ainda permanecem obscuros (Herendeen *et al.*, 2003; Wieringa & Gervais, 2003). Nosso trabalho pode contribuir para melhorar a resolução filogenética do gênero *Hymenaea* dentro do contexto de família e subfamília através da obtenção de dados moleculares inéditos. No banco de transcritos internacional “*The TIGR Plant Transcript Assemblies database*” (Childs *et al* 2006) 16 espécies de Leguminosae tem bancos de ESTs montados, no entanto todas pertencem à subfamília Papilionoideae (como soja e *Medicago truncatula*), portanto as seqüências de Jatobá são os primeiros ESTs de uma Leguminosae de outra subfamília.

Utilizamos duas proteínas ribossomais, L10 e S23, e a seqüência da subunidade G da NADH dehydrogenase, de Jatobá para produzir um bloco de 508 aminoácidos conservados entre leguminosas que aparecem nas figuras 27 e 28. O alinhamento foi

utilizado para gerar uma árvore por Neighbor-Joining (Saitou & Nei, 1987; Figura 27). Na figura 28, árvore filogenética foi gerada pelo método da Máxima Parcimônia (Eck & Dayhoff - 1966). Ambas as análises filogenéticas confirmaram a posição externa da seqüência do Jatobá às espécies da subfamília Papilionoideae, que era o esperado. A Brassicales *Arabidopsis thaliana* foi utilizada com grupo externo.



**Figura 27.** Análise filogenética, por NJ, baseada na seqüência de aminoácidos dos genes L10, S23 e Subunidade G do complexo NADH dehydrogenase. A topologia foi inferida por Neighbor-Joining (Saitou & Nei, 1987), usando o modelo de substituição de nucleotídeos *p-distance* com 1000 replicatas de *bootstraps*.



**Figura 28.** Análise filogenética, por MP, baseada na seqüência de aminoácidos dos genes L10, S23 e Subunidade G do complexo NADH dehydrogenase. A topologia foi inferida por Máxima Parcimônia (Eck & Dayhoff, 1966) com 1000 replicatas de *bootstraps*.

## **3.2. Análise filogenética das enzimas hidrolíticas de xiloglucano**

### **3.2.1. Identificação de seqüências codificantes para as enzimas ativas sobre xiloglucano**

A comparação dos genes de diferentes genomas permite uma classificação em família de genes homólogos, as quais incluem ortólogos e parálogos (Tatusov *et al.*, 1997). Um grupo de genes ortólogos assume que os genes deste grupo evoluíram de um ancestral único a partir de eventos de especiação e duplicação. Como genes ortólogos são considerados representantes de uma mesma função ancestral, a organização em grupos de ortólogos deve permitir racionalizar estudos funcionais de genes em diversas linhagens evolutivas (Tatusov *et al.*, 1997; Martinez-Castilla & Alvarez-Buylla, 2003; Nam *et al.*, 2004).

Os genomas completos disponíveis atualmente (Tabela 2) representam novas oportunidades para o estudo exaustivo da evolução de famílias multigênicas e suas funções associadas ao longo do reino Viridiplantae. Dentro deste contexto, iniciamos uma análise filogenética detalhada dos genes que compõem o mecanismo proposto para degradação de xiloglucano, de reserva ou estrutural, a fim de classificá-los em possíveis grupos de ortólogos (PoGOs).

Utilizando o protocolo descrito na figura 13 obtivemos um conjunto de 809 seqüências não redundantes, potencialmente ativas sobre xiloglucano. O conjunto total se divide em 293 XTHs, 133  $\beta$ -Galactosidases, 53  $\beta$ -Glucosidases, 24  $\alpha$ -Xilosidases, 22  $\alpha$ -Fucosidases I e 14  $\alpha$ -Fucosidases II.

**Tabela 4.** Número de seqüências, em cada organismo, para as diversas hidrolases.

<b>Genomas Completos</b>						
	<b>XTH</b>	<b><math>\beta</math>-Gal</b>	<b><math>\beta</math>-Glu</b>	<b><math>\alpha</math>-Xil</b>	<b><math>\alpha</math>-Fuc 1</b>	<b><math>\alpha</math>-Fuc 2</b>
<i>Arabidopsis thaliana</i> (eudicot)	33	17	8	2	2	1
<i>Glycine max</i> (eudicot)	35	26	14	1	5	3
<i>Populus trichocarpa</i> (eudicot)	35	20	9	2	7	4
<i>Vitis vinifera</i> (eudicot)	20	13	2	1	4	–
<i>Oryza sativa</i> (monocot)	31	15	5	1	3	2
<i>Sorghum bicolor</i> (monocot)	33	12	8	1	1	1
<i>Selaginella moellendorffii</i> (lycophyta)	16	4	2	1	–	1
<i>Physcomitrella patens</i> (bryophyta)	30	6	4	12	–	1
<i>Volvox carteri</i> (chlorophyta)	–	–	–	1	–	–
<i>Chlamydomonas reinhardtii</i> (chlorophyta)	–	–	–	1	–	–
<i>Ostreococcus lucimarinus</i> (chlorophyta)	–	–	–	–	–	–
<i>Ostreococcus tauri</i> (chlorophyta)	–	–	–	–	–	–
<i>Micromonas pusilla</i> (chlorophyta)	–	–	–	–	–	–
<b>Bancos de ESTs</b>						
	<b>XTH</b>	<b><math>\beta</math>-Gal</b>	<b><math>\beta</math>-Glu</b>	<b><math>\alpha</math>-Xil</b>	<b><math>\alpha</math>-Fuc 1</b>	<b><math>\alpha</math>-Fuc 2</b>
<i>Amborella trichopoda</i> (angio basal)	3	1	–	–	–	–
<i>Chamaecyparis obtusa</i> (coniferophyta)	–	2	–	–	–	–
<i>Chara vulgaris</i> (charophyta)	1	–	–	–	–	–
<i>Closterium peracerosum</i> (charophyta)	1	–	–	–	–	–
<i>Cryptomeria japonica</i> (coniferophyta)	4	–	–	–	–	–
<i>Cycas rumphii</i> (cycadophyta)	3	–	–	–	–	–
<i>Ginkgo biloba</i> (ginkgophyta)	–	1	–	–	–	–
<i>Helicosporidium sp.</i> (chlorophyta)	–	1	–	–	–	–
<i>Lilium longiflorum</i> (monocot)	–	–	–	–	–	1
<i>Liriodendron tulipifera</i> (magnoliids)	1	2	–	–	–	–
<i>Marchantia polymorpha</i> (marchantiophyta)	8	1	–	–	–	–
<i>Nuphar advena</i> (angio basal)	1	–	–	–	–	–
<i>Persea americana</i> (magnoliids)	1	–	–	–	–	–
<i>Picea engelmannii x stichensis</i> (coniferophyta)	1	–	–	–	–	–
<i>Picea glauca</i> (coniferophyta)	2	3	–	–	–	–
<i>Picea sitchensis</i> (coniferophyta)	10	1	–	–	–	–
<i>Pinus pinaster</i> (coniferophyta)	1	–	–	–	–	–
<i>Pinus taeda</i> (coniferophyta)	8	6	–	–	–	–
<i>Saruma henryi</i> (magnoliids)	3	2	–	–	–	–
<i>Selaginella lepidophylla</i> (lycophyta)	1	–	–	–	–	–
<i>Tortula ruralis</i> (bryophyta)	1	–	–	–	–	–
<i>Tropaeolum majus</i> (eudicot)	-	–	1	1	–	–
<i>Welwitschia mirabilis</i> (gnetophyta)	7	1	–	–	–	–
<b>Total de Seqüências</b>	293	133	53	24	22	14

### 3.2.2 Análise filogenética das XTH e organização em PoGOs.

Análises filogenéticas por NJ utilizando o conjunto de 293 XTHs obtidas, com um alinhamento de 103 posições de aminoácidos, permitiu a divisão da diversidade de XTHs em seis grupos de genes homólogos (Figura 29) e a formação de 19 PoGOs (Tabela 5; Figura 30). Um total de três seqüências pertencentes ao grupo III não

puderam ser classificadas em nenhum PoGO (Pt-XTH30 – 767014; Os-XTH29 – LOC\_Os09g23220; Sb-XTH25 – Sb02g024120) o que provavelmente se deve ao fato destas seqüências terem evoluído rapidamente, tendo se distanciado consideravelmente de suas seqüências ortólogas.

O primeiro trabalho a classificar a diversidade de XTHs (Rose *et al.*, 2002) encontrou três grupos de genes homólogos utilizando apenas as 33 seqüências de *Arabidopsis* (Grupos I, II e III). Um trabalho posterior (Yokoyama *et al.*, 2004) utilizando conjuntamente as seqüências de *Arabidopsis* e arroz sugeriu que os grupos I e II formariam um único grupo de genes homólogos (Grupo I/II), o que reduziu para dois o número de grupos de genes homólogos. No entanto nossos resultados mostraram que tanto no Grupo I quanto no Grupo II existem PoGOs que ocorrem em todas as Embryophytas: no Grupo I os PoGOs A e B e no Grupo II o PoGO F (Tabela 5; Figura 30). Este resultado sugere que os Grupos I e II surgiram de forma paralela no evento de especiação que levou à conquista do ambiente terrestre pelas plantas. Desta forma consideramos estes grupos como sendo distintos.

Nossos resultados para o Grupo III, que possui o maior nível de conservação das seqüências, foram concordantes com os trabalhos prévios (Rose *et al.*, 2002; Yokoyama *et al.*, 2004; e Baumman *et al.*, 2007). Dentro deste grupo está o PoGO L, onde estão as proteínas codificadas por At-XTH31 e At-XTH32 que possuem atividade somente hidrolítica, não possuindo atividade de transglicosilação (Baumman *et al.*, 2007). Baumman *et al.* (2007) sugerem que esta atividade seria um ganho funcional que teria surgido em angiospermas, no entanto nossos resultados mostram claramente seqüências de “gimnospermas” (Coniferophyta, Gnetophyta e Cycadophyta) dentro deste PoGO com valor de *bootstrap* de 99%. A atividade hidrolítica de xiloglucano está associada a



crescimento rápido (meristemas), germinação de sementes e degradação de reserva em sementes (Baumman *et al.* 2007), o que é compatível com um ganho funcional ocorrido no ancestral comum das plantas com semente (Spermatophyta),

O Grupo IV, que possui apenas o PoGO K, havia sido relatado como sendo o grupo ancestral das XTHs por Baumman *et al.* (2007), no entanto este trabalho somente considerou seqüências de angiospermas e não classificou estas seqüências como constituindo um grupo de homólogos distinto do Grupo I, onde este havia sido inicialmente colocado (Rose *et al.*, 2002). Nossos dados mostraram que o PoGO K possui seqüências de *Physcomitrella patens*, o que sugere que no ancestral comum das Embryophytas o grupo IV já era distinto dos Grupos I e II. Este fato nos levou a rejeitar a hipótese deste grupo de genes constituírem o grupo ancestral das demais XTHs.

O Grupo V aparece como sendo um grupo externo aos Grupos I, II, III e IV. Este grupo, constituído por dois ESTs de *Marchantia polymorpha* e um EST de *Closterium peracerosum*, existiria apenas em Marchantiophytas (hepáticas) e em algas Charophytas e teria sido perdido, posteriormente à divergência das Marchantiophytas, nas bryophytas e Tracheophytas. O fato de não haver nenhum genoma completamente seqüenciado dentro destes filos (Marchantiophyta e Charophyta) não impede de confirmar a existência deste grupo bem como determinar quanto genes ele possui. No entanto dados moleculares sugerem que as algas Charophytas seriam os ancestrais das plantas terrestres (Karol *et al.*, 2001) e que as hepáticas seriam o grupo basal das plantas terrestres (Boisselire-Dubayle, *et al.*, 2002), o que torna possível haver um grupo de ortólogos exclusivo destes dois táxons já que um evento de perda destes genes, posterior à divergência das Marchantiophytas, poderia explicar sua ausência nos outros grupos de plantas.

Os grupos I, II, III, IV e V formam um grande grupo monofilético dentro da linhagem das plantas com suporte de 99% *bootstrap* (Figura 29) e englobam todas as seqüências que já foram previamente relatadas como sendo XTHs. Sugerimos que o grupo V seja o grupo mais basal dentro desta linhagem de genes tanto pela topologia obtida como pelo fato deste grupo conter uma seqüência de uma alga Charophyta, no entanto não é possível descartar a hipótese da existência de seqüências de Charophytas que se encaixariam dentro de outros grupos, isto só poderá ser confirmado quando houver o seqüenciamento do genoma completo de algum representante deste filo de algas verdes.

O grupo VI nunca havia sido reportado anteriormente, possivelmente pelo fato da planta modelo *Arabidopsis thaliana* ter perdido o(s) gene(s) correspondente a este grupo e os primeiros trabalhos estarem baseados na resolução filogenética dos parálogos desta espécie. Este grupo possui apenas o PoGO R e pelo fato de possuir genes oriundos de todos os outros genomas completos de Embryophytas utilizados, consideramos que *Arabidopsis* perdeu este gene após a separação de sua linhagem evolutiva das demais linhagens de eudicotiledôneas consideradas em nossa análise. O PoGO R possui a seqüência parcial **Chara2** (retirada de Van Sandt *et al.* (2007), não foi depositada em bancos de dados) de *Chara vulgaris*, uma alga Charophyta, que foi obtida por Van Sandt *et al.* (2007) num trabalho onde foi demonstrado experimentalmente que esta alga possuía atividade de transglicosilação de xiloglucano nas regiões de crescimento, o que é equivalente ao que já foi demonstrado em angiospermas como sendo o produto da atividade das XTHs. Esta é a única evidência funcional conhecida dentro do PoGO R, mas sugere que as enzimas deste grupo possuem atividade equivalente às XTHs previamente reportadas.

O fato de *Arabidopsis* ter perdido seu representante dentro do PoGO R sugere uma sobreposição funcional deste grupo com as XTHs que formam o grupo monofilético I/II/III/IV/V exclusivo de plantas, o que é mais uma evidência que sustenta o fato de este grupo representar uma função catalítica equivalente à XTH. Nossas análises sugerem que o grupo VI não faz parte da linhagem evolutiva das XTHs previamente reportadas. São necessários estudos funcionais ao longo das proteínas presentes neste PoGO para confirmar a atividade de transglicosilação de xiloglucano, no entanto a presença do sítio catalítico das XTHs (DEIDFEFLG) nas proteínas deste grupo é uma forte evidência para sustentar tal atividade (Figura 31).

Buscamos utilizando Blastp e tBlastn por seqüências similares às XTHs nos genomas de outros eucariotos (animais e fungos) visando elucidar as origens evolutivas deste grupo de genes que aparentemente é exclusivo de plantas. Em animais não fomos capazes de encontrar seqüências que apresentassem qualquer similaridade com XTHs, no entanto observamos que existiam seqüências similares em genomas de fungos. Obtivemos, dos genomas completos de *Saccharomyces cerevisiae* e *Aspergillus nidulans* as seqüências que aparecem como grupo externo (*Out group*) nas Figuras 29 e 30. No total, os genomas de *S. cerevisiae* e *A. nidulans*, codificam, respectivamente, para três (*Utr2p* – NP\_010874, *Crr1p* – NP\_013314 e *Crh1p* – NP\_011705; GeneBank) e cinco (XP\_662119, XP\_664552, XP\_660657, XP\_658537, XP\_661518; GeneBank) enzimas hidrolíticas que apresentam similaridade com XTHs.

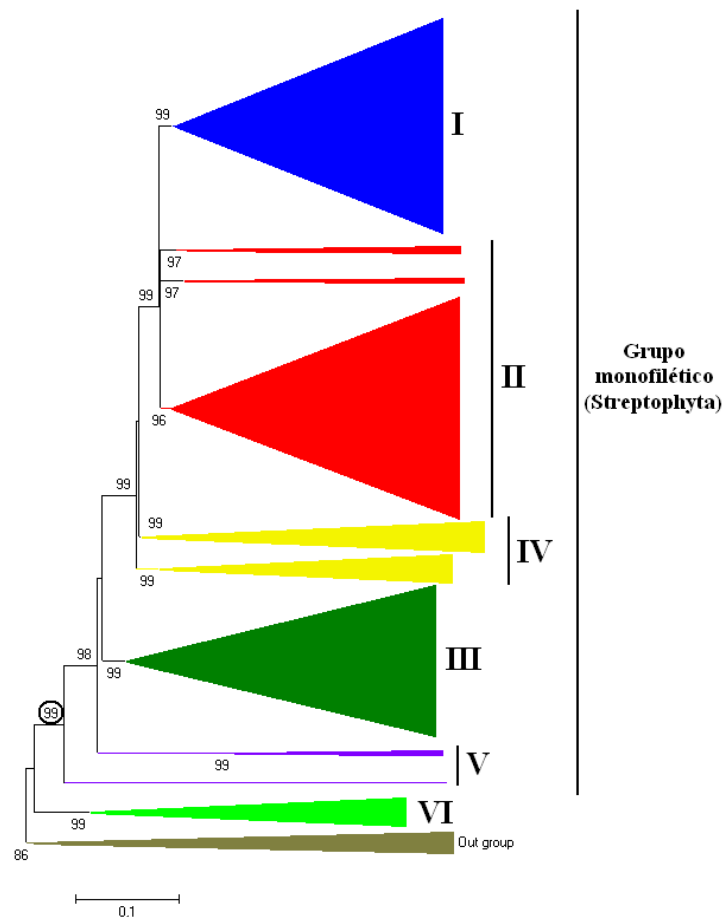
Utilizamos as seqüências de fungos juntamente com as seqüências do PoGO R, e as seqüências das XTHs At-XTH3 (Grupo IV) e At-XTH9 (Grupo I) como grupos externos numa análise filogenética por Máxima Parcimônia (MP), onde foram incluídas 105 posições alinhadas (Figura 32), com 1000 réplicas de *bootstraps*. A análise revelou

um valor de 95% de *bootstrap* unindo o PoGO R com o grupo de seqüências de fungos (Figura 33). Este resultado sugere que o Grupo VI e as XTHs (Grupo monofilético I/II/III/IV/V) exclusivas de plantas possuem uma origem comum e que a seqüência ancestral deste grupo existia no ancestral comum a todos os eucariotos.

O PoGO R possui uma seqüência de alga Charophyta (*Chara vulgaris* - **Chara 2**) assim como o PoGO Q (*Closterium peracerosum* – TA492\_34146; TIGR Plant Transcripts Assemblies) que seria o grupo basal das XTHs específicas de plantas. Isso sugere que a duplicação que deu origem às XTHs específicas de plantas ocorreu muito cedo no reino Viridiplantae, antes mesmo da conquista do meio terrestre, possivelmente no ancestral comum às Streptophyta (Embryophyta + Charophyta). Nos genomas completos de algas-verdes mais simples (*Volvox carteri*, *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*, *Ostreococcus tauri* e *Micromonas pusilla*) não fomos capazes de encontrar genes relacionados às XTHs. Esse fato pode ser interpretado como uma perda destes genes especificamente nestas linhagens basais das plantas-verdes nas quais provavelmente não era essenciais.

Os genes *UTR2* (ou *CRH2*) e *CRH1* de levedura foram caracterizados como beta-glucanases (assim como são as XTHs), possivelmente transglicosilases, que teria um papel na manutenção da organização da parede celular (Cabib *et al.*, 2007). Os mutantes individuais destes genes apresentam uma diminuição das ligações cruzadas entre quitina e beta-(1-6)-glucano e fenótipo alterado de crescimento, e no mutante duplo total ausência destas ligações cruzadas. *Crh1p* e *Crh2p* foram as primeiras proteínas diretamente associadas à formação de ligações cruzadas entre componentes da parede celular em fungos. O gene *CRR1* encoda uma Transglicosilase é expresso somente durante a esporulação da levedura e é necessário para a montagem normal da

parece celular do esporo, tendo um possível papel na formação de conexões entre as camadas de glucano e quitina (Gómez-Esquer *et al.*, 2004). Estes resultados funcionais reafirmam a possível conexão evolutiva entre o processo de transglicosilação em plantas e fungos como sendo um mecanismo basal em eucariotos para a organização da parede e crescimento celular. O genoma da levedura possui apenas três genes evolutivamente relacionados (e cinco em *Aspergillus nidulans*) que participam destes processos, enquanto *Arabidopsis*, por exemplo, possui 33 XTHs. Esta amplificação substancial em plantas pode ser uma das possíveis fontes da diversidade morfológica em células vegetais, em comparação com a organização simples, em hifas, dos fungos.



**Figura 29.** Grupos de XTHs homólogas. *Bootstrap* maiores que 50% estão mostrados nos nós. O valor de *bootstrap* circulado suporta o grupo monofilético em Streptophyta. O grupo externo é constituído pelas seqüências de fungos descritas nos resultados. Topologia obtida por NJ com distâncias obtidas por *p-distance* e 1000 réplicas de *bootstrap*. O tamanho dos triângulos corresponde à quantidade relativa de genes.

**Tabela 5.** Classificação das XTHs de em grupos de genes homólogos e PoGOs.

**Grupo I**

PoGO	XTH Nº	Código do gene	Sinônimo	Classe	Divisão ou Filo
	At-XTH4	At2g06850	EXGT-A1 , EXT		<b>Angiosperma (Magnoliophyta)</b>
	At-XTH5	At5g13870	EXGT-A4, XTR12		
	Gm-XTH2	Gm0111x00120		Eudicotiledôneas	
	Gm-XTH3	Gm0155x00001			
	Gm-XTH4	Gm0155x00002			
	Gm-XTH13	Gm0109x00119			
	Vv-XTH3	TCVV005241			
	Pt-XTH2	712150			
	Pt-XTH3	572825			
	Pt-XTH4	708892			
	Os-XTH2	LOC_Os11g33270	OsXTR4	Monocotiledôneas	
	Sb-XTH2	Sb05g020400			
	<i>Saruma henryi</i> -XTH1	DT583992 **		Magnoliids (Ord. Piperales)	
	<i>Amborella trichopoda</i> -XTH1	CK75516 **		Angio basal (Ord. Amborellales)	
<b>A</b>	<i>Picea sitchensis</i> -XTH1	ABK21095 ***		Coniferopsida	<b>Coniferophyta</b>
	<i>Picea sitchensis</i> -XTH2	ABK25883 ***			
	<i>Picea sitchensis</i> -XTH3	ABK21757 ***			
	<i>Pinus taeda</i> -XTH1	TA9639_3352 **			
	<i>Cryptomeria japonica</i> -XTH1	TA3285_3369 **			
	<i>Welwitschia mirabilis</i> -XTH1	DT590665 **		Gnetopsida	<b>Gnetophyta</b>
	<i>Welwitschia mirabilis</i> -XTH2	DT604218 **			
	Sm-XTH3	422922		Isoetopsida	<b>Lycopodiophyta</b>
	Sm-XTH5	14595			
	Sm-XTH15	426178			
	Pp-XTH6	117299		Bryopsida	<b>Bryophyta</b>
	Pp-XTH18	148165			
	Pp-XTH20	112374			
	Pp-XTH21	70081			
	Pp-XTH23	225086			
	Pp-XTH26	225082			
	<i>Marchantia polymorpha</i> -XTH1	TA1645_3197 **		Marchantiopsida	<b>Marchantiophyta</b>
	<i>Marchantia polymorpha</i> -XTH2	TA828_3197 **			
	<i>Marchantia polymorpha</i> -XTH3	BJ856950 **			

<b>B</b>	At-XTH10	At2g14620	XTR14	Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Embryophyta</b>
	Gm-XTH16	Gm0052x00399				
	Pt-XTH17	557559				
	Vv-XTH19	CAO49843				
	Sm-XTH10	267868		Isoetopsida	<b>Lycopodiophyta</b>	
	Pp-XTH12	108715		Bryopsida	<b>Bryophyta</b>	
	Pp-XTH19	112531				
	Pp-XTH22	191788				
	Pp-XTH25	140590				
	Pp-XTH30	111621				

<b>C</b>	At-XTH9	At4g03210	EXGT-A6, XTR16 BAB86890 ***	Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Tracheophyta</b>
	Gm-XTH17	Gm0364x00007				
	Gm-XTH26	Gm0082x00141				
	Pt-XTH6	571855				
	Pt-XTH15	250348				
	Vv-XTH9	VV78X212580				
	Vv-XTH12	TCVV009206				
	Os-XTH3	LOC_Os04g53950		Monocotiledôneas		
	Sb-XTH6	Sb10g028580				
	<i>Saruma henryi</i> -XTH2	DT598485 **		Magnoliids (Ord. Piperales)		
	<i>Picea sitchensis</i> -XTH4	ABK22283 ***		Coniferopsida	<b>Coniferophyta</b>	
	<i>Picea sitchensis</i> -XTH5	ABK22211 ***				
	<i>Picea glauca</i> -XTH1	TA14242_3330 **				
	<i>Pinus taeda</i> -XTH2	TA17577_3352 **				
	<i>Welwitschia mirabilis</i> -XTH3	TA906_3377 **		Gnetopsida	<b>Gnetophyta</b>	
	Sm-XTH1	140273		Isoetopsida	<b>Lycopodiophyta</b>	
	Sm-XTH9	110418				
	Sm-XTH11	118035				
Sm-XTH12	97913					
Sm-XTH13	98543					
<i>Selaginella kraussiana</i> (SkXTH1)	AAT84462 ***					
<i>Selaginella lepidophylla</i> -XTH1	BM402496 **					

<b>D</b>	At-XTH6	At5g65730	XTR10	Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Spermatophyta</b>
	At-XTH7	At4g37800	XTR15			
	Pt-XTH5	718004				
	Gm-XTH5	Gm0013x00525				
	Vv-XTH11	TCVV007457				
	<i>Persea americana</i> -XTH1	CK759615 **		(Ord. Laurales)	<b>Coniferophyta</b>	
	<i>Liriodendron tulipifera</i> -XTH1	TA1454_3415 **		Magnoliids (Ord. Piperales)		
	<i>Saruma henryi</i> -XTH3	DT576904 **		(Ord. Piperales)		
	<i>Pinus taeda</i> -XTH3	TA11677_3352 **		Coniferopsida		
	<i>Cryptomeria japonica</i> -XTH2	TA1815_3369 **				
<i>Welwitschia mirabilis</i> -XTH4	CK761545 **		Gnetopsida	<b>Gnetophyta</b>		

E	At-XTH8	At1g11545			Eudicotiledóneas	Angiosperma (Magnoliophyta)	Tracheophyta
	Gm-XTH11	Gm0052x00423					
	Pt-XTH25	196727					
	Pt-XTH28	233810					
	Vv-XTH15	VV78X251477					
	Os-XTH1	LOC_Os07g34580	OsXTR1	Monocotiledóneas			
	Sb-XTH1	Sb02g035160					
	<i>Picea sitchensis</i> -XTH6	TA25505_3332 **			Coniferopsida	Coniferophyta	
	<i>Picea sitchensis</i> -XTH7	ABK26934 ***					
	<i>Picea sitchensis</i> -XTH8	ABK23651 ***					
	<i>Pinus taeda</i> -XTH4	TA9314_3352 **					
	<i>Cryptomeria japonica</i> -XTH3	TA2201_3369 **					
	<i>Welwitschia mirabilis</i> -XTH5	TA793_3377 **		Gnetopsida	Gnetophyta		
<i>Cycas rumphii</i> -XTH1	CB093138 **		Cycadopsida	Cycadophyta			
Sm-XTH2	75136		Isoetopsida	Lycopodiophyta			

## Grupo II

F	At-XTH26	At4g28850	XTR18		Eudicotiledóneas	Angiosperma (Magnoliophyta)	Embryophyta
	Gm-XTH12	Gm0092x00041					
	Gm-XTH14	Gm0030x00696					
	Gm-XTH29	Gm0013x00085					
	Pt-XTH8	779340					
	Vv-XTH4	VV78X254686					
	Os-XTH13	LOC_Os02g17880			Monocotiledóneas		
	Os-XTH14	LOC_Os02g17900					
	Os-XTH15	LOC_Os06g22919					
	Sb-XTH9	Sb10g019390					
	Sb-XTH16	Sb04g010980					
	Sb-XTH18	Sb04g010760					
	Sm-XTH8	403283		Isoetopsida	Lycopodiophyta		
	Pp-XTH2	131542			Bryopsida	Bryophyta	
	Pp-XTH16	142529					
	Pp-XTH17	102528					
	<i>Marchantia polymorpha</i> -XTH4	TA1466_3197 **			Marchantiopsida	Marchantiophyta	
<i>Marchantia polymorpha</i> -XTH5	TA427_3197 **						
<i>Marchantia polymorpha</i> -XTH6	BJ858913 **						
<i>Marchantia polymorpha</i> -XTH7	BJ848098 **						

G	Os-XTH16	LOC_Os04g51520			Monocotiledóneas	Angiosperma (Magnoliophyta)
	Os-XTH17	LOC_Os08g13920				
	Os-XTH18	LOC_Os06g48170				
	Sb-XTH7	Sb06g029660				
	Sb-XTH17	Sb07g009410				
	Sb-XTH20	Sb06g015940				
	Sb-XTH26	Sb06g015880				
	Sb-XTH29	Sb06g015930				



H

Os-XTH4	LOC_Os08g14200					
Os-XTH5	LOC_Os08g14210					
Os-XTH6	LOC_Os04g51450					
Os-XTH7	LOC_Os04g51510					
Os-XTH8	LOC_Os08g13980					
Os-XTH9	LOC_Os04g51460			Monocotiledóneas		
Sb-XTH8	Sb07g007010					
Sb-XTH10	Sb07g007460					
Sb-XTH11	Sb06g027670					
Sb-XTH12	Sb07g007020					
Sb-XTH14	Sb06g027660					
Sb-XTH15	Sb06g027680					
Sb-XTH22	Sb07g007450					
At-XTH12	At5g57530					
At-XTH13	At5g57540					
At-XTH14	At4g25820					
Gm-XTH22	Gm0015x10189	XTR9		Eudicotiledóneas		
Gm-XTH27	Gm0004x10090					
Pt-XTH23	786577					
<i>Amborella trichopoda</i> -XTH2	CV011512 **			Angio basal (Ord. Amborellales)		
<i>Pinus taeda</i> -XTH5	TA10225_3352 **					
<i>Pinus taeda</i> -XTH6	TA19224_3352 **					
<i>Picea engelmannii</i> x <i>sitchensis</i> -XTH1	DR464422 **			Coniferopsida		
<i>Picea glauca</i> -XTH2	TA26213_3330 **					
<i>Cryptomeria japonica</i> -XTH4	AB197670 **					
<i>Cycas rumphii</i> -XTH2	DR062353 **			Cycadopsida		

I

At-XTH15	At4g14130	XTR7				
At-XTH16	At3g23730					
Gm-XTH21	Gm0139x00211					
Gm-XTH24	Gm0022x00458					
Gm-XTH25	Gm0003x00166					
Gm-XTH33	Gm0121x00022					
Pt-XTH7	205198					
Pt-XTH9	572886			Eudicotiledóneas		
Pt-XTH18	754387					
Pt-XTH19	234656					
Pt-XTH20	411405					
Pt-XTH24	711461					
Pt-XTH27	234660					
Vv-XTH8	VV78X174940					
Vv-XTH10	VV78X007417					
Vv-XTH13	A06004430					
Vv-XTH14	TCVV000308					
Os-XTH11	LOC_Os06g48160					
Os-XTH12	LOC_Os06g48180			Monocotiledóneas		
Sb-XTH3	Sb10g028550					
Sb-XTH5	Sb10g028570					

**J**

At-XTH17	At1g65310	XTR1	
At-XTH18	At4g30280		
At-XTH19	At4g30290		
At-XTH20	At5g48070		
At-XTH21	At2g18800	XTR17	
At-XTH22	At5g57560	TCH4	
At-XTH23	At4g25810	XTR6	
At-XTH24	At4g30270	Meri5	
At-XTH25	At5g57550	EXGT-A5, XTR3	
Gm-XTH6	Gm0084x00056		Eudicotiledóneas
Gm-XTH8	Gm0025x00555		
Gm-XTH9	Gm0084x00133		
Gm-XTH18	Gm0084x00058		
Gm-XTH19	Gm0025x00559		
Gm-XTH23	Gm0086x00025		
Pt-XTH10	262309		
Pt-XTH12	762218		
Pt-XTH14	570685		
Pt-XTH16	262298		
Pt-XTH22	262296		
Pt-XTH26	811167		
Vv-XTH7	VV78X108629		
Os-XTH10	LOC_Os06g48200	OsXTR2	
Sb-XTH4	Sb10g028560		Monocotiledóneas

**Angiosperma  
(Magnoliophyta)**

**Grupo III**

**L**

At-XTH31	At3g44990	ATXG, AtXTR8	
At-XTH32	At2g36870		
Gm-XTH30	Gm0118x00030		
Gm-XTH31	Gm0009x00173		
Gm-XTH32	Gm0201x10021		Eudicotiledóneas
Gm-XTH34	Gm0041x00050		
Pt-XTH32	768438		
Pt-XTH33	825278		
Pt-XTH34	282192		
Vv-XTH17	TCVV000509		
Vv-XTH18	A06008325		
Os-XTH19	LOC_Os03g01800		
Os-XTH20	LOC_Os10g39840		
Os-XTH21	LOC_Os07g29750		
Os-XTH22	LOC_Os02g57770		Monocotiledóneas
Sb-XTH21	Sb01g050070		
Sb-XTH23	Sb01g030140		
Sb-XTH27	Sb02g033240		
<i>Pinus taeda</i> -XTH7	TA7617_3352 **		
<i>Pinus taeda</i> -XTH8	TA12528_3352 **		Coniferopsida
<i>Picea sitchensis</i> -XTH9	ABK23544 ***		
<i>Welwitschia mirabilis</i> -XTH6	TA1688_337 **		Gnetopsida
<i>Cycas rumphii</i> -XTH3	CB092744 **		Cycadopsida

**Angiosperma  
(Magnoliophyta)**

**Spermatophyta**

**Coniferophyta**

**Gnetophyta**

**Cycadophyta**

<b>M</b>	At-XTH33	At1g10550	Eudicotiledóneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Embryophyta</b>
	Gm-XTH28	Gm0004x00592			
	Pt-XTH29	731607			
	Vv-XTH16	VV78X089690			
	<i>Welwitschia mirabilis</i> -XTH7	CK74448 **	Gnetopsida	<b>Gnetophyta</b>	
	Sm-XTH4	232041	Isoetopsida	<b>Lycopodiophyta</b>	
	Sm-XTH6	56616			
	Sm-XTH7	115266			
	Pp-XTH1	113013	Bryopsida	<b>Bryophyta</b>	
<i>Marchantia polymorpha</i> -XTH8	BJ872294 **	Marchantiopsida	<b>Marchantiophyta</b>		
<i>Marchantia polymorpha</i> -XTH9	BJ846938 **				

<b>Proto NOP</b>	<i>Picea sitchensis</i> -XTH10	TA21165_3332 **	Coniferopsida	<b>Coniferophyta</b>
	<i>Pinus pinaster</i> -XTH	TA3431_71647 **		

<b>N</b>	At-XTH27	At2g01850	EXGT-A3 EXGT-A2, XTR2	<b>Angiosperma (Magnoliophyta)</b>	
	At-XTH28	At1g14720			
	Gm-XTH10	Gm0022x00585			
	Gm-XTH15	Gm0049x00295			Eudicotiledóneas
	Pt-XTH11	658681			
	Pt-XTH13	720800			
	Vv-XTH6	TCVV001724	Monocotiledóneas		
	Os-XTH27	LOC_Os10g02770			
	Os-XTH28	LOC_Os03g13570			
Sb-XTH24	Sb01g041510				

<b>O</b>	Os-XTH23	LOC_Os02g46910	OsXTR3	Monocotiledóneas
	Os-XTH24	LOC_Os03g63760		
	Os-XTH25	LOC_Os10g42670		
	Sb-XTH30	Sb04g031050		
	Sb-XTH31	Sb01g027960		
	Sb-XTH32	Sb01g000710		

<b>P</b>	At-XTH29	At4g18990	XTR13 XTR4	<b>Angiosperma (Magnoliophyta)</b>
	At-XTH30	At1g32170		
	Gm-XTH20	Gm0170x00005	Eudicotiledóneas	
	Pt-XTH21	423757		
	Pt-XTH31	646298		
	Vv-XTH5	VV78X082245	Monocotiledóneas	
	Os-XTH26	LOC_Os02g03550		
	Sb-XTH28	Sb04g002290		
	<i>Amborella trichopoda</i> -XTH3	CD484138 **	Angio basal (Ord. Amborellales)	

**Grupo IV**

<b>K</b>	At-XTH1	At4g13080	Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Embryophyta</b>
	At-XTH2	At4g13090			
	At-XTH3	At3g25050			
	At-XTH11	At3g48580			
	Gm-XTH1	Gm0119x00160			
	Gm-XTH7	Gm0039x00101			
	Vv-XTH1	VV78X275540			
	Vv-XTH2	VV78X023456			
	Pt-XTH1	408935			
	Os-XTH30	LOC_Os09g23220	Monocotiledôneas		
	Sb-XTH13	Sb10g008550			
	Sb-XTH19	Sb10g008540			
	Pp-XTH3	150064	Bryopsida	<b>Bryophyta</b>	
	Pp-XTH4	130295			
	Pp-XTH5	137834			
	Pp-XTH7	1560065			
	Pp-XTH8	4860001			
	Pp-XTH9	2130114			
	Pp-XTH10	4100003			
Pp-XTH11	135678				
Pp-XTH15	168968				
Pp-XTH27	124249				
Pp-XTH28	114674				
Pp-XTH29	34357				

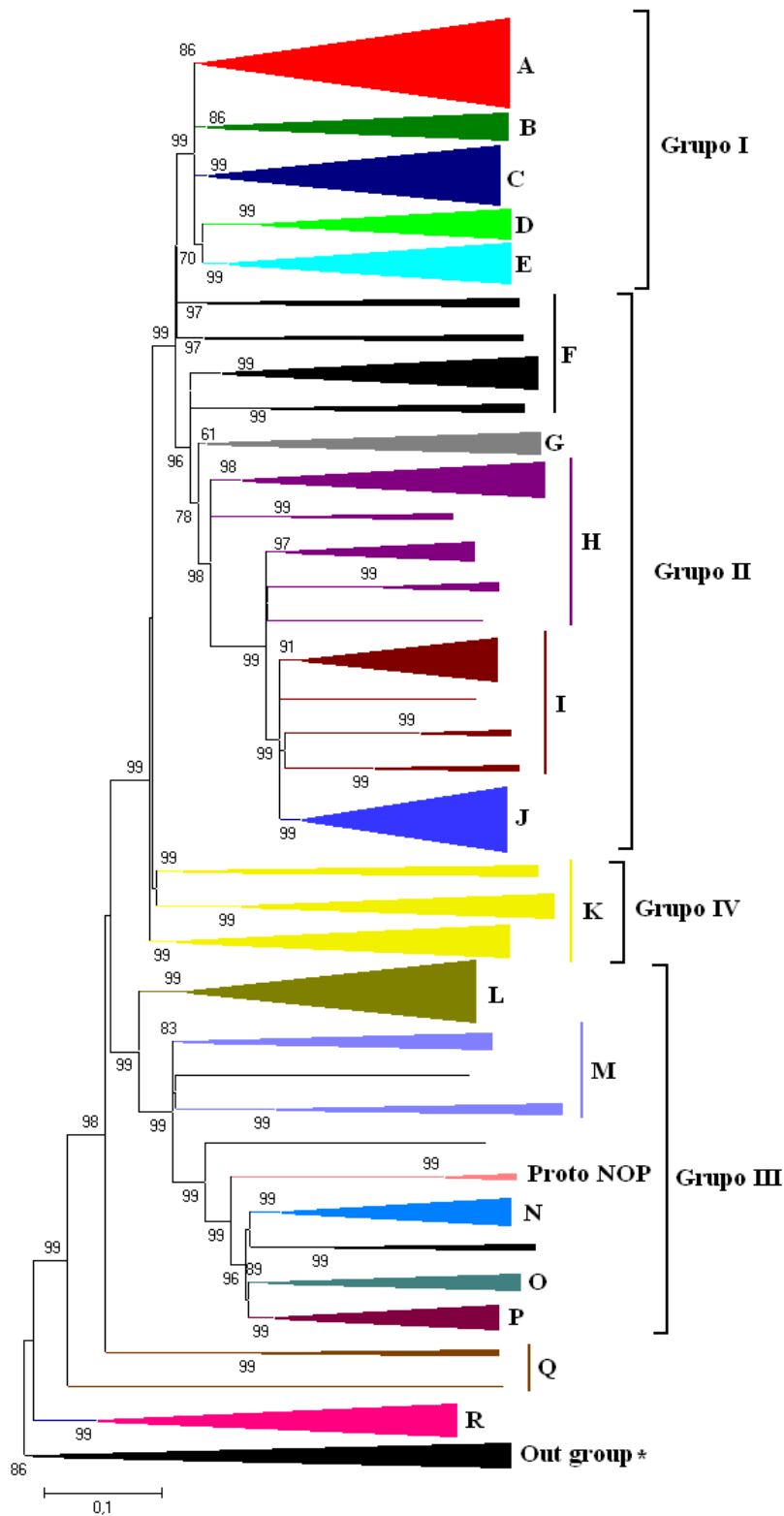
**Grupo V**

<b>Q Proto XTH</b>	<i>Marchantia polymorpha</i> -XTH10	TA608_3197 **	Marchantiopsida	<b>Marchantiophyta</b>
	<i>Marchantia polymorpha</i> -XTH11	BJ853101 **		
	<i>Closterium peracerosum</i> -XTH	TA492_34146 **		

**Grupo VI**

<b>R</b>	Gm-XTH35	Gm0108x00232	Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Streptophyta</b>
	Pt-XTH35	552110			
	Vv-XTH20	TCVV007088			
	Os-XTH31	LOC_Os03g011730	Monocotiledôneas		
	Sb-XTH33	Sb01g049420			
	Sm-XTH14	411404	Isoetopsida	<b>Lycopodiophyta</b>	
	Sm-XTH16	75086			
	Pp-XTH13	152060	Bryopsida	<b>Bryophyta</b>	
	Pp-XTH14	34640			
	Pp-XTH24	207213			
	<i>Marchantia polymorpha</i> -XTH12	TA816_3197 **	Marchantiopsida	<b>Marchantiophyta</b>	
<i>Chara vulgaris</i> -Chara2	Van Sandt <i>et al.</i> , 2007	Charophyceae	<b>Charophyta</b>		

\*\* N° de acesso: TIGR *Plant Transcript Assemblies*; \*\*\* N° de acesso: GeneBank.

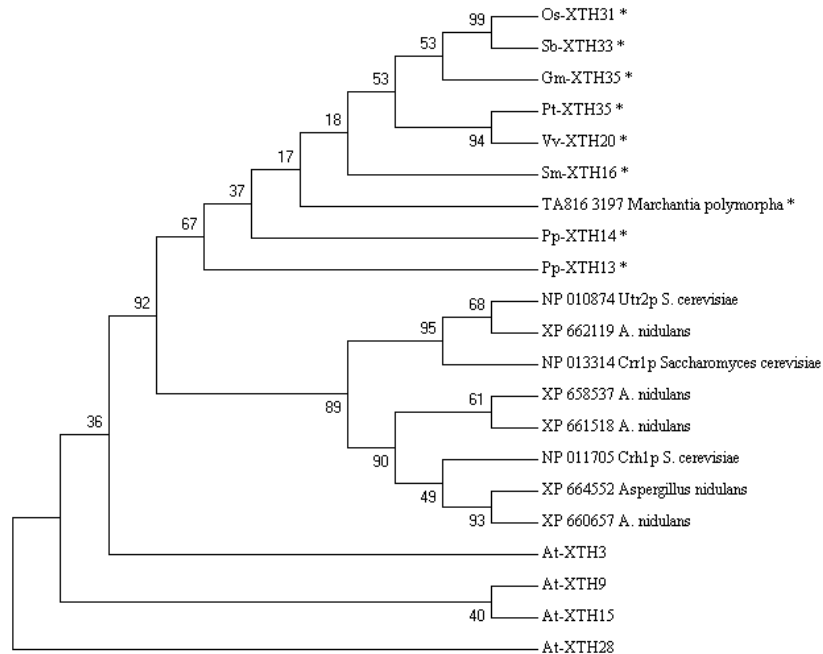


**Figura 30.** Árvore filogenética compactada mostrando os PoGOs das XTHs. Os valores de *bootstrap*, maiores que 50%, que suportam os PoGOs estão mostrados nos nós. A topologia foi obtida por NJ com distâncias calculadas por *p-distance*, foram realizadas 1000 réplicas de *bootstrap*. O tamanho dos triângulos corresponde à quantidade relativa de genes. O grupo externo (\*) é composto pelas seqüências de fungo descritas nos resultados. Está árvore, em detalhes, pode ser vista na Figura 1 do ANEXO III.



Pt-XTH35 *	L	V	P	R	G	G	A	R	R	R	F	L	V	G	E	L	I	C	G	L	F	I	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	A	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V											
Sm-XTH16 *	V	F	D	R	G	G	A	R	R	R	F	F	G	A	I	I	G	F	L	F	Y	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V											
Os-XTH31 *	V	V	D	R	G	G	A	R	R	R	F	F	G	A	I	I	G	F	L	F	Y	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V											
Sb-XTH33 *	V	V	D	R	G	G	A	R	R	R	F	F	G	A	I	I	G	F	L	F	Y	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V											
Gm-XTH35 *	L	V	P	R	G	G	A	R	R	R	F	L	V	G	E	L	I	C	G	L	F	I	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	A	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V										
Vv-XTH20 *	L	V	P	R	G	G	A	R	R	R	F	L	V	G	E	L	I	C	G	L	F	I	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V										
Pp-XTH13 *	L	V	P	R	G	G	A	R	R	R	F	L	V	G	E	L	I	C	G	L	F	I	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V										
Pp-XTH14 *	D	I	F	D	R	G	G	R	L	L	V	F	G	A	I	L	I	G	L	F	Y	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V											
TA816 3197 <i>M. polymorpha</i> *	L	V	P	R	G	G	A	R	R	R	F	L	V	G	E	L	I	C	G	L	F	I	L	L	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V										
NP_013314 <i>Crr1p S. cerevisiae</i>	V	F	I	R	A	L	L	R	R	I	I	V	A	L	A	M	R	R	G	L	I	L	F	L	G	A	G	E	I	F	E	L	G	G	L	M	A	S	V	R	G	L	D	Y	R	M	R	V	G	A	D	A	Y	H	E	I	D	R	I	V	D	G	I	A									
XP_664552 <i>Aspergillus nidulans</i>	F	I	I	R	A	L	L	R	R	F	I	F	F	G	V	V	A	V	M	A	F	A	G	I	L	V	V	L	D	L	E	I	F	E	V	G	V	V	F	F	G	G	L	V	V	F	F	G	G	S	V	D	G	A	V	A	E	I	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V
XP_658537 <i>A. nidulans</i>	F	I	I	R	A	L	L	R	R	F	I	F	F	G	V	V	A	V	M	A	F	A	G	I	L	V	V	L	D	L	E	I	F	E	V	G	V	V	F	F	G	G	L	V	V	F	F	G	G	S	V	D	G	A	V	A	E	I	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V
NP_011705 <i>Crh1p S. cerevisiae</i>	M	L	A	S	S	L	S	F	Y	I	H	Y	G	L	E	V	L	A	A	G	G	I	V	F	L	S	D	L	E	I	I	E	V	G	G	L	I	F	S	F	F	S	A	G	L	V	D	S	G	E	V	D	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V					
NP_010874 <i>Utr2p S. cerevisiae</i>	F	V	Y	A	L	L	Y	G	D	V	L	L	A	M	R	R	G	L	I	L	F	L	G	A	G	E	I	F	E	L	G	G	L	M	A	S	V	R	G	L	D	Y	R	M	R	V	G	A	D	A	Y	H	E	I	D	R	I	V	D	G	I	A											
XP_662119 <i>A. nidulans</i>	V	F	E	L	D	A	D	V	G	K	L	V	E	L	M	R	R	G	L	I	A	H	Y	I	G	K	I	D	E	I	F	E	V	G	D	L	E	V	V	F	F	G	L	V	V	G	G	V	G	G	A	V	H	E	I	D	R	I	V	D	G	V	V										
XP_661518 <i>A. nidulans</i>	F	V	A	G	A	L	I	D	I	M	F	G	V	F	V	I	A	A	G	V	G	I	V	A	V	L	D	L	E	I	F	E	V	G	V	V	F	F	G	G	A	V	A	A	A	G	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V												
XP_660657 <i>A. nidulans</i>	F	I	A	S	S	L	S	F	Y	I	H	Y	G	L	E	V	L	A	A	G	G	I	V	F	L	S	D	L	E	I	I	E	V	G	G	L	I	F	S	F	F	S	A	G	L	V	D	S	G	E	V	D	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V					
At-XTH3	L	L	M	R	G	G	F	R	A	V	G	G	N	I	V	I	G	G	G	I	V	A	F	Y	L	H	D	E	I	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V								
At-XTH9	L	L	D	V	G	A	G	F	R	S	R	L	F	G	I	I	L	V	R	G	A	G	V	A	F	Y	M	D	E	F	F	L	G	R	D	I	V	V	F	F	H	G	G	R	E	I	D	F	C	A	F	H	E	V	V	I	C	F	F	L	I	D	G	V	V								

**Figura 32.** Alinhamento utilizado para gerar a árvore da Figura 33. Um total de 105 posições conservadas entre seqüências do Grupo VI (seqüências marcadas com \*; Pt-XTH35, Sm-XTH16, Os-XTH31, Sb-XTH33, Gm-XTH35, Vv-XTH20, Pp-XTH13, Pp-XTH14 e *Marchantia polymorpha*-XTH12), as seqüências com similares a XTHs em fungos (*Crr1p*, *Crh1p* e *Utr2p* de *Saccharomyces cerevisiae* e NP\_013314, XP\_664552, XP\_658537, XP\_662119, XP\_661518, XP\_660657 [acessos do GeneBank] de *Aspergillus nidulans*) e duas XTHs pertencentes ao grupo monofilético de XTHs específicas de plantas (Grupo I/II/III/IV/V; At-XTH3 e At-XTH9) utilizadas como grupos externos. Das 105 posições, 98 são informativas e sete são idênticas entre todas as seqüências.



**Figura 33.** Árvore filogenética mostrando o agrupamento entre o Grupo VI (PoGO R) e as seqüências similares a XTHs em fungos. Análise filogenética por MP, com 1000 réplicas de *bootstrap*. As seqüências marcadas com \* fazem parte do Grupo VI. At-XTH3 (Grupo IV), At-XTH9 (Grupo I), At-XTH15 (Grupo II) e At-XTH28 (Grupo III) foram utilizadas na análise. Todos os valores de *bootstrap* estão mostrados.

### 3.2.3 – Análises filogenéticas das $\beta$ -Galactosidase e organização em PoGOs.

Análises filogenéticas por NJ permitiram a divisão da diversidade de  $\beta$ -Galactosidases, no reino Viridiplantae, em dois grupos de genes homólogos: o Grupo I que é distribuído por todos os eucariotos (plantas + animais + fungos) e contém uma única seqüência parcial de alga-verde (foi analisada separadamente por se tratar de uma seqüência parcial - Figura 34) e o Grupo II que é monofilético e específico de plantas terrestres. O Grupo I possui apenas o PoGO J, e o Grupo II possui nove PoGOs (Tabela 6; Figuras 35 e 36). No total obtivemos 133 seqüências não redundantes.

No Grupo II, que é exclusivo de plantas Embryophytas, provavelmente derivou do Grupo I no evento de especiação que levou à conquista do meio terrestre pelas plantas. O PoGO F representa a provável função ancestral do Grupo II, já que é o único dos PoGOs deste grupo que apresenta a conter genes de Marchantiophytas, Bryophytas e Lycophytas e por ter surgido por pressões seletivas por ganho de sustentação estrutural. É interessante notar, que ao contrário das XTHs, o surgimento das plantas vasculares (Tracheophytas) não apresentou amplificação dos grupos de ortólogos de  $\beta$ -Galactosidase. Este PoGO F provavelmente deu origem, no evento que levou ao surgimento das Spermatophytas, aos PoGOs C e E. As pressões seletivas que levaram ao surgimento dos PoGOs C e E provavelmente estão relacionadas ao ganho de complexidade estrutural, crescimento secundário e o surgimento da semente (Tabela 6; Figuras 35 e 36).

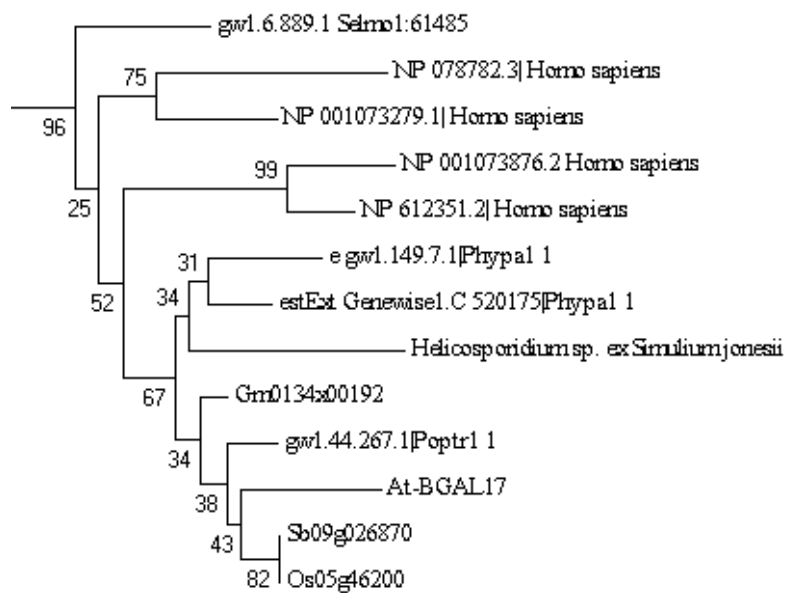
Outros cinco PoGOs (A, B, G, H e I) aparentemente surgiram no ancestral comum das Angiospermas, o que mostra uma marcante pressão seletiva para duplicações de genes codificantes para  $\beta$ -Galactosidases em Angiospermas. Este fato é



apoiado pelo ganho de importância relativa que o xiloglucano teve, como constituinte da hemicelulose, em Angiospermas, especialmente nas eudicotiledôneas. Entretanto não se pode descartar a possibilidade de alguns destes PoGOs terem surgido no ancestral comum às Spermatophytas, já que somente temos seqüências EST de “Gimnospermas” e seria necessário ter genomas completamente seqüenciado, representante deste táxon, para confirmar ou rejeitar esta hipótese.

O Grupo D aparece como sendo exclusivo de “Gimnospermas” (Coniferophyta, Gnetophyta e Ginkgophyta). Como as “Gimnospermas” formam um grupo parafilético (Raven *et al.*, 2001) é provável que as Angiospermas tenham perdido os genes deste grupo de ortólogos. No entanto outra possível explicação é que o gene ancestral dos grupos A e B (grupos irmãos) de Angiospermas poderia representar o ortólogo do grupo D. Este fato é apoiado pelo *bootstrap* de 98% que une os PoGOs A, B, C e D (Figuras 35 e 36).

Nenhuma enzima oriunda dos genomas completos de algas-verdes utilizados foi encontrada por similaridade com qualquer  $\beta$ -Galactosidase. Isso pode ser o resultado de uma perda destes genes especificamente nestas linhagens, já que uma Chlorophyta do gênero *Helicosporidium* apresenta pelo menos uma  $\beta$ -Galactosidase, no entanto esta não pertence ao grupo específico de plantas. É possível que a perda destes genes, e também das XTHs, em algas Chlorophytas se deve a uma organização diferente da parede celular. Não podemos descartar o fato de este resultado ser devido a artefatos nestes genomas completos ou mesmo na metodologia de busca que adotamos.



**Figura 34.** Análise filogenética do Grupo II das  $\beta$ -Galactosidases. A seqüência CX128430 (TIGR) da alga *Helicosporidium* faz parte deste grupo. Topologia inferida por NJ, distâncias por *p-distance* com 1000 replicas de *bootstrap* (valores mostrados nos nós).

**Tabela 6.** Classificação das  $\beta$ -Galactosidases em Grupos de homólogos e PoGOs.

PoGO	$\beta$ -Galactosidase Nº	Código do gene	Sinônimo	Classe	Divisão ou Filo	
<b>Grupo I</b>	At-BGal1				<b>Angiosperma (Magnoliophyta)</b>	
	At-BGal2					
	At-BGal4					
	At-BGal12					
	Vv-BGal1	VV78X080386				
	Vv-BGal2	VV78X168410				
	Pt-BGal1	44000028				
	Pt-BGal2	970004		Eudicotiledôneas		
	Pt-BGal3	VI000400				
	<b>A</b>	Gm-BGal1	Gm0088x00197			
		Gm-BGal2	Gm0215x00001			
		Gm-BGal3	Gm0025x00460			
		Gm-BGal4	Gm0086x00021			
		Os-BGal1	Os01g39830			
		Os-BGal2	Os02g12730			
		Os-BGal3	Os06g37560			Monocotiledôneas
		Sb-BGal1	Sb03g025990			
	Sb-BGal2	Sb04g007660				
	Sb-BGal3	Sb10g022620				
	<i>Liriodendron tulipifera</i> -BGal1	DT596898 **				Magnoliids (Ord. Magnoliales)
<b>B</b>	At-BGal3				<b>Angiosperma (Magnoliophyta)</b>	
	At-BGal5					
	Gm-BGal5	Gm0153x00072				
	Gm-BGal6	Gm0181x00052				
	Gm-BGal7	Gm0053x00360				
	Gm-BGal8	Gm0148x00130				
	Vv-BGal3	VV78X256866				
	Vv-BGal4	VV78X022431		Eudicotiledôneas		
	Pt-BGal4	LG V1511				
	Pt-BGal5	LG IV0471				
	Pt-BGal6	LG VII3760				
	Os-BGal4	Os03g06940				
	Sb-BGal4	Sb01g046160		Monocotiledôneas		
	<i>Liriodendron tulipifera</i> -BGal2	CK765873 **				Magnoliids (Ord. Magnoliales)
	<i>Saruma henryi</i> -BGal1	DT601692 **				Magnoliids (Ord. Piperales)

**C**

At-BGal10				<b>Angiosperma (Magnoliophyta)</b>	<b>Spermatophyta</b>
Gm-BGal9	Gm0065x00401				
Gm-BGal10	Gm0096x00096	Eudicotiledóneas			
Vv-BGal5	TCVV004574				
Pt-BGal7	VII000166				
Os-BGal5	Os01g65460			<b>Coniferophyta</b>	
Sb-BGal5	Sb03g041450	Monocotiledóneas			
<i>Pinus taeda</i> -BGal1	TA16991_3352 **	Coniferopsida			

**D**

<i>Picea glauca</i> -BGal1	DV996233 **			<b>Coniferophyta</b>	
<i>Picea glauca</i> -BGal2	TA20802_3330 **				
<i>Pinus taeda</i> -BGal2	TA21161_3352 **				
<i>Picea sitchensis</i> -BGal1	TA27850_3332 **				
<i>Pinus taeda</i> -BGal3	TA25885_3352 **	Coniferopsida			
<i>Chamaecyparis obtusa</i> -BGal1	TA852_13415 **				
<i>Pinus taeda</i> -BGal4	TA7564_3352 **				
<i>Picea glauca</i> -BGal3	TA28885_3330 **				
<i>Welwitschia mirabilis</i> -BGal1	TA527_3377 **	Gnetopsida		<b>Gnetophyta</b>	
<i>Ginkgo biloba</i> -BGal1	DR065132 **	Ginkgoopsida		<b>Ginkgophyta</b>	

**E**

At-BGal8				<b>Angiosperma (Magnoliophyta)</b>	<b>Spermatophyta</b>
Gm-BGal11	Gm0080x00053				
Gm-BGal12	Gm0099x00149				
Gm-BGal13	Gm0082x00342	Eudicotiledóneas			
Pt-BGal8	gw1.IX.4803.1				
Vv-BGal6	VV78X125775				
Pt-BGal9	eugene3.00110299			<b>Coniferophyta</b>	
Os-BGal6	Os03g15020				
Sb-BGal6	Sb01g040750	Monocotiledóneas			
<i>Picea sitchensis</i> -BGal2	ABK24373 ***				
<i>Pinus taeda</i> -BGal5	TA9043_3352 **	Coniferopsida			
<i>Picea glauca</i> -BGal4	TA17085_3330 **				

<b>F</b>	At-BGal9			Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Embryophyta</b>	
	Gm-BGal14	Gm0022x00436					
	Vv-BGal7	VV78X080252					
	Pt-BGal10	gw1.88.36.1					
	<i>Amborella trichopoda</i> -BGal1	TA1802_13333 **	Angio basal (Ord. Amborellales)				
	Os-BGal7	Os12g0429200		Monocotiledôneas			
	Sb-BGal7	Sb08g004410					
	Sm-BGal1	930039		Isoetopsida			<b>Lycopodiophyta</b>
	Sm-BGal2	3000021					
	Sm-BGal3	21000139					
	Pp-BGal1	gw1.24.11.1		Bryopsida	<b>Bryophyta</b>		
	Pp-BGal2	gw1.189.26.1					
	Pp-BGal3	1890048					
	Pp-BGal4	100096					
<i>Marchantia polymorpha</i> -BGal1	BJ858078 **	Marchantiopsida	<b>Marchantiophyta</b>				

<b>G</b>	At-BGal7			<b>Angiosperma (Magnoliophyta)</b>	
	At-BGal15				
	Gm-BGal15	Gm0025x00380			
	Gm-BGal16	Gm0289x00016			
	Gm-BGal17	Gm0005x00595			
	Gm-BGal18	Gm0111x00140			
	Gm-BGal19	Gm0017x10064			
	Pt-BGal11	I000098	Eudicotiledôneas		
	Pt-BGal12	eugene3.00180046			
	Pt-BGal13	eugene3.00280054			
	Vv-BGal8	VV78X105358			
	Os-BGal8	Os10g19960			Monocotiledôneas
	Os-BGal9	Os01g34920			
	Os-BGal10	Os06g42310			
	Os-BGal11	Os10g18400			
	Sb-BGal8	Sb03g023500			
	<i>Liriodendron tulipifera</i> -BGal3	DT587030 **	Magnoliids (Ord. Magnoliales)		

**H**

At-BGal6				
At-BGal16				
Gm-BGal20	Gm0065x00037			
Gm-BGal21	Gm0092x00360			
Gm-BGal22	Gm0065x00402			
Pt-BGal14	LG II000364	Eudicotiledóneas	<b>Angiosperma (Magnoliophyta)</b>	
Pt-BGal15	gw1.V.5384.1			
Pt-BGal16	gw1.41.450.1			
Pt-BGal17	gw1.V.2620.1			
Vv-BGal9	VV78X017300			
Vv-BGal10	VV78X217038			
Os-BGal12	Os05g35360			
Sb-BGal9	Sb09g021140	Monocotiledóneas		

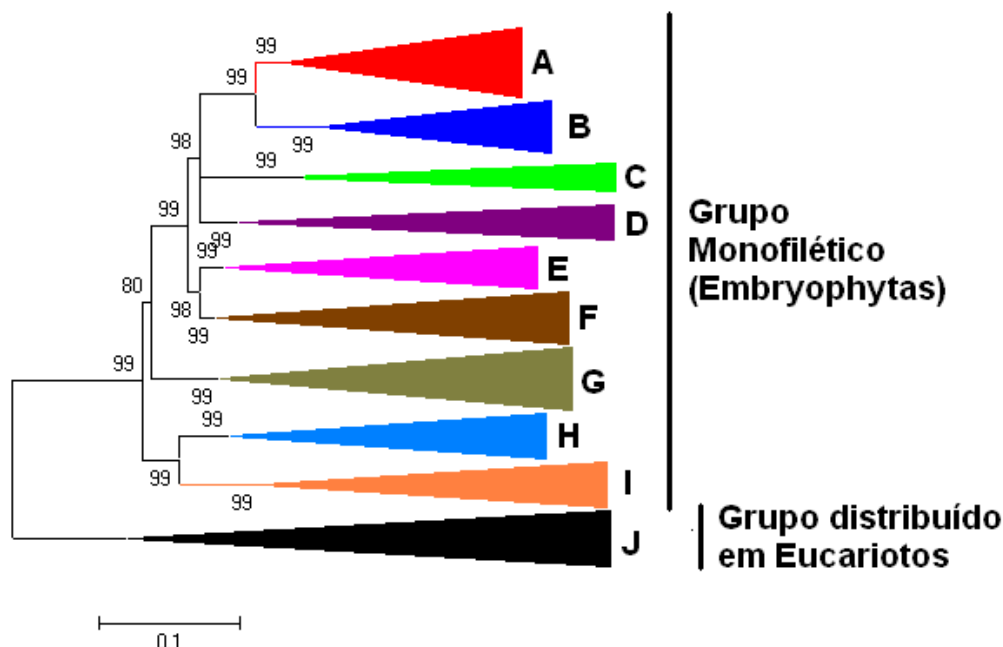
**I**

At-BGal11				
At-BGal13				
At-BGal14				
Gm-BGal23	Gm0053x10010			
Gm-BGal24	Gm0169x00134	Eudicotiledóneas	<b>Angiosperma (Magnoliophyta)</b>	
Pt-BGal18	gw1.IV.1165.1			
Pt-BGal19	IX000330			
Vv-BGal11	VV78X256435			
Vv-BGal12	VV78X101543			
Os-BGal13	Os09g36810			
Os-BGal14	Os08g43570			
Sb-BGal10	Sb02g031260	Monocotiledóneas		
Sb-BGal11	Sb07g024870			

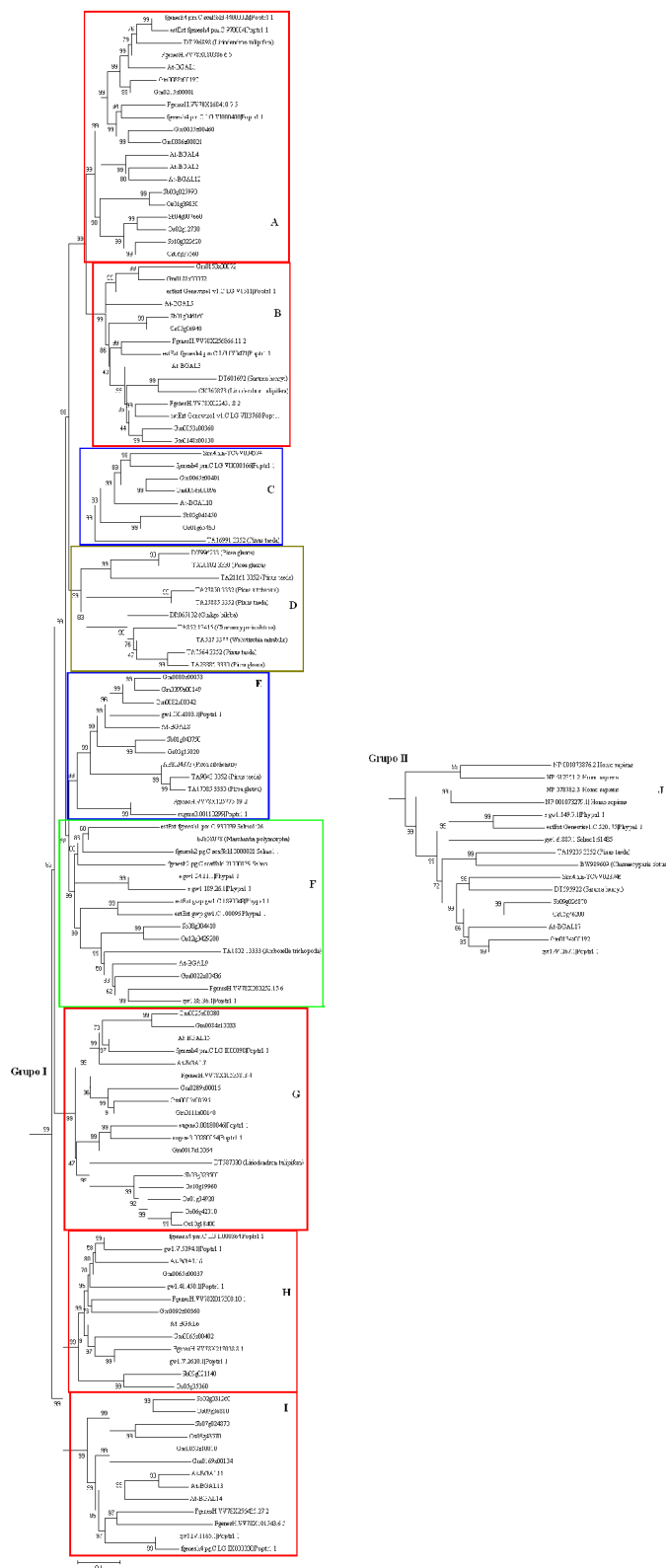
## Grupo II

	At-BGal17				
	Gm-BGal25	Gm0134x00192			
	Pt-BGal20	gw1.44.267.1	Eudicotiledóneas		
	Vv-BGal13	TCVV023946		Angiosperma	
	<i>Saruma henryi</i> -BGal2	DT595922 **	Angio basal (Ord. Amborellales)	(Magnoliophyta)	
	Os-BGal15	Os05g46200			
	Sb-BGal12	Sb09g026870	Monocotiledóneas		
	<i>Pinus taeda</i> -BGal6	TA19235_3352 **		Coniferophyta	
	<i>Chamaecyparis obtusa</i> -BGal2	BW989609 **	Coniferopsida		Eukaryota
	Sm-BGal4	61485	Isoetopsida	Lycopodiophyta	
J	Pp-BGal5	gw1.149.7.1			
	Pp-BGal6	520175	Bryopsida	Bryophyta	
	<i>Helicosporidium sp. ex Simulium jonessi</i>	CX128430 **	Trebouxiophyceae	Chlorophyta	
	GLB1L ( <i>Homo sapiens</i> )	NP_078782 ***	MGC10771		
	GLB1 ( <i>Homo sapiens</i> )	NP_001073279 ***	EBP, ELNR1		
	GLB1L3 ( <i>Homo sapiens</i> )	NP_001073876 ***	FLJ90231	Mammalia	Chordata
	GLB1L2 ( <i>Homo sapiens</i> )	NP_612351 ***	MST114, MSTP114		

\*\* N° de acesso: TIGR *Plant Transcript Assemblies*; \*\*\* N° de acesso: GeneBank.



**Figura 35.** Árvore filogenética das β-Galactosidasases. As relações filogenéticas foram inferidas por NJ, as distâncias foram calculadas por *p-distance* e 1000 réplicas de *bootstrap* utilizando 208 posições alinhadas de aminoácidos. O grupo monofilético de Embryophytas é suportado por *bootstrap* de 99%. O tamanho dos triângulos corresponde à quantidade relativa de genes.



**Figura 36.** Árvore filogenética das  $\beta$ -Galactosidasas. Os PoGOs, que aparecem compactados na figura 34, estão mostrados de maneira detalhada. As linhas vermelhas mostram grupos Angiosperma-específicos, as azuis Spermatophyta-específicos, a verde Embryophyta-específico e a linha marrom o grupo “Gimnosperma”-específico. Topologia calculada por NJ, distâncias por *p-distance* e 1000 réplicas de *bootstrap*.



### 3.2.4. Análises filogenéticas das $\beta$ -Glucosidases para a formação de PoGOs

Para elucidar as relações filogenéticas entre as  $\beta$ -Glucosidases ativas na desmontagem de xiloglucano utilizamos a seqüência da  $\beta$ -Glucosidase de *Tropaeolum majus* (CAA07070 – GeneBank), que foi caracterizada funcionalmente como parte do mecanismo de degradação de xiloglucano de reserva (semente) nesta espécie (Crombie *et al.*, 1998), para buscar por seqüências similares em genomas completos e bancos de ESTs através do protocolo mostrado na Figura 13. Obtivemos um total de 53 seqüências não redundantes. Análises por NJ dividiram as seqüências em dois grupos de genes homólogos, cada um contendo um único PoGO.

O PoGO A encontra-se distribuído ao longo das Embryophytas enquanto o PoGO B, que provavelmente emergiu do grupo A, ocorre somente em Angiospermas (Tabela 7; Figuras 37 e 38).

No PoGO A encontramos quatro genes de Arabidopsis (At5g20940, At5g20950, At5g04885 e At3g62710) que codificam proteínas que foram caracterizadas por Iglesias *et al* (2005) como ativas na degradação de oligossacarídeos de xiloglucano e localizadas no fluido apoplástico. O fato de nosso protocolo ter recuperado estas enzimas a partir de buscas realizadas com a seqüências de *T. majus* sugere que o protocolo foi eficiente e que provavelmente o PoGO A contem os conjuntos não redundantes de  $\beta$ -Glucosidase ativas sobre xiloglucano presentes nos genomas completos analisados.

O PoGO B possui quatro parálogos de Arabidopsis que se apresentam *in tandem* no cromossomo 3 (At3g47000, At3g47010, At3g47040 e At3g47050) o que sugere que se originaram por duplicações oriundas de erros de recombinação meiótica. Nenhum trabalho foi encontrado na literatura referente a genes do PoGO B, portanto estudos

funcionais são necessários para caracterizar a atividade hidrolítica e a especificidade de substrato das enzimas deste grupo que é específico de Angiospermas.

Assim como nos caso das XTHs e  $\beta$ -Galactosidases, nenhuma seqüência codificante para uma  $\beta$ -Glucosidase foi encontrada nos genomas completos de algas-verdes analisados. Buscas contra genomas de fungos e animais falharam em encontrar genes evolutivamente relacionados, porém obtivemos seqüências com alta similaridade em bactérias e escolhemos as duas mais similares para análises conjuntas com as enzimas encontradas em plantas (YP\_001634696 de *Chloroflexus aurantiacus* e YP\_001108596 de *Saccharopolyspora erythraea*; números de acesso do GeneBank). O resultado de Blastp 2x2 (bl2seq; Tatusova *et al.*, 1999) entre a seqüência de *C. aurantiacus* e a  $\beta$ -Glucosidase de *T. majus* pode ser visto na figura 39.

Análises por NJ e MP não conseguiram conectar estas seqüências com as  $\beta$ -Glucosidases específicas de plantas, porém o fato dessas enzimas aparentemente não estarem presentes em fungos ou animais (estes táxons possuem enzimas com atividade  $\beta$ -Glucosidase, porém não são evolutivamente relacionadas às  $\beta$ -Glucosidase ativas sobre xiloglucano exclusivas de plantas terrestres) e ocorrerem somente em bactérias sugere que as plantas podem ter herdado este grupo de genes da(s) bactéria(s) endossimbionte(s) fotossintética(s) que deu/deram origem ao cloroplasto. Esta suposição é reforçada pela seqüência bacteriana que apresentou maior similaridade (*e-value* de  $2e-145$ ), quando comparada com CAA07070, pertencer a uma bactéria fotossintetizante (*Chloroflexus aurantiacus*).

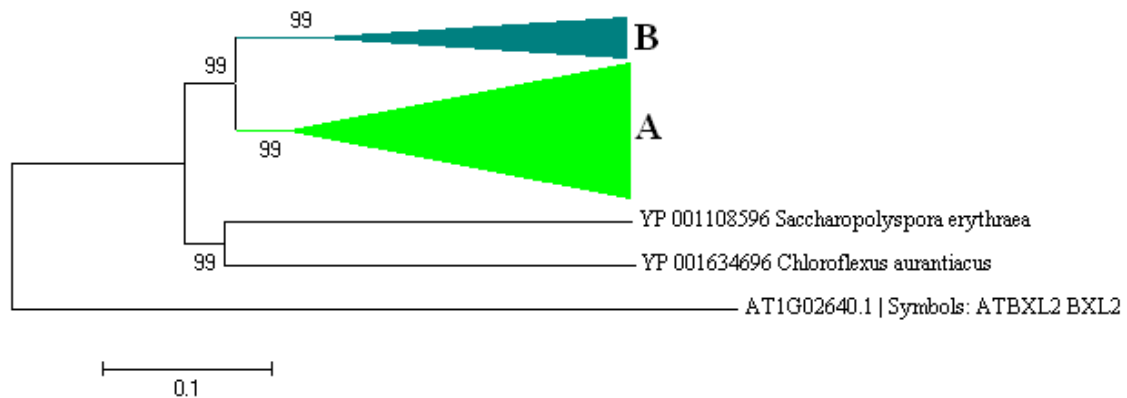
**Tabela 7.** Classificação das  $\beta$ -Glucosidases em PoGOs.

PoGO	$\beta$ -Glucosidase Nº	Código do gene	Sinônimo	Classe	Divisão ou Filo
	At-BGlu1	At5g20950			Embryophyta
	At-BGlu2	At5g20940			
	At-BGlu3	At5g04885			
	At-BGlu4	At3g62710			
	Pt-BGlu1	821016			
	Pt-BGlu2	826149			
	Pt-BGlu3	420479			
	Pt-BGlu4	200083		Eudicotiledôneas	
	Pt-BGlu5	421249			
	Gm-BGlu1	Gm0182x00021			
	Gm-BGlu2	Gm0020x00069			
	Gm-BGlu3	Gm0109x00047			
	Gm-BGlu4	Gm0031x00150			
	Gm-BGlu5	Gm0159x00091			
	Vv-BGlu1	VV78X205133			
	<i>T. majus</i> -BGlu1	CAA07070 ***			
A	Os-BGlu1	LOC_Os03g53800			
	Os-BGlu2	LOC_Os03g53790			
	Os-BGlu3	LOC_Os02g03870			
	Os-BGlu4	LOC_Os01g56510			
	Os-BGlu5	LOC_Os03g53860			
	Sb-BGlu1	Sb01g008040		Monocotiledôneas	
	Sb-BGlu2	Sb01g008050			
	Sb-BGlu3	Sb04g002570			
	Sb-BGlu4	Sb04g002560			
	Sb-BGlu5	Sb04g002580			
	Sb-BGlu6	Sb03g035970			
	Sb-BGlu7	Sb01g008030			
	<i>Welwitschia mirabilis</i> -BGlu1	TA1028_3377 **		Gnetopsida	Gnetophyta
	<i>Ceratopteris richardii</i> -BGlu1	BE643358 **		Filicopsida	Filicophyta
Sm-BGlu1	74114				
Sm-BGlu2	157260		Isoetopsida	Lycopodiophyta	
Pp-BGlu1	194433				
Pp-BGlu2	61002				
Pp-BGlu3	228446		Bryopsida	Bryophyta	
Pp-BGlu4	162943				
<i>Marchantia polymorpha</i> -BGlu1	TA1919_3197 **		Marchantiopsida	Marchantiophyta	

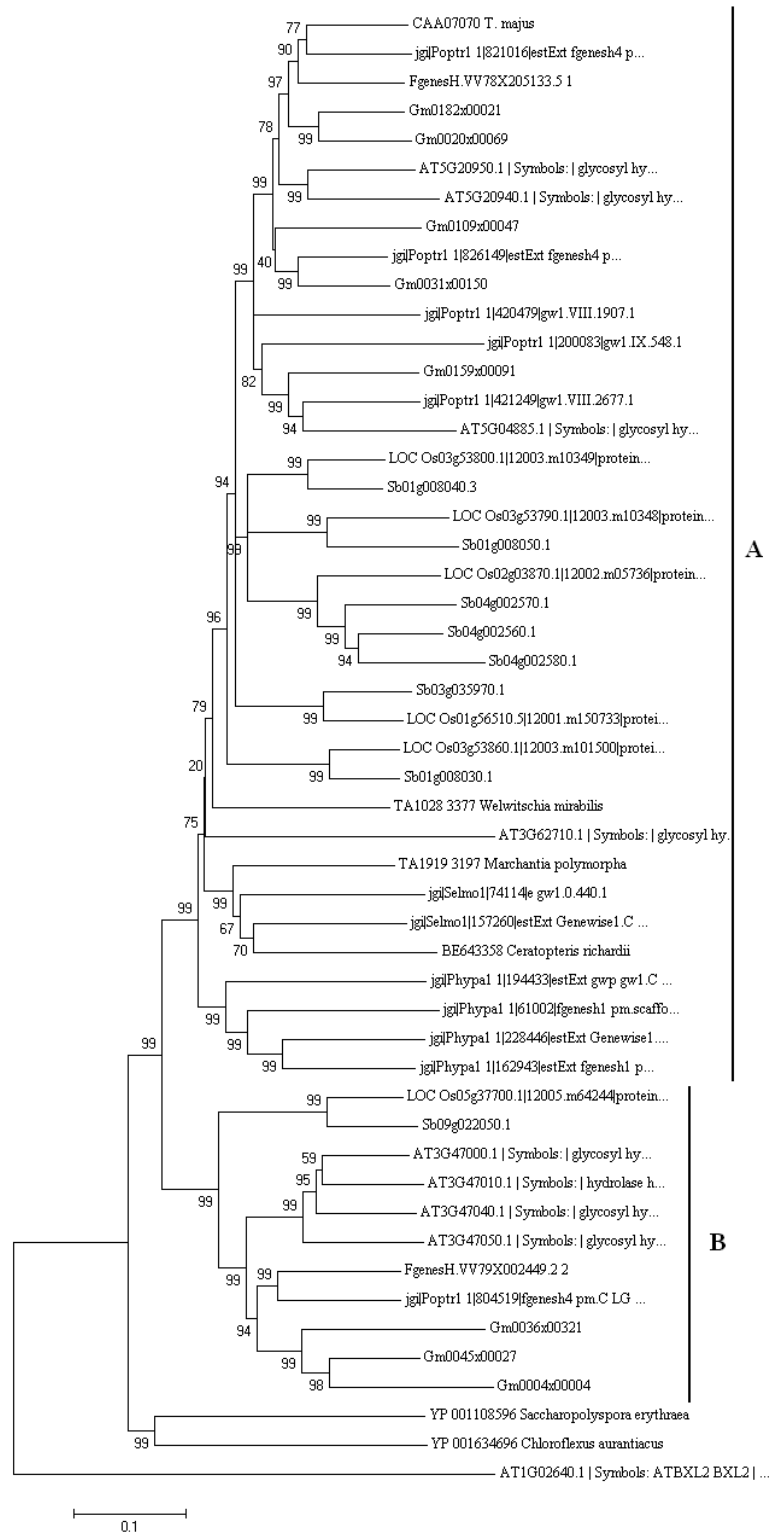
**B**

<i>Arabidopsis thaliana</i>	At3g47000		Angiosperma (Magnoliophyta)
<i>Arabidopsis thaliana</i>	At3g47010		
<i>Arabidopsis thaliana</i>	At3g47040		
<i>Arabidopsis thaliana</i>	At3g47050		
<i>Vitis vinifera</i>	VV79X002449	Eudicotiledôneas	
<i>Populus trichocarpa</i>	804519		
<i>Glycine max</i>	Gm0036x00321		
<i>Glycine max</i>	Gm0045x00027		
<i>Glycine max</i>	Gm0004x00004		
<i>Oryza sativa</i>	LOC_Os05g37700		
<i>Sorghum bicolor</i>	Sb09g022050	Monocotiledôneas	

\*\* N° de acesso: TIGR *Plant Transcript Assemblies*; \*\*\* N° de acesso: GeneBank.



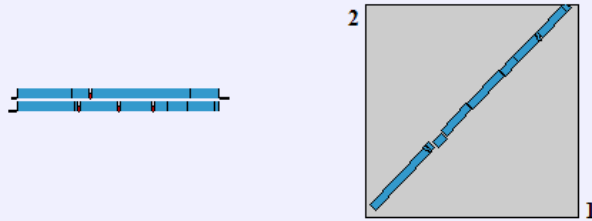
**Figura 37.** Árvore filogenética compactada das  $\beta$ -Glucosidasas específicas de plantas. Os PoGOs A e B formam um grupo monofilético em plantas Embryophytas. O tamanho dos triângulos corresponde à quantidade relativa de genes. A topologia foi inferida por NJ com distâncias obtidas por *p-distance*. Foram realizadas 1000 réplicas de *bootstrap*. AT1G02640 é uma  $\beta$ -xilosidase (ATBXL2), que apresenta similaridade em nível de aminoácidos com as  $\beta$ -Glucosidasas, foi incluída na análise.



**Figura 38.** Árvore filogenética das  $\beta$ -Glucosidasases específicas de plantas. Os PoGOs, que aparecem compactados na figura 34, estão mostrados. A seqüência do gene At1g02640 (Beta-xilosidase 2) de Arabidopsis foi utilizada como grupo externo na tentativa, sem sucesso, de buscar a relação filogenética das seqüências Embryophyta-específicas com as seqüências bacterianas YP\_001108596 e YP\_001634696. A topologia foi inferida por NJ, com distâncias calculadas por *p-distance* e 1000 réplicas de *bootstrap*.

**Sequence 1:** CAA07070 beta-glucosidase [*Tropaeolum majus*]  
 Length = 654 (1 .. 654)

**Sequence 2:** YP\_001634696 *Chloroflexus aurantiacus*  
 Length = 619 (1 .. 619)



**NOTE:** Bitscore and expect value are calculated based on the size of the nr database.



Score = 521 bits (1341), Expect = 2e-145  
 Identities = 292/611 (47%), Positives = 380/611 (62%), Gaps = 35/611 (5%)

```

Query 24  AEYMRYKDPKKPLNVRIKDLMSRMTLAEKIGQMTQIERKEATPDVISKYFIGSVLSGGGS 83
          A + Y+DP P+ R+ DL+ RMTLAEKIGQMT IE+ TPD++ IG VLSGGG
Sbjct 31  ATVITYRDPSPAPIAERVNDLLQRMTLAEKIGQMTLIEKNSLTPDLVRDLAIGGVLSGGGG 90

Query 84  VPAPKASPEAWVDLVNGMQKAALSTRLGIPMIYGDVAVHGHNNVYNATIFPHNVGLGVTR 143
          P + SP AW +VN Q+AALSTRLGIP+IYG D VHGHN+Y A IFPHN+GLG
Sbjct 91  YPQAENSPAAWAAMVNEFQQAALSTRLGIPLIYGADGVHGHNNLYGAVIFPHNIGLGAAN 150

Query 144 DPALIKRIGEATALECRATGIPYAFAPCIAVCRDPRWGRCYESYSE--DHTIVQAMTEII 201
          +P L+++IG ATALE ATG+ + +AP + V D RWGR YE Y+E DH A + +
Sbjct 151 NPQLMEQIGRATALEMAATGVFNWYAPAVMVPLDVRWGRTYEGYAERPDPHVAALA-SAFL 209

Query 202  PGLQGDVPPDVKKGVPFVGGKTKVAACAKHFVGDGGTTKG-----IDENNTVIDSRG 253
          GLQ P + +V KHF+GDGGT G +D+ T D
Sbjct 210  RGLQ-----APDIAAPNRVIGTPKHFLDGGTAWGSSTENYKLDQGETFGDEAF 259

Query 254  LFSIHMPAYHDSIKKGVATVMVSYSSWNGLRMHANRDLVTGYLKNKLFKRFVVISDWEGI 313
          L ++H+P Y I +G T+M SYSSWNG +MHA+ +T LK +L F GFV+SDW I
Sbjct 260  LRAVHLPPYQALIAEGAQTIMASYSSWNGQKMHASSYWLTDVLKRELGFAGFVVDWAAI 319

Query 314  DRITDPPGRNYSYSVEAGVGAGIDMIMVPEDFTKFLNELTSQVKKNIIPMSRIDDAVKRI 373
          D+I+ +Y +V + AGIDM MVP D +F++ LT V++ + RIDDAV+RI
Sbjct 320  DQISF----DYDQAVITAINAGIDMNMVYDAQRFDLSLTRAVERGAVSEERIDDAVRI 375

Query 374  LRVKFMVGLFESPLADYSLANQLGSQEHRLAREAVRKSLSVLLKNGESADKPFVPLPKNA 433
          L VKF MGLFE P A +L++Q+GS +HR LAR AV +SLVLLKN D +PLPK+
Sbjct 376  LTVKFAMGLFEQPFHAHTALSQIGSAQHRQLARTAVAQSLVLLKN----DANLLPLPKDI 431

Query 434  KKILVAGSHADNLGRQCGGWTIEWQGVNNDLTTGTTILNAIKKTVDEPTTQVIYNENPDS 493
          + + G A +LG Q GGWTIEWQG G + GTTIL I+ V P T V YN++
Sbjct 432  GHLYIGGQAHDLGIQAGGWTIEWQGRIG-PIIPGTTILEGIQAAVSPQTVVEYNQHGRF 490

Query 494  NYVKTNSFDYAIVVVGEPYAEQGDSEFNLTIEPQGPTTI--SSVCGAVKCVVVVISGRP 551
          + I VVGE PYAE +GDS +L++P P + V+ VVV+++GRP
Sbjct 491  TGDPGAADAVCIAVVGELPYAEGRGDSASLSLP-PAENRVLRRMEEACVRLVVVLVAGRP 549

Query 552  VVLQPYVSYMDALVAAWLPGTEGQGVTDVLFVGDYGFYGTGKLARTWFKITVDQLPMNVGDKHY 611
          +++ + DALV AWLPG+EG GV DVLFGD F G+L TW +++DQLP+ GD
Sbjct 550  LLVTDDLKWDALVMAWLPGSEGAADVLFVGDQPFGRGLPVTWPRSLDQLPVGSGDG-- 607

Query 612  DPLFFFGFGLT 622
          PLFP+GFGLT
Sbjct 608  QPLFFPYGFGLT 618
  
```

**Figura 39.** Resultado do alinhamento por Blastp 2x2 entre as seqüências YP\_001634696 (*Chloroflexus aurantiacus*) e CAA07070 (*Tropaeolum majus*). As seqüências apresentam sobreposição de 62% e identidade de 47% com *e-value* de 2e-145.

### 3.2.5. Análises filogenéticas das $\alpha$ -Xilosidases e $\alpha$ -Glucosidases e organização em PoGOs

Para elucidar a história evolutiva das  $\alpha$ -Xilosidases realizamos buscar ao longo dos genomas disponíveis de plantas-verdes usando como “isca” as seqüências peptídicas da  $\alpha$ -Xilosidase que faz parte do mecanismo de degradação de xiloglucano de reserva em *Tropaeolum majus* (CAA10382 – GeneBank; Crombie *et al.*, 2002) e do gene ATXYL (At1g68560; Sampedro *et al.*, 2001) de *Arabidopsis*. As buscas resultaram em 24 genes evolutivamente relacionados que formam um grupo monofilético no reino Viridiplantae e divididos em 3 PoGOs (Tabela 8; Figura 40).

O PoGO C contém as duas seqüências caracterizadas como  $\alpha$ -Xilosidases utilizadas para realizar as buscas. O gene ATXYL possui um parólogo (At3g45940) que não possui qualquer tipo de caracterização funcional, mas que possivelmente encoda uma proteína com a mesma atividade. Encontramos, ainda, o EST TA3201\_71647 (TIGR) da conífera *Pinus pinaster*, o que sugere que este grupo existia no ancestral comum das Spermatophytas.

Interessantemente, o PoGO A que é grupo irmão do PoGO C, apresenta um único gene de *Arabidopsis* (AGLU1 – At5g11720) que foi anotado, apenas por comparação, como sendo uma  $\alpha$ -Glucosidase e é um candidato a caracterização funcional por possivelmente desempenhar uma função única em *Arabidopsis*. Para validar o PoGO A como tendo atividade de  $\alpha$ -Glucosidase buscamos na literatura por proteínas caracterizadas com essa atividade em plantas e encontramos a proteína AAB02985 (GeneBank) de cevada (*Hordeum vulgare*; Tibbot & Skadsen, 1996). Esta proteína agrupou-se nas análises por NJ e MP no PoGO A, juntamente com as

seqüências ortólogas de sorgo e arroz com valor de *bootstrap* de 100% em ambas as análises.

Encontramos ainda, usando uma análise por MP, o PoGO B, que contém genes das algas Chlorophytas *Volvox carteri* e *Chlamydomonas reinhardtii* e da Bryophyta *Physcomitrella patens*. (Figura 41). Infelizmente nenhum destes genes possui qualquer tipo de estudo funcional, portanto não se sabe se possuem atividade  $\alpha$ -Glucosidase ou  $\alpha$ -Xilosidase.

Visando elucidar qual das duas atividades hidrolítica é a mais derivada (mais recente) na história evolutiva do reino Viridiplantae, buscamos pelas seqüências peptídicas e genômicas mais similares em genomas dos fungos *Pichia stipitis* e *Schizosaccharomyces pombe* e dos cordados *Dario rerio* (peixe, paulistinha) e humano. Todas as seqüências recuperadas com alta similaridade correspondiam a proteínas com atividade  $\alpha$ -Glucosidase. As proteínas XP\_001383835 (*P. stipitis*) e NP\_593996 (*S. pombe*) agruparam, com valor de *bootstrap* de 53% em análises por NJ, externamente a todas as seqüências do reino Viridiplantae em questão. As seqüências de animais permaneceram externas a estas, no entanto não obtivemos *bootstrap* significativo (> 50%). Concluimos, com estes resultados, que a atividade  $\alpha$ -Glucosidase é a atividade original deste grupo de hidrolases relacionadas evolutivamente. A atividade  $\alpha$ -Xilosidase dentro deste grupo pode ter surgido como uma neofuncionalização a partir da duplicação do gene original, uma  $\alpha$ -Glucosidase, dentro da linhagem das plantas. Como a Lycophyta *Selaginella moellendorffii*, que está na base das plantas Tracheophytas, possui somente uma gene dentro deste grupo e encontramos um EST da conífera *Pinus pinaster*, que agrupou com valor de 100% no grupo das  $\alpha$ -Xilosidasas, sugerimos que o surgimento da atividade derivada ( $\alpha$ -Xilosidase) ocorreu no ancestral

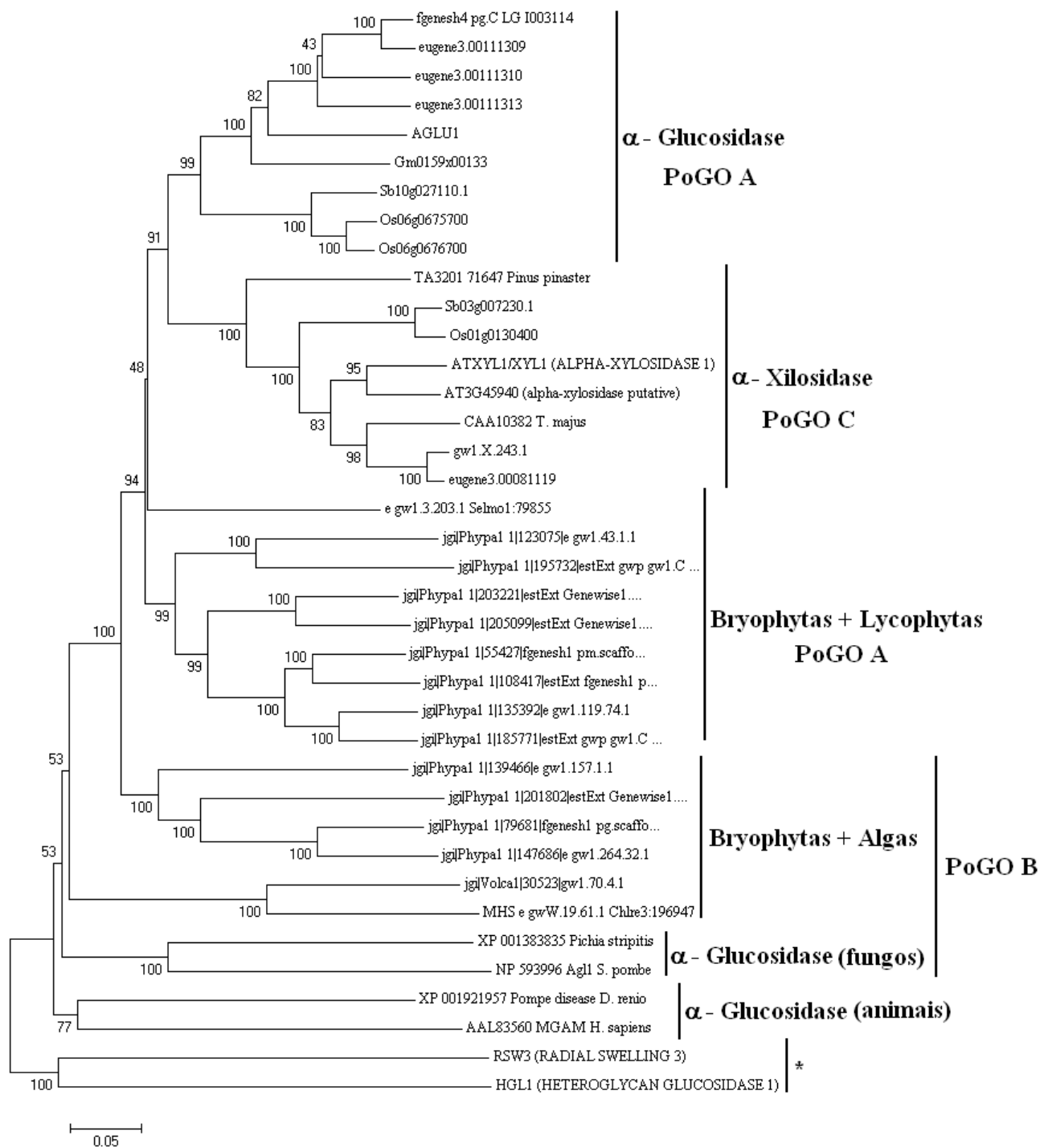


comum das Spermatophytas e pode estar relacionada ao surgimento da semente ou ao ganho de complexidade estrutural que marcam o grupo das plantas com semente.

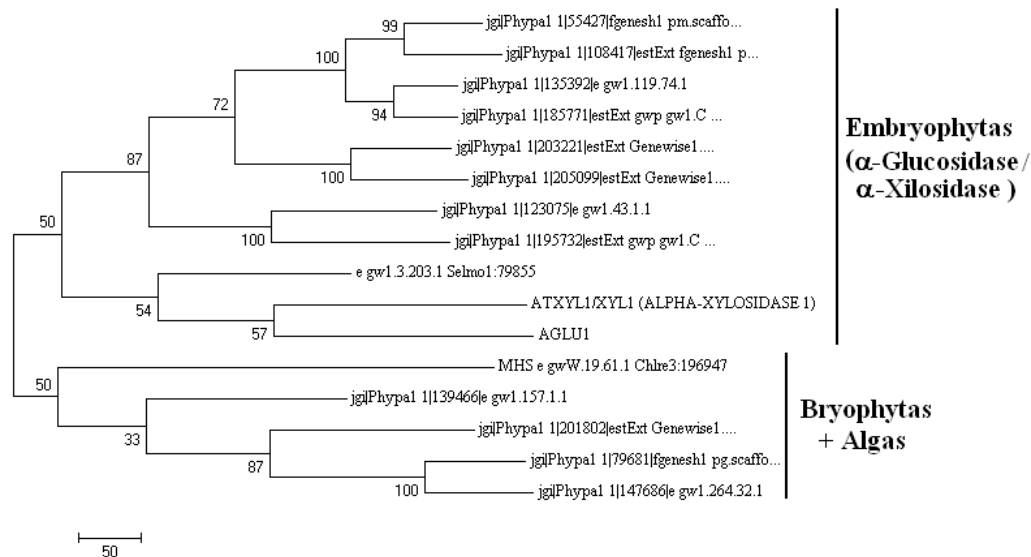
**Tabela 8.** Agrupamento das  $\alpha$ -Xilosidases e  $\alpha$ -Glucosidases em PoGOs.

PoGO	$\alpha$ -Glucosidase / $\alpha$ -Xilosidase N°	Código do gene	Sinônimo	Classe	Divisão ou Filo	
<b>A</b>	At-AGlu1	At5g11720	AGLU1	Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Embryophyta</b>
	Gm-AGlu1	Gm0159x00133				
	Pt-AGlu1	I003114				
	Pt-AGlu3	111309				
	Pt-AGlu3	111310				
	Pt-AGlu4	111313				
	Os-AGlu1	Os06g0675700		Monocotiledôneas		
	Os-AGlu2	Os06g0676700				
	Sb-AGlu1	Sb10g027110				
	Sm-AGlu1	79855		Isoetopsida	<b>Lycopodiophyta</b>	
	Pp-AGlu1	123075		Bryopsida	<b>Bryophyta</b>	
	Pp-AGlu2	195732				
	Pp-AGlu3	203221				
	Pp-AGlu4	205099				
	Pp-AGlu5	55427				
Pp-AGlu6	108417					
Pp-AGlu7	135392					
Pp-AGlu8	185771					
<b>B</b>	Pp-Agl1	139466		Bryopsida	<b>Bryophyta</b>	<b>Bryophyta + Algae</b>
	Pp-Agl2	201802				
	Pp-Agl3	79681				
	Pp-Agl4	147686				
	Vc-Agl1	30523		Chlorophyceae	<b>Chlorophyta</b>	
	Cr-Agl1	196947				
CGA1 ( <i>Pichia stipitis</i> )	XP_001383855 ***		Saccharomycetes		<b>Fungi</b>	
Agl1 ( <i>Schizosaccharomyces pombe</i> )	NP_593996 ***		Schizosaccharomycetes	<b>Ascomycota</b>		
<b>C</b>	At-Xyl1	At1g68560	ATXYL1	Eudicotiledôneas	<b>Angiosperma (Magnoliophyta)</b>	<b>Spermatophyta</b>
	At-Xyl2	At3g45940				
	Gm-Xyl1	Gm0223x10016				
	Gm-Xyl2	Gm0159x00133				
	Pt-Xyl1	gw1.X.243.1				
	Pt-Xyl2	81119				
	Os-Xyl1	Os01g0130400		Monocotiledôneas		
	Sb-Xyl1	Sb03g007230				
	<i>Pinus pinaster</i> -Xyl1	TA3201_71647 **				

\*\* N° de acesso: TIGR *Plant Transcript Assemblies*; \*\*\* N° de acesso: GeneBank.



**Figura 40.** Análise filogenética das  $\alpha$ -Xilosidasas e  $\alpha$ -Glucosidasas. A topologia foi inferida por NJ, as distâncias calculadas por *p-distance*. Foram realizadas 1000 réplicas de *bootstrap*. As proteínas RSW3 (At5g63840) e HGL1 (At3g23640) foram utilizadas como grupo externo(\*).

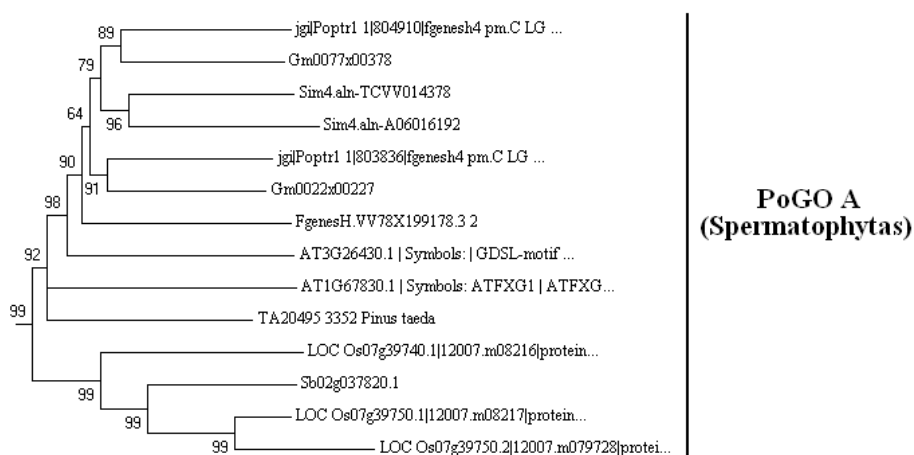


**Figura 41.** Análise filogenética por Máxima Parcimônia das seqüências de Bryophytas. A análise por MP mostrou a existência de dois grupos distintos de genes em Bryophytas. Um deles forma um grupo monofilético com as  $\alpha$ -Xilosidasas e  $\alpha$ -Glucosidasas de Angiospermas e o outro possivelmente representa um grupo basal de  $\alpha$ -Glucosidasas em Eucariotos.

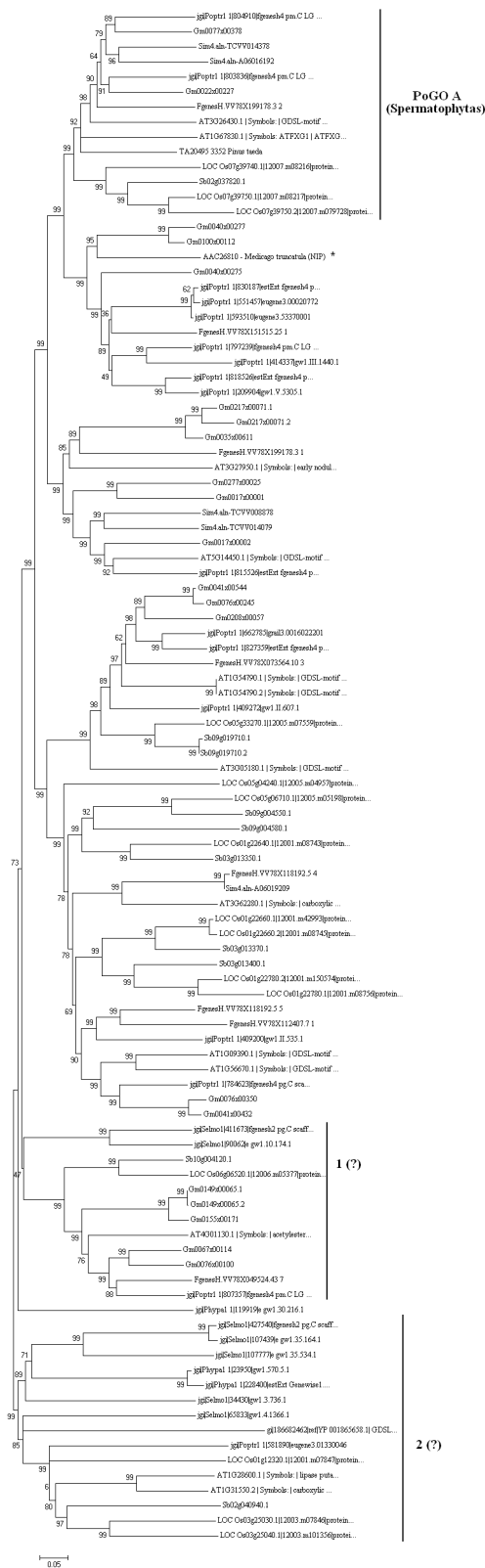
### 3.2.6. Análises filogenéticas das $\alpha$ -Fucosidasas e organização em PoGOs.

Inicialmente buscamos na literatura por enzimas que apresentassem a capacidade de liberar fucose de oligossacarídeos de xiloglucanos. Encontramos duas enzimas que foram caracterizadas funcionalmente com esta atividade ( $\alpha$ -(1-2)fucosidase): ATFXG1 de *Arabidopsis thaliana* (At1g67830; de La Torre *et al.*, 2002) e EBM II (BAF85832 - GeneBank; Ishimizu *et al.*, 2007) de *Lilium longiflorum*. Curiosamente estas proteínas apresentavam baixa similaridade entre si e nossas análises mostraram que se trata de proteínas não relacionadas evolutivamente. Ishimizu *et al.* (2007) demonstraram que EMBII representa uma nova classe de  $\alpha$ (1-2)fucosidase capaz de agir sobre oligossacarídeos de xiloglucano. Utilizamos estas duas enzimas nas buscas para encontrar genes relacionados ao longo dos genomas completos do reino Viridiplantae e dos bancos de ESTs de nosso interesse.

ATFXG1 mostrou-se fazer parte de um grande grupo de proteínas do qual não fomos capazes de encontrar qualquer outra evidência funcional. Estas proteínas, em *Arabidopsis thaliana* (os diversos genes parálogos), encontram-se anotadas no banco de dados de *Arabidopsis*, TAIR ([www.arabidopsis.org](http://www.arabidopsis.org)), genericamente como “GDSL-*motif proteins*” com atividade “lipase/hidrolase *putative*”. Pela total ausência de conhecimento a cerca destas hidrolases optamos por classificar somente o PoGO onde se encontra ATFXG1. Este grupo contém seqüências dos genomas completos de Angiospermas e um EST da conífera *Pinus taeda* (Grupo específico de Spermatophyta) e foi nomeado como PoGO A (Tabela9; Figura 42). Seqüências peptídicas das plantas mais basais *Selaginella* e *Physcomitrella* foram encontradas, mas pertencendo a outros possíveis grupos de ortólogos (Figura 43). Ainda são necessários estudos funcionais envolvendo os representantes desta família multigênica de prováveis hidrolases (GDSL-*motif proteins*) para que a atividade  $\alpha(1-2)$ fucosidase possa ser estendida para ramos mais basais da linhagem das plantas terrestres e para que possamos lançar luz sobre como este grupo de enzimas evoluiu.



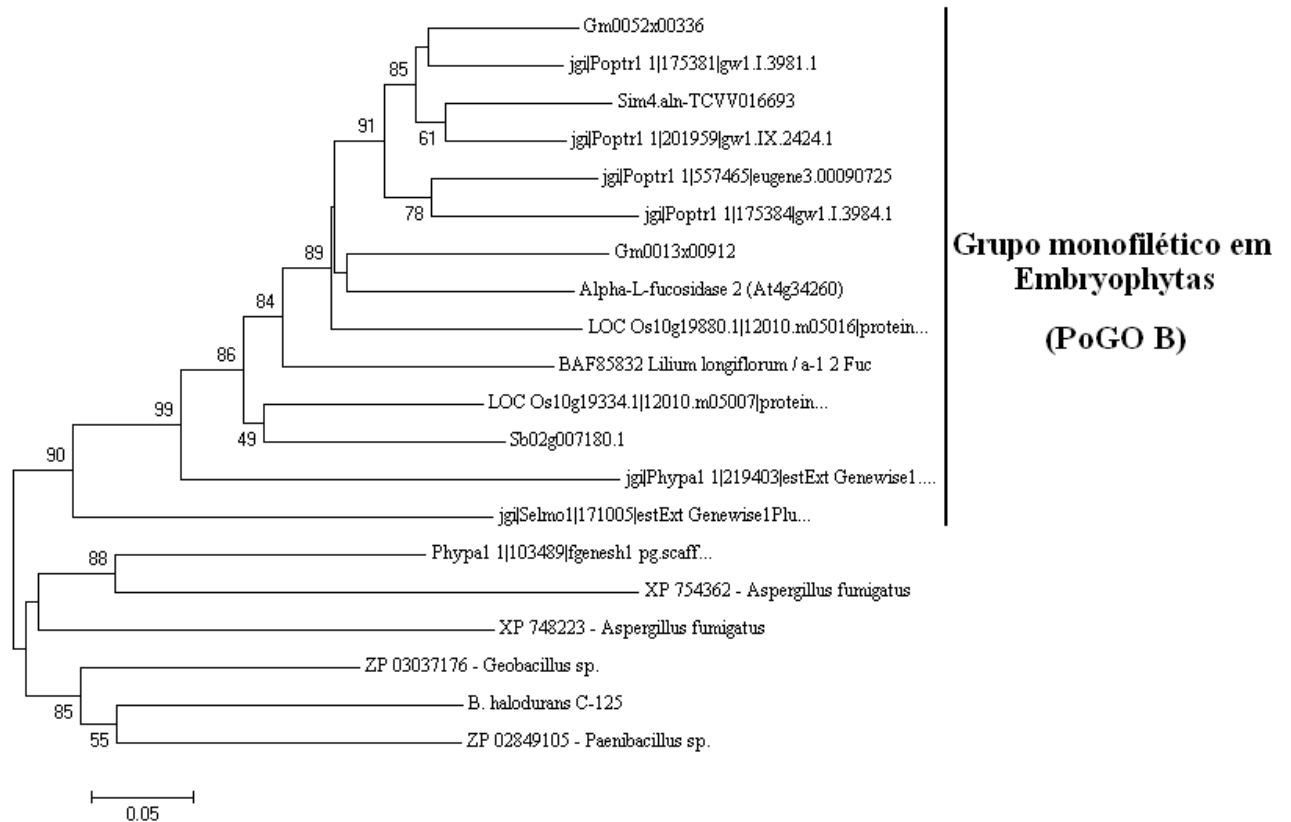
**Figura 42.** Possível grupo de ortólogos de ATFXG1. A figura é uma representação parcial da árvore filogenética dos 100 primeiros *hits* de Blast ATFXG1 contra os genomas completos utilizados (ver material e métodos). A topologia foi inferida por NJ, com distâncias obtidas por *p-distance* e foram realizadas 1000 replicatas de *bootstrap*. Os valores de *bootstrap* (em porcentagem) estão mostrados nos ramos da árvore.



**Figura 43.** Árvore filogenética dos 100 primeiros *hits* de Blastp e tBlastn de ATRXG1. O PoGO A (figura 42) aparece no topo da árvore filogenética. 1 e 2 mostram PoGOs contendo proteínas de, respectivamente, Tracheophytas (incluindo *Selaginella*) e Embryophytas (incluindo *Physcomitrella*). A topologia foi inferida por NJ, com distâncias obtidas por *p-distance*. Foram realizadas 1000 réplicas de *bootstraps*.

Utilizando o outro tipo de  $\alpha(1-2)$ fucosidase (chamaremos de  $\alpha$ -fucosidase II) ativa sobre oligos de xiloglucano descoberta recentemente (EBM II, BAF85832 - GeneBank; Ishimizu *et al.*, 2007) como “isca” para encontrar seus ortólogos, obtivemos um único possível grupo de ortólogos (PoGO B; Figura 44), que se encontra distribuído ao longo das Embryophytas e contém as únicas seqüências com similaridade à BAF85832 contidas no genomas destas plantas. Uma grande quantidade de seqüências similares existente em genomas de bactérias (a grande parte bactérias patogênicas de plantas, como *Xanthomonas*, por exemplo) e em alguns fungos foi encontrada. A seqüência que apresentou maior similaridade com EBM II foi a proteína hipotética NP\_412708 (GeneBank; Figura 45) da bactéria do solo *Bacillus halodurans*. Esta bactéria possui em seu genoma 112 genes codificantes para transposases, que provavelmente tem um importante papel na transferência horizontal de genes (Takami *et al.*, 2000). Isso, somado ao fato de este tipo de  $\alpha(1-2)$ fucosidase aparentemente não existe em algas e aparecer essencialmente em genomas de bactérias patogênicas de plantas (estas enzimas podem ser usadas para degradação da parede celular de plantas), sugere que este grupo de  $\alpha(1-2)$ fucosidase II, que parece ser monofilético (ter uma única origem) em plantas terrestres, surgiu nesta linhagem por transferência horizontal de um gene relacionado a patogenicidade em bactérias. Este evento de transferência, que provavelmente ocorreu muito cedo na evolução das Embryophytas, e o fato destes genes terem se mantido funcionais por todo este tempo e sofrido processos de duplicação (quatro genes em *Populus*, por exemplo) sugere que conferiram vantagens adaptativas na linhagem das plantas terrestres. Não se pode excluir o fato deste grupo de genes terem sido simplesmente perdido em algas o que pode excluir a possibilidade de transferência horizontal.

Os genes que compõem os PoGOs A e B podem ser vistos na Tabela 9.



**Figura 44.** Árvore filogenética mostrando o grupo monofilético (PoGO B) de Embryophytas. A topologia foi inferida por NJ e as distâncias calculadas com *p-distance*. Para testar estatisticamente o peso dos nós, 1000 réplicas de *bootstrap* foram conduzidas.



**Figura 45.** Resultado do alinhamento por Blastp 2x2 entre as seqüências NP\_412708 (*Bacillus halodurans*) e BAF85832 (*Lilium longiflorum*). As seqüências apresentam sobreposição de 57% e identidade de 42% com *e-value* de 5e-167.



**Tabela 9.** Agrupamento das  $\alpha$ -Fucosidases em PoGOs.

PoGO	$\alpha$ -Fucosidase I / $\alpha$ -Fucosidase II N <sup>o</sup>	Código do gene	Sinônimo	Classe	Divisão ou Filo	
<b>A</b>	At-Fuc1	At1g67830		Eudicotiledôneas	Angiosperma (Magnoliophyta)	Spermatophyta
	At-Fuc2	At3g26430				
	Gm-Fuc1	Gm0077x00378				
	Gm-Fuc2	Gm0022x00227				
	Pt-Fuc1	804910				
	Pt-Fuc2	803836				
	Vv-Fuc1	TCVV014378				
	Vv-Fuc2	A06016192				
	Vv-Fuc3	VV78x199178				
	Os-Fuc1	LOC_Os07g39740		Monocotiledôneas		
	Os-Fuc2	LOC_Os07g39750				
	Sb-Fuc1	Sb02g037820				
		<i>Pinus taeda</i> -Afuc1	TA9043_3352 **		Coniferopsida	Coniferophyta

<b>B</b>	At-FucII1	At4g34260		Eudicotiledôneas	Angiosperma (Magnoliophyta)	Embryophyta
	Gm-FucII1	Gm0052x00336				
	Gm-FucII2	Gm0013x00912				
	Pt-FucII1	175381				
	Pt-FucII2	201959				
	Pt-FucII3	557465				
	Pt-FucII4	175384				
	Os-FucII1	LOC_Os10g19880		Monocotiledôneas		
	Os-FucII2	LOC_Os10g19334				
	Sb-FucII1	Sb02g007180				
		<i>Lilium longiflorum</i> -FucII1	BAF85832***	EBM II		
	Sm-FucII1	171005		Isoetopsida	Lycopodiophyta	
	Pp-FucII1	219403		Bryopsida	Bryophyta	

\*\* N<sup>o</sup> de acesso: TIGR *Plant Transcript Assemblies*; \*\*\* N<sup>o</sup> de acesso: GeneBank.

## **4. Conclusões e Perspectivas**

### **4.1. Bibliotecas de cDNA**

Obtivemos 103 seqüências (em sua maioria parciais) de cDNA de Jatobá, o que corresponde aos primeiros ESTs seqüenciados desta espécie. Ainda é necessário seqüenciar um número mais amplo de clones para melhor avaliar o real índice de novidade das bibliotecas.

Como perspectiva de curto prazo está o seqüenciamento de um maior número de clones das bibliotecas que temos e a deposição das seqüências obtidas no banco de dados NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Estas seqüências poderão ser usadas para ampliar o conhecimento sobre as relações filogenéticas dentro das leguminosas.

Como perspectivas de médio o seqüenciamento de um número amplo de clones permitirá a montagem do transcriptoma do Jatobá, algo inédito para uma árvore tropical nativa do Brasil. Em longo prazo a informação destas seqüências poderá ser utilizada na produção de microarranjos de cDNA para abordagens rápidas na avaliação da resposta do transcriptoma de uma árvore tropical, que compõe importantes biomas de interesse internacional como a Amazônia e a Mata Atlântica, aos estresses oriundos do aquecimento global e efeito estufa, o que permitirá obter as primeiras informações moleculares sobre os efeitos do aquecimento global e aumento da concentração de CO<sub>2</sub> sobre o transcriptoma de uma árvore tropical.

## **4.2. Análise filogenética das enzimas hidrolíticas de xiloglucano no reino Viridiplantae**

As enzimas necessárias para degradação de xiloglucano são encontradas ao longo de todos os grupos de Spermatophytas. Todas as enzimas, menos as  $\alpha$ -Xilosidases e as  $\alpha$ -Fucosidases, são encontrada em todas as Embryophytas. A quantidade de grupos de ortólogos em Angiospermas, de forma geral, sempre foi maior que nos outros táxons, o que sugere um ganho progressivo de complexidade do metabolismo e estrutura da parede celular ao longo da evolução das plantas terrestres.

Apesar de o xiloglucano ser pouco abundante na hemicelulose de plantas não-eudicotiledôneas uma grande quantidade de genes relacionados à degradação de xiloglucano foram encontrados em todos os grupos de Embryophytas. Isso sugere que estas enzimas, além do próprio xiloglucano, possam ter atividade em diversos outros polímeros constituintes da hemicelulose, como galactomananos e arabinogalactanos, já que são específicas de ligação entre açúcares (ligação  $\beta(1-4)$  entre glicoses, por exemplo) que não são exclusivas de xiloglucano.

A conexão entre as XTHs exclusivas de plantas e o grupo que encontramos de  $\beta(1-4)$ -glucanases/transglicosilases que ocorre em plantas e fungos sugere que a transglicosilação de polímeros de glicose é um processo essencial no controle da plasticidade morfológica e no crescimento celular das duas principais linhagens de eucariotos que possuem parede celular. Evidências recentes mostram que os genes de fungos supostamente relacionados às XTHs também participam destes processos celulares, porém este grupo de enzimas foi muito pouco amplificado nos genomas de fungos que avaliamos, ao contrário do que aconteceu com as XTHs em plantas

terrestres, o que pode ser uma das explicações moleculares para a maior complexidade e plasticidade morfológica de células e tecidos em vegetais, quando comparada com a organização baseada em hifas, muito mais simples, dos fungos. Também foi demonstrado que em fungos estes genes participam de processos onde ocorre a interação das cadeias de  $\beta$ -glucanos com os filamentos ordenados de quitina, o é semelhante ao que acontecem em plantas, mas estas, ao contrário dos fungos, possuem celulose como polímeros filamentosos de açúcar, que interagem com xiloglucanos e as demais hemiceluloses no processo de controle estrutural da parede celular. Seriam duas estratégias diferentes, mas possivelmente relacionadas para o controle da plasticidade adaptativa da parede celular.

As  $\beta$ -Galactosidases exclusivas de plantas apresentaram grande amplificação do número de cópias ao longo da evolução das Embryophytas, ao contrário do único grupo de ortólogos que ocorre em todos os eucariotos. Isso sugere que estas hidrolases participam de processos exclusivos da linhagem das plantas terrestres, especialmente nas Angiospermas.

As  $\beta$ -Glucosidases ativas sobre xiloglucano ocorrem em todas as Embryophytas e apresentam alta similaridade com seqüências de bactérias fotossintéticas, o que pode significar que estes genes, em plantas, foram originados da transferência de genes oriundos da linhagem de bactérias fotossintéticas que foram englobadas por células eucarióticas do *pool* inicial que originou o reino Viridiplantae.

As  $\alpha$ -Xilosidases ocorrem em plantas Spermatophytas e provavelmente são um caso de neofuncionalização que ocorreu a partir de genes que codificavam para  $\alpha$ -Glucosidases que tem similaridade em todos os eucariotos.

Duas linhagens de genes em plantas possuem atividade  $\alpha(1-2)$ -Fucosidase ativa sobre oligossacarídeos de xiloglucano. O PoGO das  $\alpha$ -Fucosidases I que encontramos ocorre em Spermatophytas e faz parte de uma extensa família multigênica (*GDSL-motif proteins*) que ocorre em plantas Embryophytas, no entanto carecemos de dados funcionais para estender a existência desta atividade hidrolítica para os grupos mais basais de Embryophytas, bem como ligar esta atividade a enzimas que eventualmente ocorram em outros reino biológicos. A  $\alpha$ -Fucosidases II ocorre em Embryophytas e possui somente um PoGO que possui seqüências com alta similaridade com proteínas hidrolíticas de bactérias do solo e/ou patogênicas de plantas, o que pode significar uma possível transferência horizontal de bactérias destas linhagens para as plantas terrestres. Este evento provavelmente ocorreu muito cedo na história evolutiva das Embryophytas, já que está presente no musgo *Physcomitrella patens*. No entanto não podemos descartar o fato de que estes genes tenham sido perdidos em algas e que, na verdade, não se trate de um processo de transferência entre bactéria e planta, mas sim que estes genes tenham um ancestral comum entre todos os seres-vivos e tenham sido perdidos nas demais linhagens de eucariotos e hoje existem apenas em Embryophytas e bactérias.

Como perspectivas deste trabalho estão a racionalização dos estudos funcionais ao longo das hidrolases ativas sobre xiloglucano que podem, em longo prazo, contribuir com processos biotecnológicos que passam pela modificação da parede celular vegetal.

## Referências

1. Adams, M.D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M, Utterback, T.R., Nagle, J. W., Fields, C., Venter, J.C. (1991) Sequence identification of 2.375 man brain genes. *Nature* 355 (6361); 632-634.
2. Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377 (6547 Suppl):3-174.
3. Ahn YO, Zheng M, Bevan DR, Esen A, Shiu SH, Benson J, Peng HP, Miller JT, Cheng CL, Poulton JE, Shih MC. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35. *Phytochemistry*. 2007 Jun;68(11):1510-20. Epub 2007 Apr 26.
4. Aidar, M.P.M.; Martinez, C.A.; Costa, A.C.; Costa, P.M.F.; Dietrich, S.M.C. & Buckeridge, M.S. 2002. Effect of atmospheric CO<sub>2</sub> enrichment on the establishment of seedlings of jatobá, *Hymenaea courbaril* L. (Leguminosae, Caesalpinioideae). *Biota Neotropica* 2(1)
5. Albert M, Werner M, Proksch P, Fry SC, Kaldenhoff R (2004) The cell wall-modifying xyloglucan endotransglycosylase/hydrolase LeXTH1 is expressed during the defence reaction of tomato against the plant parasite *Cuscuta reflexa*. *Plant Biol* 6(4):402-7.
6. Altschul S.F., Lipman D.J. 1990. Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A*. 87:5509-13.
7. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389-402.
8. Alvarez-Buylla, E.R.; Pelaz S., Liljegren S.J., Gold S.E., Burgeff, C., Ditta G.S., Ribas de Pouplana L., Martinez-Castilla L., Yanofsky M.F. (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *PNAS* 97(10):5328-5333.
9. Andrews, J., Bouffard, G.G., Cheadle, C., Lü, J. Becker, K.G., Oliver, B. (2000). Gene discovery using computational and microarray analysis of transcription in *Drosophila melanogaster*. *Tetis. Genome Research*. 6(9): 791-806.
10. Appel AM, Newell R, DuBois DL, Rakowski DuBois M. Concentration of carbon dioxide by electrochemically modulated complexation with a binuclear copper complex. *Inorg Chem*. 2005 May 2;44(9):3046-56.
11. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000 Dec 14;408(6814):796-815.
12. Balsler, C. 1960. Notes on the resin in aboriginal Central America. *Akt. 34. Int. Amerik. Kong. Ein.* p. 374-380.
13. Baumann J. M., Jens M. Eklöf, Gurvan Michel, Åsa M. Kallas, Tuula T. Teeri, Mirjam Czjzek, and Harry Brumer, III. Structural Evidence for the Evolution of Xyloglucanase Activity from Xyloglucan Endo-Transglycosylases: Biological Implications for Cell Wall Metabolism. *Plant Cell*. 2007 June; 19(6): 1947–1963.
14. Bennetzen, J. (2002). Opening the door to comparative plant biology. *Science* (296):60-63.

15. Bernard P, Couturier M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol.* 1992 Aug 5;226(3):735-45.
16. Bläsing O.E., Gibon Y., Günther M., Höhne M., Morcuende R., Osuna D., Thimm O., Usadel B., Scheible W.R., Stitt M. 2005. Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell.* 17: 3257-3281.
17. Boisselier-Dubayle MC, Lambourdière J, Bischler H. 2002 . Molecular phylogenies support multiple morphological reductions in the liverwort subclass Marchantiidae (Bryophyta). *Mol Phylogenet Evol.*24:66-77.
18. Brown DD, Wensink PC, Jordan E. A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J Mol Biol.* 1972 Jan 14;63(1):57-73.
19. BUCKERIDGE, M.S.; CROMBIE, H.J.; MENDES, C.J.M.; REID, J.S.G.; GIDLEY, M.J. & VIEIRA, C.C.J. A new family of oligosaccharides from the xyloglucan of *Hymenaea courbaril* L. (Leguminosae) cotyledons. *Carbohydrate Research*, 303:233-237, 1997.
20. Cabib E, Blanco N, Grau C, Rodríguez-Peña JM, Arroyo J. Crh1p and Crh2p are required for the cross-linking of chitin to beta(1-6)glucan in the *Saccharomyces cerevisiae* cell wall. *Mol Microbiol.* 2007 Feb;63(3):921-35.
21. CAVALLI-SFORZA, L.L. & Edwards (1967). Phylogenetic analysis: models and estimation procedures. *Am. J. Hum. Genet.* (19): 233-257.
22. Carpita NC, Gibeaut DM. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J*, 1993 - Blackwell Synergy
23. Carpita, N.C. 1996. Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Physiol. Plant Molec. Biol.*47, 445-476
24. Carpita N, McCann M (2000) The cell wall. In *Biochemistry and Molecular Biology of Plants*. Edited by Buchanan BB, Gruissem W, Jones RL. pp. 52-108. American Society of Plant Physiologists, Rockville, Maryland.
25. Chase MW, Fay MF. Ancient flowering plants: DNA sequences and angiosperm classification.
26. *Genome Biol.* 2001;2(4):REVIEWS1012. Epub 2001 Mar 22. Review.
27. Chen, Xiao-hong, Chen Zhi, Yao Hang-ping, Chen Feng, Zhu, Hai-hong, Zhou, Hong-juan. 2005. Construction and characterization of a cDNA library from human liver tissue with chronic hepatitis B. *Journal of Zhejiang University SCIENCE.* 2005 6B(4):288-294
28. Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP. 2006. The TIGR Plant Transcript Assemblies database. *Nucleic Acids Res.* 2006 Nov 6
29. Clark P. U., McCabe A. M., Mix A. C., Weaver A. J. Rapid Rise of Sea Level 19,000 Years Ago and Its Global Implications. *Science* 21 May 2004;Vol. 304. no. 5674, pp. 1141 – 1144
30. Cosgrove DJ (1997) Relaxation in a high-stress environment: the molecular bases of extensible cell walls and cell enlargement. *Plant Cell* 9: 1031-1041

31. Cox, P. M., R.<sup>a</sup> Betts, C.D. Jones, S.A. Spall & I. J. Totterdell, 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408: 184-187
32. Crombie HJ, Chengappa S, Hellyer A, Reid JS. A xyloglucan oligosaccharide-active, transglycosylating beta-D-glucosidase from the cotyledons of nasturtium (*Tropaeolum majus* L) seedlings--purification, properties and characterization of a cDNA clone. *Plant J.* 1998 Jul;15(1):27-38.
33. Daohong W, Bochu W, Biao L, Chuanren D, Jin Z. Extraction of total RNA from *Chrysanthemum* containing high levels of phenolic and carbohydrates. (2004). *Coll. Surf. B: Bio* 36: 111-114.
34. DAYHOFF, M.O., Schwartz, R.C.; Orcutt, B.C. (1978). A model of evolutionary change in proteins. In: *Atlas of protein sequence and structure* (M.O. Dayhoff, ed.), pp. 301-310. National Biomedical Research Foundation, Silver Spring, MD.
35. de la Torre F., Sampedro J., Zarra I, Revilla G. AtFXG1, an Arabidopsis Gene Encoding -L-Fucosidase Active against Fucosylated Xyloglucan Oligosaccharides. *Plant Physiol*, January 2002, Vol. 128, pp. 247-255.
36. Derelle E., Ferraz C., Rombauts S., Rouzé P., Worden A.Z., Robbens S., Partensky F., Degroeve S., Echeynié S., Cooke R., Saeys Y., Wuyts J., Jabbari K., Bowler C., Panaud O., Piégu B., Ball SG, Ral J.P., Bouget F.Y., Piganeau G., De Baets B, Picard A., Delseny M., Demaille J., Van de Peer Y., Moreau H. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci U S A.* 103:11647-52.
37. Dopazo, J. Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. *Journal of Molecular Evolution.* Volume 38, Number 3 / March, 1994. 300-304.
38. ECK, R.V. & Dayhoff, M.O. (1967). *Atlas of protein sequence and structures.* National Biomedical Research Foudation, Silver Springs, MD.
39. Edwards AWF, Cavalli-Sforza LL. *The reconstruction of evolution - Heredity*, 1963.
40. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading- *Syst. Zool*, 1978 – JSTOR
41. FELSENSTEIN, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* (17): 368-376.
42. FELSENSTEIN, J. *Confidence limits on phylogenies: an approach using the bootstrap - Evolution*, 1985.
43. Field C.B., Behrenfeld M. J., Randerson J.T., Falkowski P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* 10 July 1998: Vol. 281. no. 5374, pp. 237 – 240
44. FITCH, W.M. (2000). Homology: a personal view on some of the problems. *TIG* 16(5): 227-231.
45. Fry, S. C. In-vivo formation of xyloglucan nonasaccharide: A possible biologically active cell-wall fragment (1986) *Planta* 169, 443-453



46. Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J.-P., Kato, Y., Lorences, E. P., Maclachlan, G. A., McNeil, M., Mort, A. J., Reid, J. S. G., Seitz, H. U., Selvendran, R. R., Voragen, A. G. J., and White, A. R. (1993) An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiol. Plant.* 89: 1-3.
47. Gesteira A.S., Micheli F., Ferreira C.F. and Cascardo C.M. 2003. Isolation and purification of functional total RNA from different organs of cacao tree during its interaction with the pathogen *Crinipellis perniciososa*. *BioTech.* 35: 494-500.
48. GOFF, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Prus, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Felhaus, J., Malcama, T., Oliphant, A. & Briggs, S. (2002). A draft of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 296: 92-100.
49. Gómez-Esquer F, Rodríguez-Peña JM, Díaz G, Rodríguez E, Briza P, Nombela C, Arroyo J. CRR1, a gene encoding a putative transglycosidase, is required for proper spore wall assembly in *Saccharomyces cerevisiae*. *Microbiology*. 2004 Oct;150(Pt 10):3269-80.
50. Hayashi T (1989) Xyloglucans in the primary cell wall. *Annu Rev Plant Physiol Plant Mol Biol* 40: 139-168.
51. Hedges SB, Blair JE, Venturi ML, Shoe JL. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol*. 2004 Jan 28;4:2.
52. Henderson IM, Hendy MD, Penny D. Influenza viruses, comets and the science of evolutionary trees. *J Theor Biol*. 1989 Oct 9;140(3):289-303.
53. HENIKOFF, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K., Hood, L. (1997). Gene families: the taxonomy of protein paralogs and chimeras. *Science* (278): 609-614.
54. HENNING, W. (1966). *Phylogenetic systematics*. University of Illinois Press, Urbana.
55. Herendeen, P. S., A. Bruneau, G. P. Lewis. 2003. Phylogenetic relationships in caesalpinoid legumes: a preliminary analysis based on morphological and molecular data. Pages 37-62 in *Advances in Legume Systematics, part 10, higher level systematics* (B. B. Klitgaard and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.
56. Hood L., Campbell J.H., Elgin S.C. 1975. The organization, expression, and evolution of antibody genes and other multigene families. *Annu Rev Genet.* 9:305-53. Review.
57. Huang X., Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9:868-77.
58. Hughes A.L., Nei M. 1990. Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. *Mol Biol Evol.* 7:491-514.
59. Hwang, D. M., Dempsey, A. A., Wang, R. X., Rezvani, M., Barrans, J. D., Dai, K. S., Wang, H. Y., Ma, H., Cukerman, E., Liu, Y. Q., Gu, J.R., Zhang, J. H. Tsui, S. K., Wayne, M. M., Fung, K. P., Lee, C. Y., Liew, C. C. (1997). A genome-based resource for molecular cardiovascular medicine toward a compendium of cardiovascular genes. *Circulation* 96 (12):4146-4203.

60. Iglesias N., Abelenda J. A., Rodiño M., Sampedro J., Revilla G., Zarra I. Apoplastic Glycosidases Active Against Xyloglucan Oligosaccharides of *Arabidopsis thaliana*. *Plant and Cell Physiology* 2006 47(1):55-63.
61. Ingram V.M. 1961. Gene evolution and the haemoglobins. *Nature*. 189:704-8.
62. Ishimizu T, Hashimoto C, Takeda R, Fujii K and Hase S. A Novel 1,2-L-Fucosidase Acting on Xyloglucan Oligosaccharides is Associated with Endo- $\beta$ -Mannosidase. *Journal of Biochemistry* 2007 142(6):721-729.
63. Jaillon O., Aury J.M., Noel B., Policriti A., Clepet C., Casagrande A. Choisne N., Aubourg S., Vitulo N., Jubin C., Vezzi A., Legeai F., Huguency P., Dasilva C., Horner D., Mica E., Jublot D., Poulain J., Bruyère C., Billault A., Segurens B., Gouyvenoux M. Ugarte E., Cattonaro F., Anthouard V., Vico V., Del Fabbro C., Alaux M., Di Gaspero G., Dumas V., Felice N., Paillard S., Juman I., Moroldo M., Scalabrin S., Canaguier A., Le Clainche I., Malacrida G., Durand E., Pesole G., Laucou V., Chatelet P., Merdinoglu D., Delledonne M., Pezzotti M., Lecharny A., Scarpelli C., Artiguenave F., Pè M.E., Valle G., Morgante M., Caboche M., Adam-Blondon A.F., Weissenbach J., Quétier F., Wincker P. French-Italian Public Consortium for Grapevine Genome Characterization. 2007.
64. Kallas AM, Piens K, Denman SE, Henriksson H, Faldt J, Johansson P, Brumer H, Teeri TT (2005) Enzymatic properties of native and deglycosylated hybrid aspen (*Populus tremulaxtremuloides*) xyloglucan endotransglycosylase 16A expressed in *Pichia pastoris*. *Biochem J*. 390 (Pt 1): 105-13.
65. Karol KG, RM McCourt, MT Cimino, CF Delwiche. The Closest Living Relatives of Land Plants. - *Science*, 2001 - sciencemag.org - *Science*, Vol 294, Issue 5550, 2351-2353
66. Kellis, M.; Birren, B. W.; Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* (428):617-624.
67. Kenzelmann M, Mühlemann K. Substantially enhanced cloning efficiency of SAGE (Serial Analysis of Gene Expression) by adding a heating step to the original protocol. *Nucleic Acids Res*. 27:917-8.
68. Kidd KK, Sgaramella-Zonta LA. Phylogenetic analysis: concepts and methods. *Am J Hum Genet*. 1971 May;23(3):235-52.
69. Kim CB, Moon SY, Gelder SR, Kim W. Phylogenetic relationships of annelids, molluscs, and arthropods evidenced from molecules and morphology. *J Mol Evol*. 1996 Sep;43(3):207-15.
70. Kishino H., T Miyata, M Hasegawa. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts - *Journal of Molecular Evolution*, 1990 – Springer
71. Langenheim, J. H. Botanical source of Amber from Chipas, Mexico. *Ciencia, Mex*. Volume:24, 1966.
72. Lawton-Rauh, A. (2003) Evolutionary dynamics of duplicated genes in plants. *Mol Phylogenet Evol*. 29(3):396-409.
73. LE COINTE, P. Árvores e plantas úteis (indígenas e aclimadas). São Paulo, Companhia Editora Nacional, 1947. 505p.

74. Lee EJ, Matsumura Y, Soga K, Hoson T, Koizumi N. Glycosyl hydrolases of cell wall are induced by sugar starvation in Arabidopsis. *Plant Cell Physiol.* 2007 Mar;48(3):405-13. Epub 2007 Jan 18.
75. Lee. Y. T.; Langenheim, J. H. 1975. Systematics of the genus *Hymenaea* L. (Leguminosae - Caesalpinioideae) Detarieae University of California Press - London, p.77-79.
76. Lespinet O, Wolf YI, Koonin EV, Aravind L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 12(17):1048-1059.
77. Lewis, S.L., Phillips, O.L., Sheil, D., Vinceti, B., Baker, T.R., Brown, S., Graham, A.W., Higuchi, N., Hilbert, D.W., Laurance, W.F., Lejoly, J., Malhi, Y., Monteagudo, A., Vargas, P.N., Sonké, B., Supardi, N., M.N., Terborgh, J.W., Martínez, R.V. 2004. Tropical forest tree mortality, recruitment and turnover rates: calculation, interpretation and comparison when census intervals vary. *Journal of Ecology* 92: 929-944.
78. Liew, C. C., Hwang, D. M., Fung, Y.W., Laurensen, C., Cukerman, E., Tsui, S. and Lee, C.Y. (1994). A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc. Natl. Acad. Sci. USA.* 91 (22):10645-10649.
79. Liu, Y.B., Lu S. M., Zhang J. F., Liu S., Lu Y.T. A xyloglucan endotransglucosylase/hydrolase involves in growth of primary root and alters the deposition of cellulose in Arabidopsis. *Planta.* Springer Berlin / Heidelberg. Volume 226, Number 6 / November, 2007. 1547-1560.
80. Lima, P. F., 1950. A canoa de casca de jatobá entre os índios do Xingu. *Revista, São Paulo, Brasil Museu Paulista.* N. S. 4:369-380.
81. Linch M. (2002). Genomics. Gene duplication and evolution. *Science* 297(5583):945-947.
82. Loureiro, A. A.; Silva, M. F.; Alencar, J. C. 1979. Essências madeireiras da Amazônia. Vol. 1. Instituto Nacional de Pesquisas da Amazônia. Manaus, Amazonas p.193-196.
83. Mackinder, B. 2005. Detarieae. Pages 69-109 *Legumes of the world* (Lewis, G., B. Schrire, B. MacKinder, and M. Lock, eds.). The Royal Botanical Gardens, Kew, UK.
84. MARTINEZ-CASTILLA LP, Alvarez-Buylla ER.(2003). Adaptive evolution in the Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny. *PNAS* 100(23):13407-13412
85. McCann MC, Wells B, Roberts K (1990) Direct visualization of cross-links in the primary cell wall. *J Cell Sci* 96: 323-334 .
86. Melillo JM, McGuire AD, Kicklighter DW, Moore III B, Vorosmarty CJ, Schloss AL (1993). Global climate change and terrestrial net primary production. *Nature.* 363:234-240.
87. Merchant S.S., Prochnik S.E., Vallon O., Harris E.H., Karpowicz S.J., Witman G.B., Terry A., Salamov A., Fritz-Laylin L.K., Maréchal-Drouard L., Marshall W.F., Qu L.H., Nelson D.R., Sanderfoot A.A., Spalding M.H., Kapitonov V.V., Ren Q., Ferris P., Lindquist E., Shapiro H, Lucas S.M., Grimwood J., Schmutz J., Cardol P, Cerutti H, Chanfreau G., Chen C.L., Cognat V., Croft M.T., Dent R., Dutcher S., Fernández E., Fukuzawa H, González-Ballester D, González-Halphen D, Hallmann A, Hanikenne M, Hippler M., Inwood W., Jabbari K., Kalanon M., Kuras R., Lefebvre P.A., Lemaire S.D., Lobanov A.V, Lohr M., Manuell A., Meier I. Mets L., Mittag M., Mittelmeier T., Moroney J.V., Moseley J., Napoli C., Nedelcu A.M., Niyogi K, Novoselov S.V., Paulsen I.T., Pazour G., Purton S., Ral J.P., Riaño-Pachón D.M., Riekhof W., Rymarquis L., Schroda M., Stern D., Umen J., Willows R., Wilson N., Zimmer S.L., Allmer J., Balk J.,

- Bisova K., Chen C.J., Elias M., Gendler K., Hauser C., Lamb M.R., Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E., Terauchi A.M., Yang P., Ball S., Bowler C., Dieckmann C.L., Gladyshev V.N., Green P., Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martínez D, Ngau WC, Otilar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*. 318:245-50.
88. Meyerowitz, E.M. (2002). Plants compared to animals: the broadest comparative study of development. *Science* (295): 1482 – 1485.
  89. Muir G., Schlötterer C. 1999. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Mol Ecol*. 2005 14:549-61.
  90. Nam J, dePamphilis CW, Ma H, Nei M. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol Biol Evol*. 2003 Sep;20(9):1435-47. Epub 2003 May 30.
  91. NAM J, Kim J, Lee S, An G, Ma H, Nei M. (2004). Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *PNAS* 101(7):1910-1915.
  92. Nei M., F Tajima, Y Tateno. Accuracy of estimated phylogenetic trees from molecular data - *Journal of Molecular Evolution*, 1983 - Springer
  93. Nei M. *PHYLOGENETIC ANALYSIS IN MOLECULAR EVOLUTIONARY GENETICS* - *Annual Reviews in Genetics*, 1996 - Annual Reviews
  94. NEI, M., Kumar, S., Takahashi, K. (1998). The optimization principle in phylogenetic analysis tend to give incorrect topologies when the number of nucleotides or aminoacids is small. *PNAS* (95): 12390-12397.
  95. NEI, M. & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press.
  96. Nei M., Rooney A.P. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*. 39:121-52.
  97. Ohta T. Further study on the genetic correlation between members of a multigene family. *Genetics*. 1981 Nov-Dec;99(3-4):555-71.
  98. Ota T, Nei M. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol*. 11:469-82.
  99. Palenik B., Grimwood J., Aerts A., Rouzé P., Salamov A., Putnam N., Dupont C., Jorgensen R., Derelle E., Rombauts S., Zhou K., Otilar R., Merchant S.S., Podell S., Gaasterland T., Napoli C., Gendler K., Manuell A., Tai V., Vallon O., Piganeau G. Jancek S., Heijde M., Jabbari K., Bowler C., Lohr M., Robbens S., Werner G., Dubchak I., Pazour GJ., Ren Q., Paulsen I., Delwiche C., Schmutz J., Rokhsar D., Van de Peer Y., Moreau H., Grigoriev IV. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A*. 104:7705-10.
  100. Paterson AH. 2008. Genomics of sorghum. *Int J Plant Genomics*. 362451.

101. Pennachio, L.A (2003) Insights from human/mouse genome comparisons. *Mammalian Genome*. (14): 429-436.
102. Perteua G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 2003 Mar 22;19(5):651-2.
103. Prentice, I.C., Farquhar, G.D., Fasham, M.J.R., Goulden, M.L., Heimann, M., Jaramillo, V.J., Kheshgi, H.S., Le Quere, C., Scholes, R.J., Wallace, D.W.R., Archer, D., Ashmore, M.R., Aumont, O., Baker, D., Battle, M., Bender, M., Bopp, L.P., Bousquet, P., Caldeira, K., Ciais, P., Cox, P.M., Cramer, W., Dentener, F., Enting, I.G., Field, C.B., Friedlingstein, P., Holland, E.A., Houghton, R.A., House, J.I., Ishida, A., Jain, A.K., Janssens, I.A., Joos, F., Kaminski, T., Keeling, C.D., Keeling, R.F., Kicklighter, D.W., Hohfeld, K.E., Knorr, W., Law, R., Lenton, T., Lindsay, K., Maier-Reimer, E., Manning, A.C., Mearns, R.J., McGuire, A.D., Melillo, J.M., Meyer, R., Mund, M., Orr, J.C., Piper, S., Plattner, K., Rayner, P.J., Sitch, S., Slater, R., Taguchi, S., Tans, P.P., Tian, H.Q., Weirig, M.F., Whorf, T. and Yool, A. 2001. The carbon cycle and atmospheric carbon dioxide. Chapter 3 of the Third Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change 2001: The Scientific Basis*. Cambridge University Press, Cambridge, UK, pp. 183-238.
104. Raven, P. H.; Evert, R. F.; Eichhorn, S. E.; *Biology of Plants*, sixth edition. W.H. FREEMAN AND COMPANY. 1999.
105. Record, S. J. & Mell, C. D. 1924. *Timbers of tropical America*. New Haven, Yale University Press.
106. Rensing S.A., Lang D., Zimmer A.D., Terry A., Salamov A., Shapiro H., Nishiyama T., Perroud P.F., Lindquist E.A., Kamisugi Y., Tanahashi T., Sakakibara K., Fujita T., Oishi K., Shin-I. T., Kuroki Y., Toyoda A., Suzuki Y., Hashimoto S., Yamaguchi K., Sugano S., Kohara Y., Fujiyama A., Anterola A., Aoki S., Ashton N., Barbazuk W.B., Barker E., Bennetzen J.L., Blankenship R., Cho S.H., Dutcher S.K., Estelle M., Fawcett J.A., Gundlach H., Hanada K., Heyl A., Hicks K.A., Hughes J., Lohr M., Mayer K., Melkozernov A., Murata T., Nelson D.R., Pils B., Prigge M., Reiss B, Renner T., Rombauts S, Rushton PJ, Sanderfoot A., Schween G, Shiu S.H., Stueber K., Theodoulou F.L., Tu H., Van de Peer Y., Verrier P.J., Waters E., Wood A., Yang L., Cove D., Cuming A.C., Hasebe M., Lucas S., Mishler B.D., Reski R. Grigoriev I.V., Quatrano R.S., Boore J.L. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*. 319:64-9.
107. Rose JKC, Braam J, Fry SC, Nishitani K (2002). The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. *Plant and Cell Physiol*. 43 (12): 1421-1435.
108. Royer, D.L., Berner, R.A., Beerling, D.J., 2001a, Phanerozoic CO<sub>2</sub> change: Evaluating geochemical and paleobiological approaches: *Earth-Science Reviews*, 54, p. 349–392.
109. Ruddiman, W.F. *The early anthropogenic hypothesis a year later*. Climate Change, 2005. Springer.
110. Rzhetsky A, Nei M. Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *J Mol Evol*. 1992 Oct;35(4):367-75.
111. Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* (24): 184-204.

112. Sampedro J., Sieiro C., Revilla G., González-Villa T., Zarra I. Cloning and Expression Pattern of a Gene Encoding an  $\alpha$ -Xylosidase Active against Xyloglucan Oligosaccharides from *Arabidopsis*. *Plant Physiol*, June 2001, Vol. 126, pp. 910-920.
113. Sandwith, N.Y., 1938. New species and records from British Guiana: XV. Contributions to the flora of tropical America: XXX. *Bull. Misc. Inf. Roy. Bot. Gard. Kew* 1931: 366-367.
114. Simpson, B. B., L. L. Larkin, and A. Weeks. 2003. Progress towards resolving the relationships of the *Caesalpinia* group (Caesalpinieae: Caesalpinioideae: Leguminosae). Pages 123-148 in *Advances in Legume Systematics*, part 10, higher level systematics (B. B. Klitgaard and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.
115. Sneath PHA, Sokal RR. Numerical taxonomy. The principles and practice of numerical classification. - 1973 - A Series of Books in Biology.
116. Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci - Bull*, 1958
117. Takami H, Nakasone Y, Takaki G, Maeno R, Sasaki R. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis* - *Nucleic Acids Research*, 2000. Vol. 28, No. 21 4317-4331 - Oxford Univ Press.
118. Takezaki N., Nei M. Genetic Distances and Reconstruction of Phylogenetic Trees From Microsatellite DNA - *Genetics*, 1996 - Genetics Soc America.
119. Takezaki N, Nei M. Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant. *J Mol Evol*. 1994 Aug;39(2):210-8.
120. Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596-1599.
121. Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Münster T, Theissen G, Nozaki H, Ito M. Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. *Proc Natl Acad Sci U S A*. 2005 Feb 15;102(7):2436-41. Epub 2005 Feb 7.
122. Tatusov, R.L., Koonin, E.V. & Lipman, D.J. (1997). A genomic perspective on protein families. *Science*, 278: 631-637.
123. Tatusova TA, Karsch-Mizrachi I, Ostell JA. Complete genomes in WWW Entrez: data representation and analysis. *Bioinformatics*. 1999 Jul-Aug;15(7-8):536-43.
124. Theissen G. 2001. Genetics of identity. *Nature*. 29;414(6863):491.
125. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994 Nov 11;22(22):4673-80.
126. Thornton, J. & DeSalle, R. (2000) Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genomics Hum. Genet.*, 1:41-73.
127. Tibbot BK, Skadsen RW. Molecular cloning and characterization of a gibberellin-inducible, putative  $\alpha$ -glucosidase gene from barley. *Plant Mol Biol*. 1996 Jan;30(2):229-41.

128. TINÉ, M.A.S., CORTELAZZO, A.L. & BUCKERIDGE, M.S. (2000) Occurrence of xyloglucan containing protuberances in the storage cell walls of cotyledons of *Hymenaea courbaril* L. *Revista Brasileira de Botanica* 23:413-417.
129. TINÉ, M.A.S., SILVA, C.O., LIMA D.U., CARPITA, N.C. & BUCKERIDGE, M.S. (2006) Fine structure of a mixed-oligomer storage xyloglucan from seeds of *Hymenaea courbaril* Carbohydrate Polymers.
130. Tuskan G.A., Difazio S., Jansson S., Bohlmann J., Grigoriev I. Hellsten U., Putnam N., Ralph ., Rombauts S., Salamov A., Schein J, Sterck L., Aerts A., Bhalerao R.R., Bhalerao R.P., Blaudez D., Boerjan W., Brun A., Brunner A., Busov V., Campbell M., Carlson J., Chalot M., Chapman J., Chen G.L., Cooper D., Coutinho P.M., Couturier J., Covert S., Cronk Q., Cunningham R., Davis J., Degroeve S., Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehrling J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryabov D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 313:1596-604.
131. Tschirch, A. & Stock, E., 1936. *Die Hazar*. 3rd ed. Berlin.
132. Van Sandt VS, Suslov D, Verbelen JP, Vissenberg K. Xyloglucan endotransglucosylase activity loosens a plant cell wall. *Ann Bot (Lond)*. 2007 Dec;100(7):1467-73. Epub 2007 Oct 4.
133. Vandepoele, K., Simillion C., Van de Peer, Y. (2003). Evidence that rice and other cereals are ancient aneuploids. *Plant Cell*. 15(9):2192-2202.
134. VINCENTZ, M., Bandeira-Kobarg, C., Gauer, L., Schlögl, P. & Leite, A. (2003). Evolutionary pattern of Angioperm bZIP factors homologous to the maize Opaque2 regulatory protein. *Journal of Molecular Evolution* 55: 1–12.
135. Vincentz M., Cara F.A., Okura V.K., da Silva F.R., Pedrosa G.L., Hemerly A.S., Capella A.N., Marins M., Ferreira P.C., Franca S.C., Grivet L., Vettore A.L., Kemper E.L., Burnquist W.L., Targon M.L., Siqueira W.J., Kuramae E.E., Marino C.L., Camargo L.E., Carrer H., Coutinho L.L., Furlan L.R., Lemos M.V., Nunes L.R., Gomes S.L., Santelli R.V., Goldman M.H., Bacci M. Jr., Giglioti E.A., Thiemann O.H., Silva F.H., Van Sluys M.A., Nobrega F.G., Arruda P., Menck C.F. (2004). Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. *Plant Physiol*. 134(3): 951-959.
136. Yokoyama R, Rose JKC, Nishitani K (2004) A surprising diversity and abundance of xyloglucan endotransglucosylase/hydrolases in rice. Classification and expression analysis. *Plant Physiol*. 134(3):1088-99.
137. York WS, van Halbeek H, Darvill AG, Albersheim P (1990) Structure of plant cell walls: XXIX. Structural analysis of xyloglucan oligosaccharides by <sup>1</sup>H-NMR spectroscopy and fast atom bombardment mass spectrometry. *Carbohydr Res* 200: 9-31.
138. Weigel D., Meyerowitz E.M. 1994. The ABCs of floral homeotic genes. *Cell*. 78:203-209

139. Wendel J. F. (2000). Genome evolution in polyploids. *Plant Mol Biol* 42(1):225-249.
140. Wieringa, J. J., and G. Y. F. Gervais. 2003. Phylogenetic analysis of combined morphological and molecular data sets on the *Aphanocalyx-Bikinia-Tetraberlinia* group (Leguminosae, Caesalpinioideae, Detarieae s.l.). Pages 181-196 in *Advances in Legume Systematics, part 10, higher level systematics* (B. B. Klitgaard and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.
141. Zeleny R., Leonard R., Dorfner G., Dalik T., Kolarich D. and Altmann F. Molecular cloning and characterization of a plant 1,3/4-fucosidase based on sequence tags from almond fucosidase I. *Phytochemistry*. Volume 67, Issue 7, April 2006, Pages 641-648.
142. Zhang, Z., Schwartz, S, Wagner, L., and Miller, W. (2000), "A greedy algorithm for aligning DNA sequences", *J Comput Biol* 2000; 7(1-2):203-14.
143. Zhang D, Fujio K, Jiang Y, Zhao J, Tada N, Sudo K, Tsurui H, Nakamura K, Yamamoto K, Nishimura H, Shira T, Hirose S. Dissection of the role of MHC class II A and E genes in autoimmune susceptibility in murine lupus models with intragenic recombination. *Proc Natl Acad Sci U S A*. 2004 Sep 21;101(38):13838-43.
144. Zharkikh A, Li WH. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Mol Biol Evol*. 1992 Nov;9(6):1119-1147.



**ANEXO I** – Código fonte, em linguagem JAVA®, do programa BTF (*Blast to Fasta*; Del Bem & de Mello, 2008 – não publicado).

```
package gfmello.basttofasta;

import java.io.BufferedReader;

import java.io.BufferedWriter;

import java.io.File;

import java.io.FileInputStream;

import java.io.FileOutputStream;

import java.io.InputStreamReader;

import java.io.OutputStreamWriter;

import java.util.ArrayList;

import java.util.Collection;

import java.util.HashMap;

import java.util.Iterator;

import java.util.Map;

/**

 * Classe para ler sequencias de um resultado de blast local e localiza-las em um banco

de sequencias.

 * @author

 *

 */

public class BTF {
```

```

private File blastFileName;

private File fastaFileName;

private String newFileName;

private static final String STR_START_SEQUENCE = "Sequences producing significant
alignments";

private Collection listaNomesSequencias;

private Map mapNomesSequencias; //codigo do nome das sequencias

private Map mapNomesSequenciasCompleto; //nome da sequencia conforme esta no
arquivo do fasta

private static final String FASTA_MARCADOR = ">";

/**
 * @param args
 */

public static void main(String[] args) {

    try {

        if (!argumentosValidos(args)) {

            showHelp();

        }

        else {

            BTF btf = new BTF();

            btf.run(args);

        }

    }

    catch(Exception e) {

```

```

        System.out.println("Ocorreu um erro.");

        //e.printStackTrace();

    }

}

/**
 * Executa a funcao principal da aplicacao, coordenando a busca do blast,
 * no fasta, e geracao do arquivo final.
 *
 * @param args
 * @throws Exception
 */
private void run(String[] args) throws Exception {

    long tempoInicio = System.currentTimeMillis();

    //Inicializa nomes dos arquivos

    blastFileName = new File(args[0]);

    fastaFileName = new File(args[1]);

    newFileName = generateOutFileName();

    System.out.println("lendo blast...");

    //Carrega nomes de sequencias do blast

    loadBlastSequences();

    System.out.println("localizando no fasta...");

    //Carrega sequencias do fasta que estavam no blast

```

```

        findFastaSequences();

        System.out.println("gerando output..." + newFileName);

        //Gera o arquivo com resultados

        generateOutputFile();

        long tempoFim = System.currentTimeMillis();

        System.out.println("concluido em " + (tempoFim-tempoInicio) + " ms!");
    }

    /**
     * Gera nome do arquivo de resposta
     * @return
     */
    private String generateOutFileName() throws Exception {

        try {

            //Retira o .txt

            int index = blastFileName.getName().indexOf(".txt");

            String aux = blastFileName.getName().substring(0, index);

            String outFileName = "fasta_from_" + aux + ".fasta";

            return outFileName;

        }

        catch(Exception e) {

```

```

        System.out.println("Verifique os nomes dos arquivos! Nao estao em
ordem trocada?");

        throw e;

    }

}

/**
 * Gera o arquivo de resposta com as sequencias encontradas
 * @throws Exception
 */

private void generateOutputFile() throws Exception {

    if (listaNomesSequencias!= null && !listaNomesSequencias.isEmpty()) {

        FileOutputStream fout = new FileOutputStream(newFileName);

        OutputStreamWriter ow = new OutputStreamWriter(fout);

        BufferedWriter bw = new BufferedWriter(ow);

        String auxName;

        String auxCompleteName;

        String auxSequencia;

        for (Iterator it = listaNomesSequencias.iterator(); it.hasNext();
) {

            auxName = (String) it.next();

            auxCompleteName = (String)
mapNomesSequenciasCompleto.get(auxName);

            auxSequencia = (String) mapNomesSequencias.get(auxName);

            bw.write(auxCompleteName);

```

```

        bw.write("\n");

        bw.write(auxSequencia);

    }

    bw.close();

    ow.close();

    fout.close();

}

else {

    throw new Exception("Nenhuma sequencia encontrada para gerar o
arquivo de saida.");

}

}

/**
 * Carrega sequencias do fasta, cujos nomes estavam no blast
 * @throws Exception
 */

private void findFastaSequences() throws Exception {

    FileInputStream fin = new FileInputStream(fastaFileName);

    InputStreamReader sr = new InputStreamReader(fin);

    BufferedReader bf = new BufferedReader(sr);

    String linha;

    String auxSequenceName = null; //codigo da sequencia

```

```

        String auxCompleteSequenceName = null; //codigo da sequencia com nome
completo do fast

        StringBuffer auxSequence = null;

        boolean lendoSequencia = false;

        while ((linha = bf.readLine()) != null) {

            //Verifica se nova sequencia esta se iniciando

            if (linha.startsWith(FASTA_MARCADOR)) {

                if (lendoSequencia) {

                    //Se outra sequencia se inicia, vamos gravar o que
foi lido da anterior.

                    mapNomesSequencias.put(auxSequenceName,
auxSequence.toString());

                    mapNomesSequenciasCompleto.put(auxSequenceName,
auxCompleteSequenceName);

                }

                auxSequenceName = getFastaSequenceName(linha); //so o
codigo do nome da sequencia

                auxCompleteSequenceName = linha; //guarda o nome completo
da sequencia

                //Se a sequencia do fasta existe no blast, vamos carrega-la

                if (mapNomesSequencias.get(auxSequenceName) != null) {

                    lendoSequencia = true;

                    auxSequence = new StringBuffer(10000); //inicializa
sequencia lida

                }

                else {

```

```
lendoSequencia = false; //se a sequencia do fast
nao esta no blast, nao vamos carrega-la.
```

```
    }

    }

    else if (lendoSequencia) {

        auxSequence.append(linha).append("\n");

    }

}

/**
 * Identifica nome da sequencia do fasta
 * @param linha
 * @return
 */

private String getFastaSequenceName(String linha) {

    int indexEspaco;

    indexEspaco = linha.indexOf(" ");

    if (indexEspaco == -1) {

        indexEspaco = linha.length(); //Se nao tiver ' ', o nome da
sequencia é a linha toda

    }

    String nome = linha.substring(1,indexEspaco);

    return nome;

}
```



```

/**
 * Carrega nomes de sequencias que estao no arquivo blast
 *
 * @throws Exception
 */
private void loadBlastSequences() throws Exception {

    //Abre o arquivo do blast para leitura

    FileInputStream fin = new FileInputStream(blastFileName);

    InputStreamReader sr = new InputStreamReader(fin);

    BufferedReader bf = new BufferedReader(sr);

    String linha;

    int indiceFimNome;

    String auxNomeSequencia;

    listaNomesSequencias = new ArrayList(1000);

    mapNomesSequencias = new HashMap(1000);

    mapNomesSequenciasCompleto = new HashMap(1000);

    while ((linha = bf.readLine()) != null) {

        if (isInicioSequencia(linha)) {

            //Inicio de sequencia a ser analisada

            bf.readLine(); //pula uma linha

            //Le as sequencias ate encontrar linha em branco indicando
o fim da sequencia

            while ((linha = bf.readLine()) != null) {

                if (linha.trim().equalsIgnoreCase("")) {

```

```

//Se encontrar uma linha em branco indica o
fim da sequencia de nomes

break;

}

indiceFimNome = linha.indexOf(" "); //Procura
'espaco' indicando fim do nome da sequencia

if (indiceFimNome == -1) throw new Exception("Nao
foi possivel extrair o nome da sequencia!\n[" + linha + "]");

auxNomeSequencia = linha.substring(0,
indiceFimNome);

if (mapNomesSequencias.get(auxNomeSequencia) ==
null) {

mapNomesSequencias.put(auxNomeSequencia,
""); //guarda no map como nome ja localizado

listaNomesSequencias.add(auxNomeSequencia);

//adiciona na lista para manter ordem q foi encontrado

}

} //end While

} //end if

} //end while

}

/**
 * Verifica se a String informada indica inicio de sequencia
 * @param linha
 * @return

```

```

*/

private boolean isInicioSequencia(String linha) {

    boolean encontrou = false;

    if (linha.startsWith(STR_START_SEQUENCE)) {

        encontrou = true;

    }

    return encontrou;

}

/**
 * Verifica se os argumentos passados sao validos
 *
 * @param args
 *
 * @return
 */

private static boolean argumentosValidos(String[] args) {

    boolean validos = false;

    if (args != null && args.length == 2) {

        try {

            File f1 = new File(args[0]);

            File f2 = new File(args[1]);

            if (f1.exists() && f2.exists()) {

                validos = true;

            }

        }

        else {


```

```

        System.out.println("Arquivo nao encontrado.[" +
(f1.exists()? f2.getName() : f1.getName()) + "]);

    }

}

catch(Exception e){

    System.out.println("Nome de arquivo invalido.");

    e.printStackTrace();

}

}

return validos;

}

/**
 * Exibe ajuda sobre utilizacao do programa
 *
 */

private static void showHelp() {

    System.out.println("\n\nUso: java -jar BTF.jar blastFile fastaDatabase");

    System.out.println("\n ou: BTF.jar blastFile fastaDatabase (Se a
'Java(TM) Platform' esta configurada para abrir arquivos .jar");

    System.out.println("\nO novo arquivo sera gerado no diretorio atual.");

}

}

```

**ANEXO II** – Código fonte, em linguagem JAVA®, do programa ETTool (*ESTs Translator Tool*; Del Bem & de Mello, 2008 – não publicado)

```
package com.gmail.at.gfmello;

import java.io.BufferedReader;

import java.io.BufferedWriter;

import java.io.File;

import java.io.FileInputStream;

import java.io.FileOutputStream;

import java.io.InputStream;

import java.io.InputStreamReader;

import java.io.OutputStreamWriter;

import java.util.ArrayList;

import java.util.HashMap;

import java.util.LinkedHashMap;

import java.util.List;

import java.util.Map;

/**

 *

 * Classe para ler sequencias de um resultado de blast local (ESTs, sequencias

nucleotídicas), localiza-las em seu respectivo banco de sequencias,

 *

 *          traduzi-las (testando as traduções contra bancos de proteínas) e

entregar os ESTs traduzidos em arquivo FASTA.

 *

 */
```

```

*/

public class ETTool {

    /**
     * Classe que representa uma sequencia, ex, DT576904+1
     GLCPPGNVSPRLSSHMV*FSYKTD...etc
     * @author
     * email:
     *
     */

    private class Sequencia {

        private String nomeSequencia;

        private char[] letrasSequencia;

        /**
         * Inicializa o nome e letras de uma sequencia
         * @param nome
         * @param sequencia
         */

        public Sequencia(String nome, char[] sequencia) {

            nomeSequencia = nome;

            letrasSequencia = sequencia;

        }

        public char[] getLetrasSequencia() {

```

```

        return letrasSequencia;
    }

    public String getNomeSequencia() {

        return nomeSequencia;
    }
}

//Arquivos Lidos

private File blastFileName; //Nome do Arquivo do blast que vamos carregar os
nomes das sequencias

private File fastaFileName; //Nome do Arquivo fasta com as sequencias de
aminoacidos

//Arquivos Escritos

private String nomeOutputBlast = "_outputBlast-" + System.currentTimeMillis() +
".txt"; //Nome do Arquivo que o Blast vai gerar apos ser executado

private String nomeTmpInputBlast = "_inputBlast.txt"; //Nome do Arquivo
temporario utilizado para alimentar o Blast

private String nomeOutputFile;

//Variaveis

private String blastPath; //Caminho da pasta onde o Blast esta instalado

private String nomeBancoSeqProteinas; //Nome do banco de proteinas a ser usado
qdo executar o Blast

private static final String STR_START_SEQUENCE = "Sequences producing significant
alignments";

```

```

private static final String STR_START_QUERY_SEQUENCE_NAME = "Query= ";

private static final String FASTA_MARCADOR = ">";

private static final String BLAST_EXE = "blastall.exe";

private int tabelaAminoacidos;

/**
 * Porta de entrada da aplicacao
 * @param args
 */
public static void main(String[] args) {

    try {

        //Valida os argumentos passados na linha de comando

        if (!argumentosValidos(args)) {

            //Exibe ajuda caso comandos estejam errados

            showHelp();

        }

        else {

            //Executa a aplicacao pois os parametros da linha de
comando sao validos

            ETTool e = new ETTool();

            //Inicializa parametros da aplicacao

            e.loadParameters(args);

            e.run();

        }
    }
}

```



```

    }

    catch(Exception e) {

        log("Ocorreu um erro.");

        e.printStackTrace();

    }

}

/**

 * Armazena em variaveis os parametros passados pelo usuario ao executar a
aplicacao

 * @param args

 * @throws Exception

 */

private void loadParameters(String[] args) throws Exception {

    blastFileName = new File(args[0]);

    fastaFileName = new File(args[1]);

    blastPath = args[2];

    nomeBancoSeqProteinas = args[3];

    nomeOutputFile = args[4];

    tabelaAminoacidos = Integer.parseInt(args[5]);

}

/**

 * Executa a funcao principal da aplicacao, coordenando a busca do blast,

 * no fasta, e geracao do arquivo final.

```

```

* @param args

* @throws Exception

*/

private void run() throws Exception {

    long tempoInicio = System.currentTimeMillis();

    //Map com nomes de sequencia, e sequencias como value do map

    Map<String, String> listaNomesSequencias;

    //Map com todas sequencias deslocadas que serviram de input ao blast

    Map<String, Sequencia> sequenciasInput;

    //Carrega apenas os nomes de sequencias do arquivo do blast

    listaNomesSequencias = getListaNomesSequenciasFromBlastFile();

    //Carrega os dados das sequencias do fast, cujo nome esta no map

    findFastaSequences(listaNomesSequencias);

    //Gera o arquivo de input do blast realizando os 6 deslocamentos de
leitura da sequencia

    sequenciasInput = traduzir(listaNomesSequencias);

    //Executa o blast para o arquivo de input temporario gerado.

    validacaoBlast();

    //Le o arquivo gerado pelo Blast

```

```

List<String> queryBlastEncontradas = extraiResultados();

//Cria um arquivo contendo somente as sequencias encontradas no blast

createOutputESTTool(queryBlastEncontradas, sequenciasInput);

long tempoFim = System.currentTimeMillis();

log("Concluído em " + ((tempoFim-tempoInicio)/1000) + " s!");
}

/**
 * Cria o arquivo final que contem somente as sequencias deslocadas que o blast
foi capaz de
 * encontrar um mapeamento.
 * @param lista
 * @param sequenciasInput
 */
private void createOutputESTTool(List<String> lista, Map<String, Sequencia>
sequenciasInput) throws Exception {
    if (lista != null) {
        log("Gerando arquivo de resposta em " + nomeOutputFile + " ...");

        //Cria o arquivo de saida

        FileOutputStream fout = new FileOutputStream(nomeOutputFile);

        OutputStreamWriter ow = new OutputStreamWriter(fout);

        BufferedWriter bw = new BufferedWriter(ow);

```

```

        for (String querySequence : lista) {

            Sequencia seq = sequenciasInput.get(querySequence);

            bw.write(">" + querySequence + "\n");

            bw.write(new String(seq.getLetrasSequencia()) + "\n");

        }

        //Fecha o arquivo

        bw.close();

        ow.close();

        fout.close();

        log("OK\n");

    }

}

/**
 * Executa o blast e aguarda sua finalizacao
 * @throws Exception
 */

private void validacaoBlast() throws Exception {

    StringBuilder sb = new StringBuilder();

    sb.append(BLAST_EXE)

        .append(" -p blastp -d ")

        .append(nomeBancoSeqProteinas)

        .append(" -i ")

```

```

        //.append(blastPath + File.separator + "bin" + File.separator +
nomeTmpInputBlast)

        .append(nomeTmpInputBlast)

        .append(" -o ")

        .append(nomeOutputBlast)

        .append(" -b 0 -v 5 -e 1e-10");

        String tmp = "cmd.exe /C cd " + blastPath + File.separator + "bin && " +
sb.toString();

        log("Executando Blast: " + tmp + "\nAguarde...");

        Process p = Runtime.getRuntime().exec(tmp);

        InputStream in = p.getInputStream();

        while(in.read() != -1) {} //Aguarda a execucao do blast

        //Remove o arquivo utilizado de input ao blast

        File tmpInBlast = new File(blastPath + File.separator + "bin" +
File.separator + nomeTmpInputBlast);

        //tmpInBlast.delete();

        log("OK\n");

    }

    /**
     * Le o resultado do blast procurando as sequencias com resultados.
     */

    private List<String> extraiResultados() throws Exception {

```

```

//Abre o arquivo de resultado blast para leitura

FileInputStream fin = new FileInputStream(blastPath

        + File.separator

        + "bin"

        + File.separator

        + nomeOutputBlast);

InputStreamReader sr = new InputStreamReader(fin);

BufferedReader bf = new BufferedReader(sr);

String linha;

List<String> listaSequenciasComResposta = new ArrayList<String>(1000);

String querySequence = null;

while ((linha = bf.readLine()) != null) {

        querySequence=getIfQuerySequenceName(linha,
querySequence);//Guarda o nome da sequencia que vamos analisar a resposta

        if (isInicioSequencia(linha)) {

                listaSequenciasComResposta.add(querySequence);

        } //end if

} //end while

return listaSequenciasComResposta;

}

/**

* Gera um arquivo temporario com os 6 deslocamentos de leitura da sequencia,

* e que sera utilizado de input para a execucao do blast.

```

```

    * @param listaNomesSequencias

    * @throws Exception

    */

    private Map<String, Sequencia> traduzir(Map<String, String> listaNomesSequencias)
throws Exception {

        log("Gerando input do Blast...");

        Map<String, Sequencia> result = new HashMap<String, Sequencia>();

        //Cria o arquivo temporario que vai ser o input do Blaster

        FileOutputStream fout = new FileOutputStream(blastPath + File.separator +
"bin" + File.separator + nomeTmpInputBlast);

        OutputStreamWriter ow = new OutputStreamWriter(fout);

        BufferedWriter bw = new BufferedWriter(ow);

        //Para cada sequencia do fasta

        for (String nomeSequencia : listaNomesSequencias.keySet()) {

            String sequencia = listaNomesSequencias.get(nomeSequencia);

            //Gera um array com os 6 deslocamentos de leitura da sequencia

            Sequencia[] sequencias = getAminoacidosDeslocados(nomeSequencia,
sequencia);

            for (int i = 0; i < 6; i++) {

                //Monta o arquivo com os 6 deslocamentos de leitura
possiveis

                bw.write(">" + sequencias[i].getNomeSequencia() + "\n");

```

```

        bw.write(new String(sequencias[i].getLetrasSequencia()) +
"\n");

        //Cria um map com todas sequencias gravadas no arquivo

        result.put(sequencias[i].getNomeSequencia(),
sequencias[i]);

    }

}

//Fecha o arquivo gerado

bw.close();

ow.close();

fout.close();

log("OK\n");//(Arquivo de input gerado em:" + blastPath + File.separator
+ "bin" + File.separator + nomeTmpInputBlast + ")\n");

return result;

}

/**
 * Gera um codon caso a sequencia possua 3 elementos, senao retorna null
 * @param seqArray
 * @param i
 * @return
 */

private String verificaCodon(char[] seqArray, int i) {

```



```

        String result = null;

        if (!(i+3>seqArray.length)) {

            result = new String(new char[]{ seqArray[i], seqArray[i+1],
seqArray[i+2] });

        }

        return result;

    }

    /**
     * Retorna uma lista dos aminoacidos encontrados com deslocamento +1,+2,+3, -1, -
2, -3
     * para a sequencia informada.
     * @param seq
     * @return
     * @throws Exception
     */

    private Sequencia[] getAminoacidosDeslocados(String nomeSeq, String seq) throws
Exception {

        Sequencia[] sequencias = new Sequencia[6];

        char[] seqArray = seq.toCharArray();

        sequencias[0] = new Sequencia( nomeSeq + " plus 1",
mapearAminoacidos(seqArray, 0));

        sequencias[1] = new Sequencia( nomeSeq + " plus 2",
mapearAminoacidos(seqArray, 1));

        sequencias[2] = new Sequencia( nomeSeq + " plus 3",
mapearAminoacidos(seqArray, 2));

```

```

        sequencias[3] = new Sequencia( nomeSeq + " minus 1",
mapearAminoacidos(inverte(seqArray), 0));

        sequencias[4] = new Sequencia( nomeSeq + " minus 2",
mapearAminoacidos(inverte(seqArray), 1));

        sequencias[5] = new Sequencia( nomeSeq + " minus 3",
mapearAminoacidos(inverte(seqArray), 2));

    return sequencias;

}

/**
 * Inverte o vetor a
 *
 * @param a
 *
 * @return
 */
private char[] inverte(char[] a) {

    char[] b = new char[a.length];

    for (int i = 0; i < a.length; i++) {

        b[b.length-i-1] = a[i];

    }

    return b;

}

/**
 * Mapeia para aminoacidos a sequencia fornecida e realizando o deslocamento
informado.

```

```

* @param seqArray

* @param shift

* @return

*/

private char[] mapearAminoacidos(char[] seqArray, int shift) {

    List<String> result = new ArrayList<String>();

    for(; shift < seqArray.length; shift+=3) {

        String codon = verificaCodon(seqArray, shift);

        if (codon != null) {

            String aminoName = ETToolConstants.getAminoacidoName(codon,
tabelaAminoacidos);

            result.add(aminoName);

        }

    }

    char[] seq_mais1 = new char[result.size()];

    int aux_i = 0;

    for(String amino : result ) {

        seq_mais1[aux_i++] = amino.toCharArray()[0];

    }

    return seq_mais1;

}

```

```

/**
 * Le o arquivo do blast e gera uma lista com os nomes das sequencias encontradas
 * @return Map de nomes de sequencias encontradas no arquivo do blast
 * @throws Exception
 */
private Map<String, String> getListaNomesSequenciasFromBlastFile() throws
Exception {

    log("Carregando Nomes de Sequencias do Blast...");

    //List<String> listaNomesSequencias = new ArrayList<String>(1000);
//Variavel de Retorno

    Map<String, String> listaNomesSequencias = new LinkedHashMap<String,
String>(1000);

    //Abre o arquivo do blast para leitura

    FileInputStream fin = new FileInputStream(blastFileName);

    InputStreamReader sr = new InputStreamReader(fin);

    BufferedReader bf = new BufferedReader(sr);

    String linha; //linha atual do arquivo

    int indiceFimNome; //tamanho do nome da sequencia sendo lida

    String auxNomeSequencia; //nome da sequencia encontrada e que sera
adicionada na lista de retorno

    while ((linha = bf.readLine()) != null) {

        if (isInicioSequencia(linha)) {

            //Inicio de sequencia a ser analisada

```

```

        bf.readLine(); //pula uma linha

        //Le as sequencias ate encontrar linha em branco indicando
o fim da sequencia

        while ((linha = bf.readLine()) != null) {

            if (linha.trim().equalsIgnoreCase("")) {

                //Se encontrar uma linha em branco indica o
fim da sequencia de nomes

                break;

            }

            indiceFimNome = linha.indexOf(" "); //Procura
'espaco' indicando fim do nome da sequencia

            if (indiceFimNome == -1) throw new Exception("Nao
foi possivel extrair o nome da sequencia!\n[" + linha + "]);

            auxNomeSequencia = linha.substring(0,
indiceFimNome);

            listaNomesSequencias.put(auxNomeSequencia, "");

        } //end While

    } //end if

} //end while

    log("OK\n");

    return listaNomesSequencias;

}

/**
 * Verifica se a String informada indica inicio de sequencia

```

```

    * @param linha

    * @return

    */

private boolean isInicioSequencia(String linha) {

    boolean encontrou = false;

    if (linha.startsWith(STR_START_SEQUENCE)) {

        encontrou = true;

    }

    return encontrou;

}

/**

 * Verifica se a linha passada contem o nome Query da sequencia. Caso contrario,
retorna

 * o nome passado em querySequence.

 * @param linha

 * @param querySequence

 * @return

 */

private String getIfQuerySequenceName(String linha, String querySequence) {

    String retorno = querySequence; //Retorna o valor atual caso a linha nao
contenha outro nome

    if (linha != null && linha.startsWith(STR_START_QUERY_SEQUENCE_NAME)) {

        retorno = linha.substring(STR_START_QUERY_SEQUENCE_NAME.length(),
linha.length());

    }

```

```

        return retorno;
    }

/**
 * Identifica nome da sequencia do fasta
 *
 * @param linha
 *
 * @return
 */

private String getFastaSequenceName(String linha) {

    int indexEspaco;

    indexEspaco = linha.indexOf(" ");

    if (indexEspaco == -1) {

        indexEspaco = linha.length(); //Se nao tiver ' ', o nome da
sequencia é a linha toda

    }

    String nome = linha.substring(1,indexEspaco);

    return nome;

}

/**
 * Verifica se os argumentos passados sao validos
 *
 * @param args
 *
 * @return
 */

private static boolean argumentosValidos(String[] args) {

```

```

boolean validos = false;

if (args != null && args.length == 6) {

    try {

        File f1 = new File(args[0]);

        File f2 = new File(args[1]);

        if (f1.exists() && f2.exists()) {

            validos = true;

        }

        else {

            log("Arquivo nao encontrado.[" + (f1.exists()?
f2.getName() : f1.getName()) + "]\n");

        }

        File f3 = new File(args[2] + File.separator + "bin" +
File.separator + BLAST_EXE);

        if (!f3.exists()) {

            validos = false;

            log("Diretorio informado para a aplicação Blast é
inválido!\n");

        }

        try {

            Integer.parseInt(args[5]);

        }

        catch(Exception e) {

```



```

        validos = false;

        log("O valor informado para a tabela de aminoácidos
não é numerico\n");

    }

}

catch(Exception e){

    log("Nome de arquivo invalido.");

    e.printStackTrace();

}

}

return validos;

}

/**

 * Exibe ajuda sobre utilizacao do programa

 *

 */

private static void showHelp() {

    log("\n\nUso:   java   -jar   ETTool.jar   blastFile   fastaDatabase
diretorio_install_blast nome_banco_proteina nome_final_output nro_tabela");

    log("O novo arquivo sera gerado no diretorio atual.");

}

/**

 * Carrega as sequencias do arquivo fasta, cujos nomes estavam no blast

```

```

    * @throws Exception

    */

    private void findFastaSequences(Map<String, String> listaNomesSequencias) throws
Exception {

        log("Localizando no fasta...");

        //Abre o arquivo do fasta para leitura

        FileInputStream fin = new FileInputStream(fastaFileName);

        InputStreamReader sr = new InputStreamReader(fin);

        BufferedReader bf = new BufferedReader(sr);

        String linha; //linha sendo lida do arquivo

        String auxSequenceName = null; //codigo da sequencia

        StringBuffer auxSequence = null;

        boolean lendoSequencia = false;

        while ((linha = bf.readLine()) != null) {

            //Verifica se nova sequencia esta se iniciando

            if (linha.startsWith(FASTA_MARCADOR)) {

                if (lendoSequencia) {

                    //Se outra sequencia se inicia, vamos gravar o que
foi lido da anterior.

                    listaNomesSequencias.put(auxSequenceName,
auxSequence.toString());

                }

                auxSequenceName = getFastaSequenceName(linha); //so o
codigo do nome da sequencia

```

```

//Se a sequencia do fasta existe no blast, vamos carrega-la

if (listaNomesSequencias.get(auxSequenceName) != null) {

    lendoSequencia = true;

    auxSequence = new StringBuffer(10000); //inicializa
sequencia lida

    }

else {

    lendoSequencia = false; //se a sequencia do fast
nao esta no blast, nao vamos carrega-la.

    }

else if (lendoSequencia) {

    auxSequence.append(linha);

    }

}

log("OK\n");

}

/**
 * Exibe no console as mensagens do sistema
 * @param msg
 */

private static void log(String msg) {

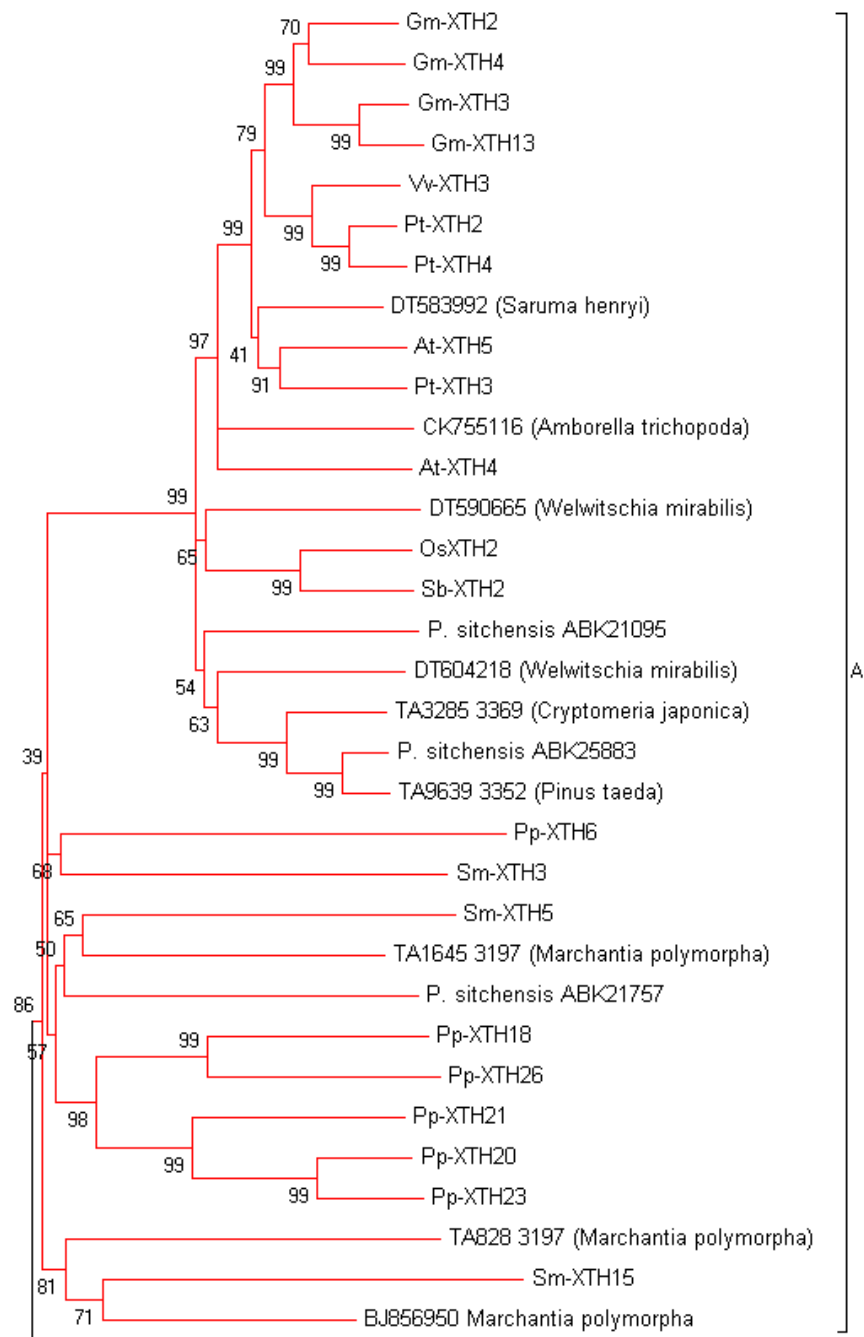
    System.out.print(msg);

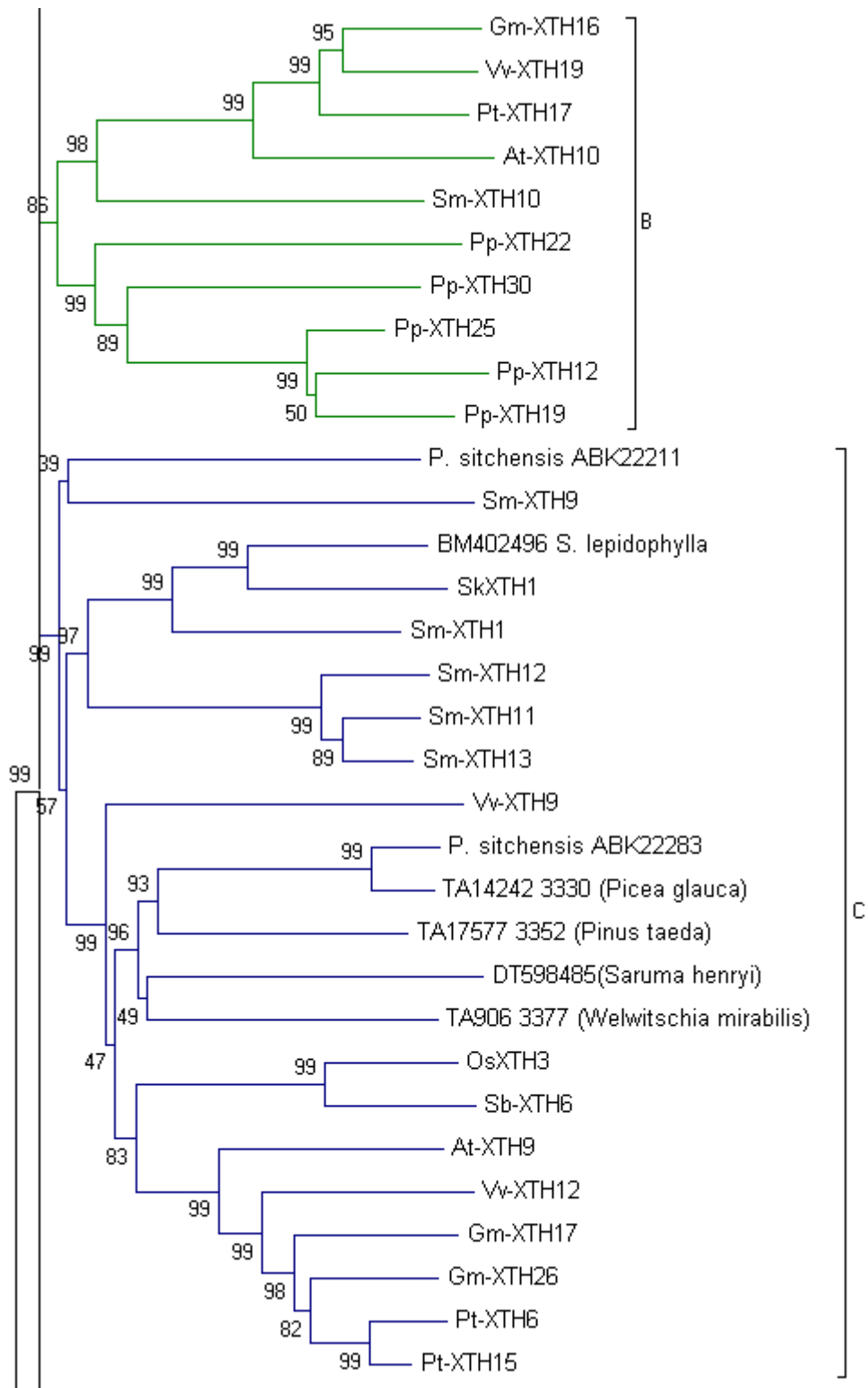
}

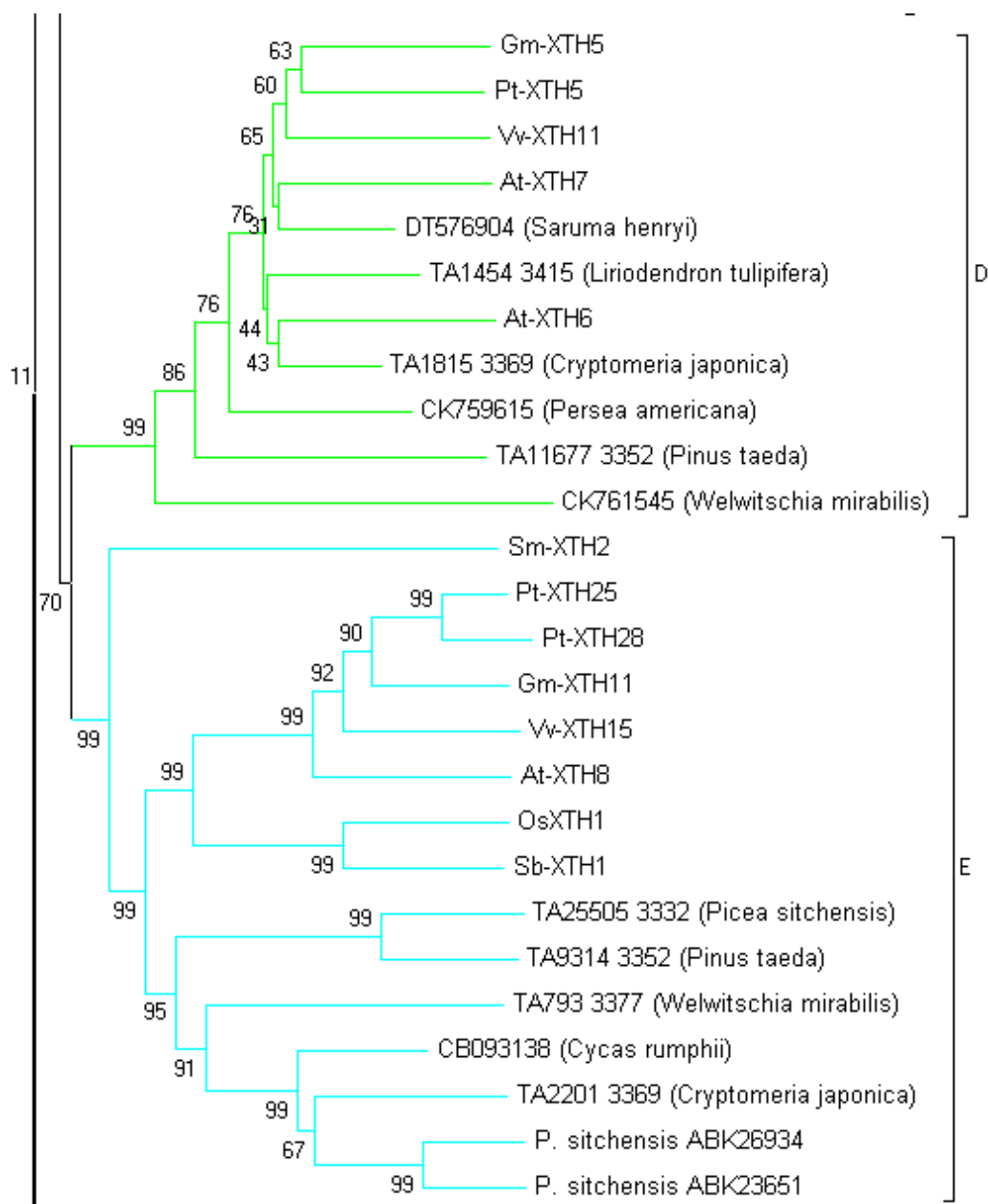
}

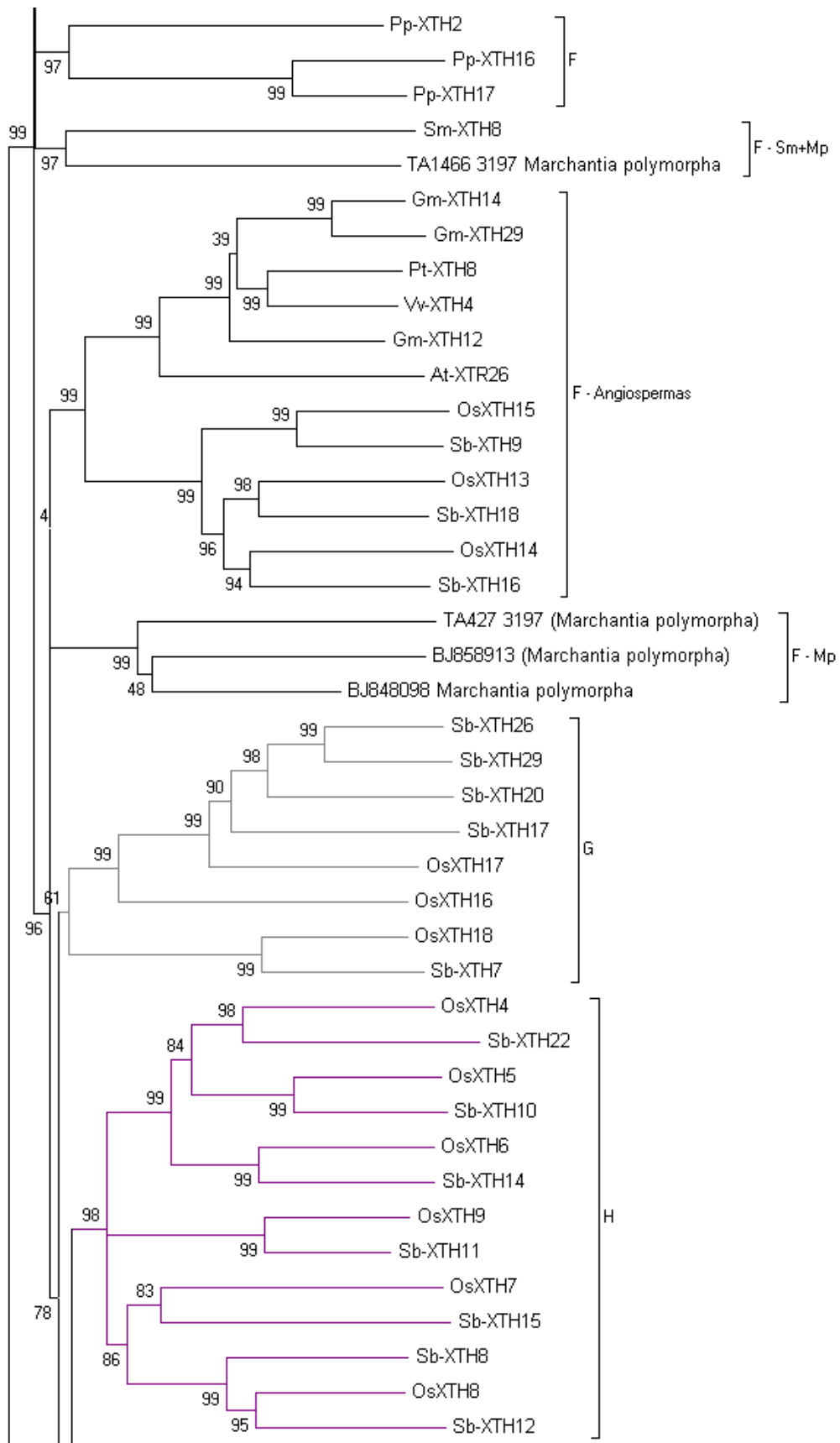
```

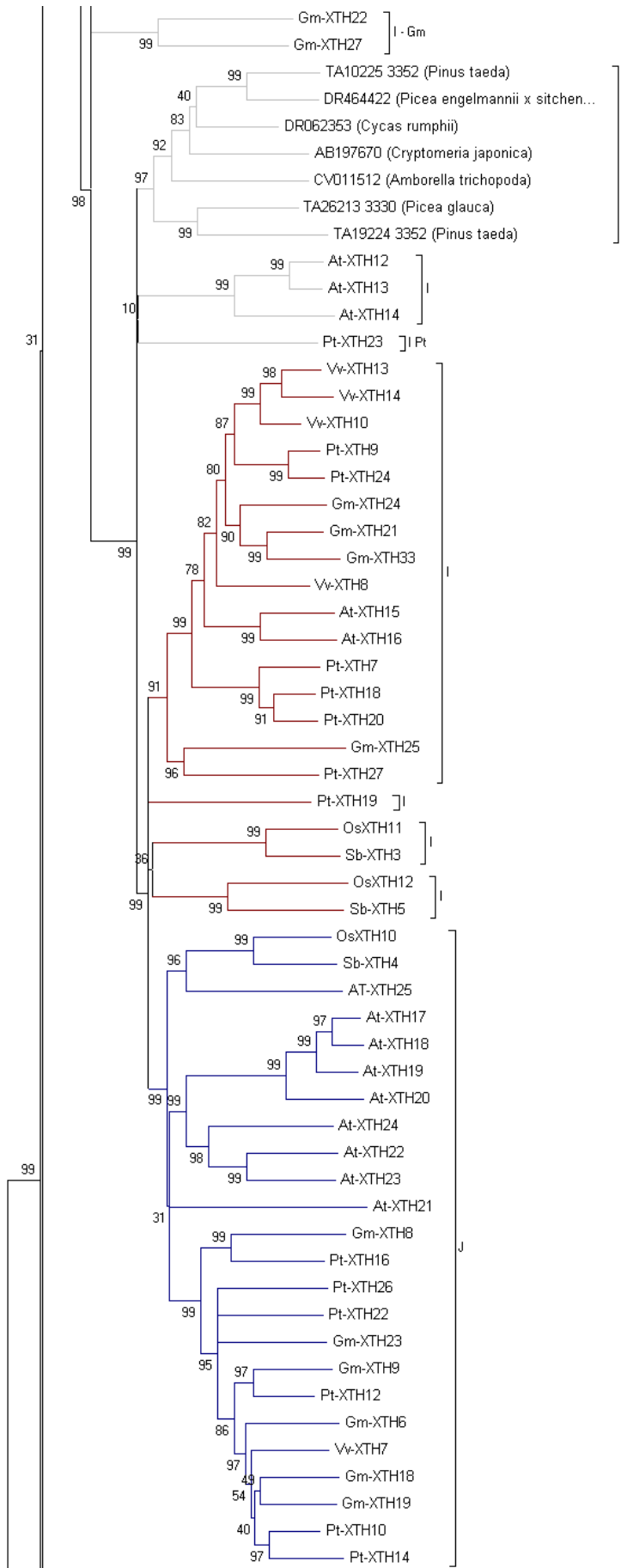
ANEXO III – Árvore filogenética detalhada das XTHs.



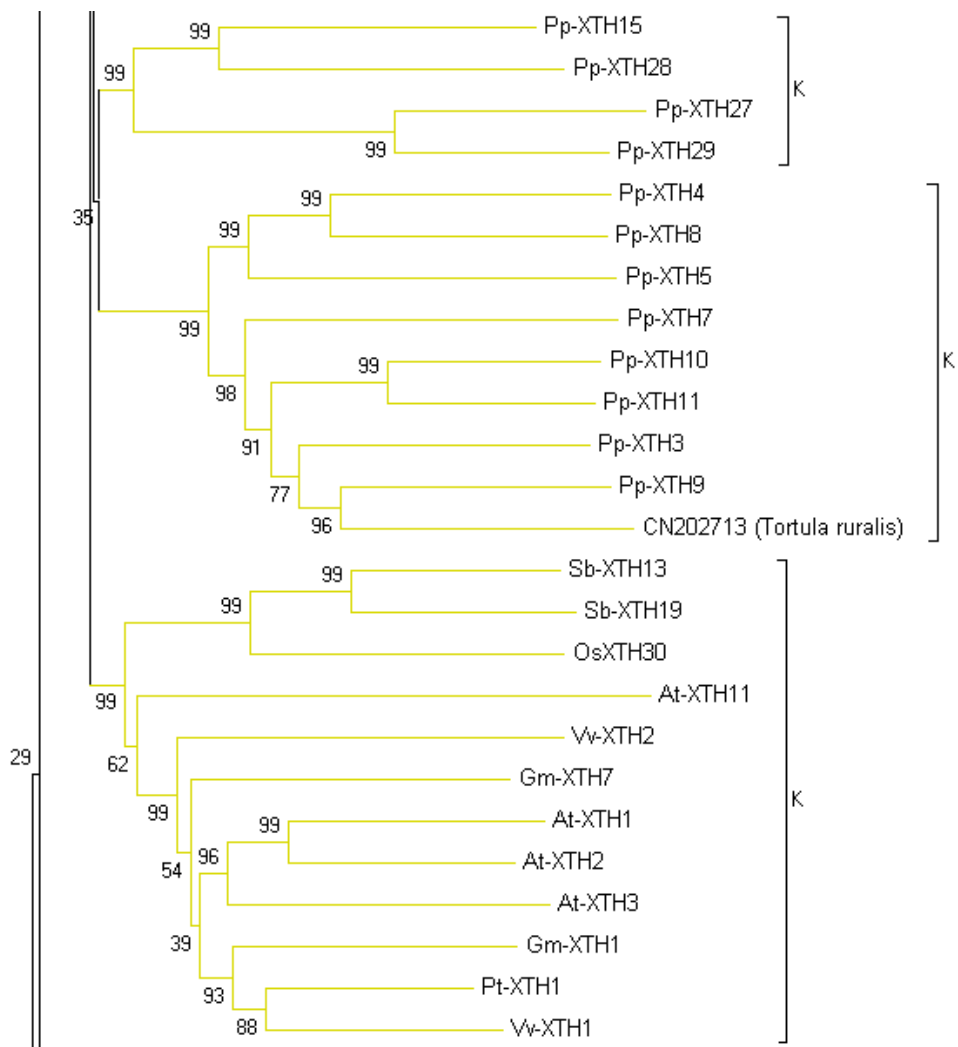


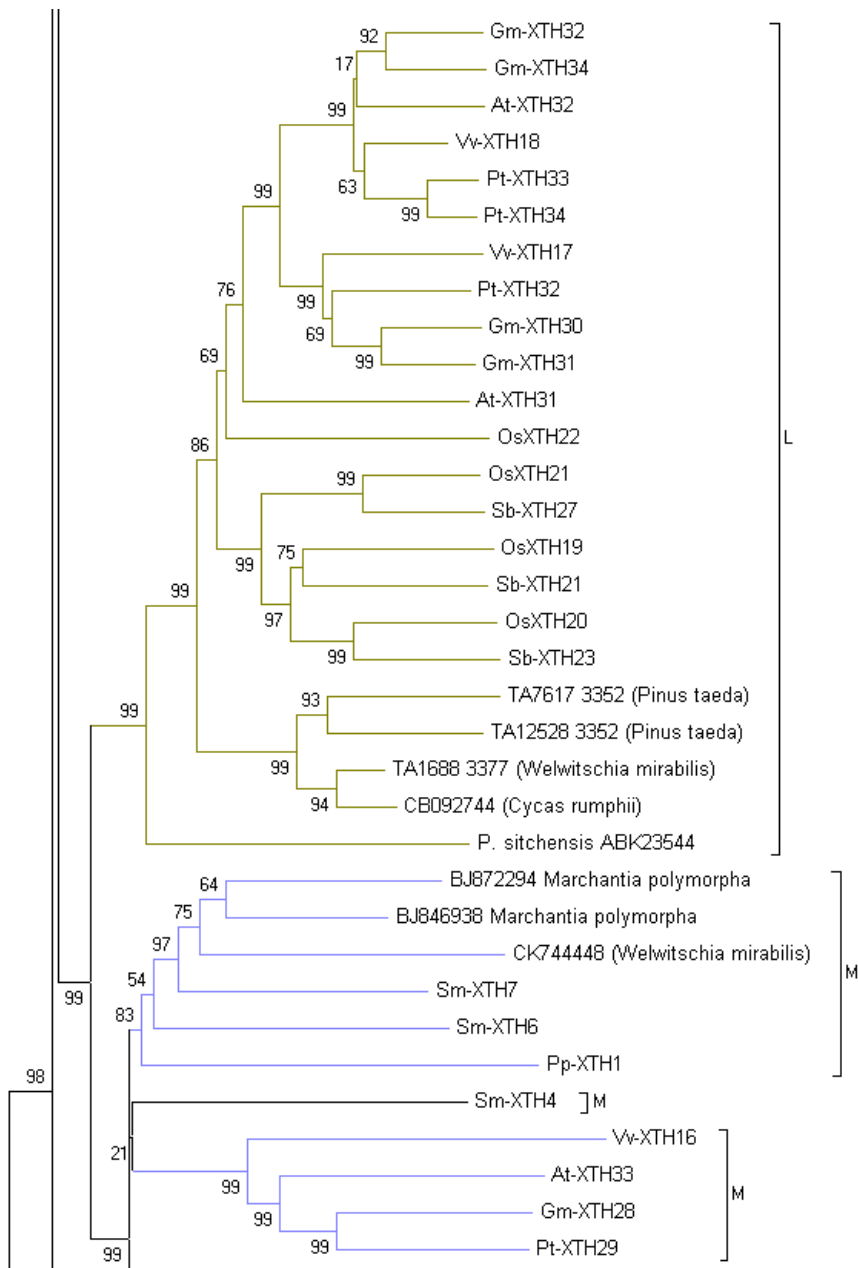


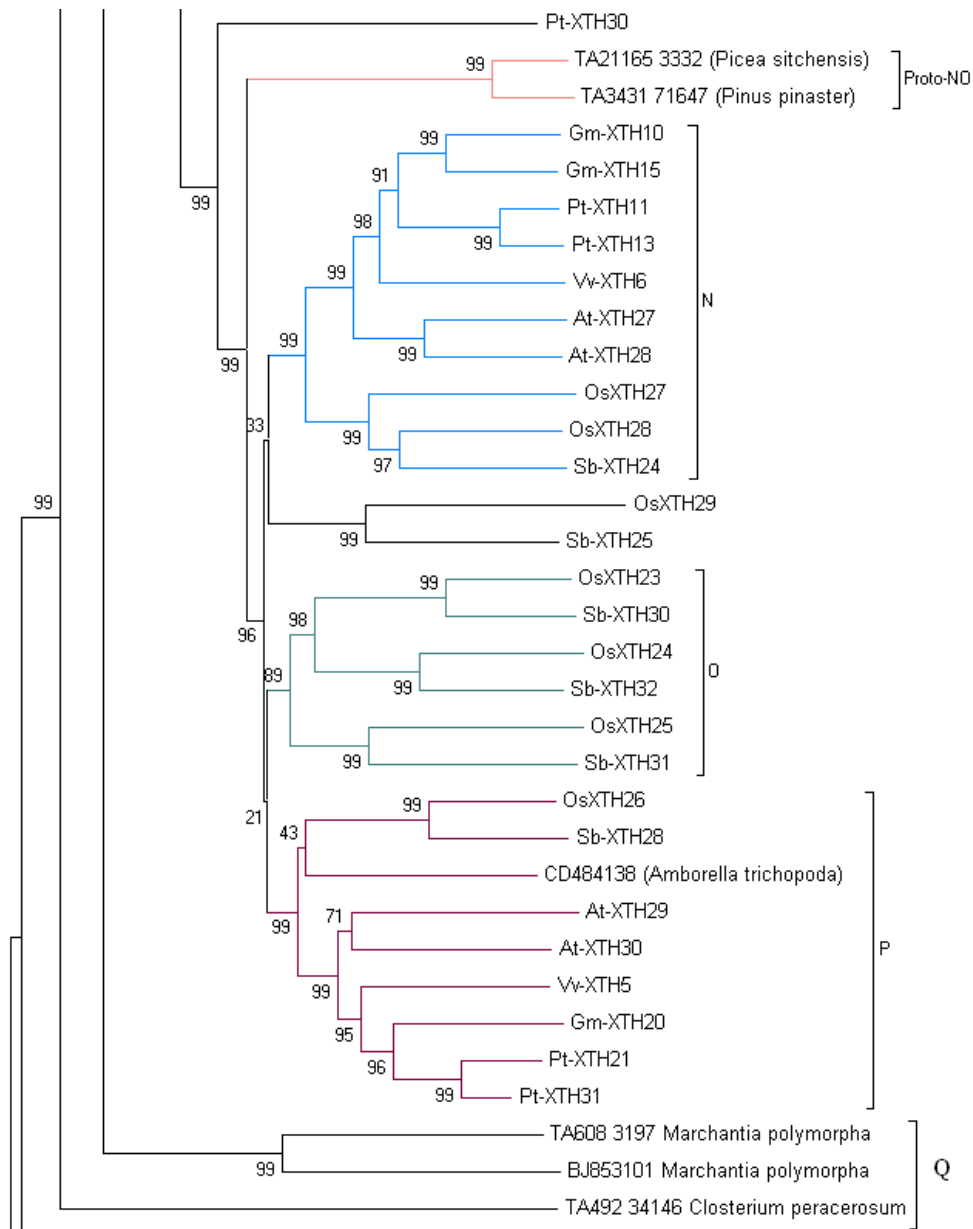


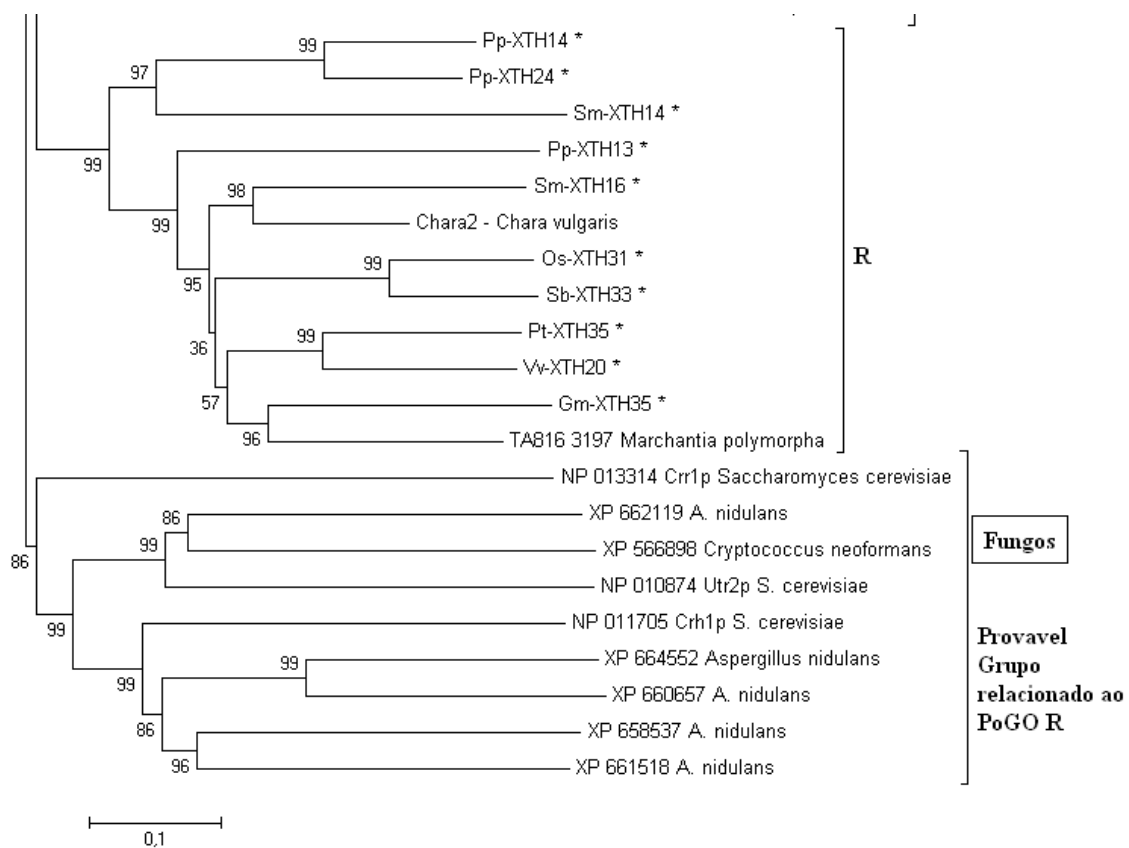












**Figura 1 (ANEXO III).** Árvore filogenética contendo todas as XTHs analisadas. Os nomes dos PoGOs estão indicados pelas letras ao lado das chaves que os delimitam. A topologia foi inferida por NJ, as distâncias calculadas por *p-distance*. Foram realizadas 1000 réplicas de *bootstrap*. Não fixamos *out group*. Todos os valores de *bootstrap*, entre 0 e 100%, estão mostrados nos nós.