

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

NATALIA BALBINOTT

Evolução molecular das cistatinas de plantas: ênfase em fitocistatinas do tipo II

Dissertação de Mestrado

PORTO ALEGRE

2021

CIP - Catalogação na Publicação

Balbinott, Natalia
Evolução molecular das cistatinas de plantas:
ênfase em fitocistatinas do tipo II / Natalia
Balbinott. -- 2021.
83 f.
Orientador: Rogério Margis.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Centro de Biotecnologia do Estado
do Rio Grande do Sul, Programa de Pós-Graduação em
Biologia Celular e Molecular, Porto Alegre, BR-RS,
2021.

1. fitocistatinas. 2. evolução molecular. I.
Margis, Rogério, orient. II. Título.

NATALIA BALBINOTT

Evolução molecular das cistatinas de plantas: ênfase em fitocistatinas do tipo II

Dissertação submetida ao Programa De Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Mestre em Biologia Celular e Molecular.

Orientador: Rogério Margis

Porto Alegre

Abril de 2021

Este trabalho foi desenvolvido no Laboratório de Genomas e Populações de Plantas (LGPP), localizado no Departamento de Biofísica da Universidade Federal do Rio Grande do Sul, com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES).

“

The only trust required is to know that when there is one ending
there will be another beginning.

”

Clarissa Pinkola Estés

AGRADECIMENTOS

Agradeço inicialmente ao meu orientador, professor Dr. Rogério Margis, pelo apoio e pela confiança depositada em mim para realizar este e outros trabalhos paralelos. Pela abertura para discutir ideias, relacionadas com ciência ou não, pela paciência e incentivo. Sou grata pela excelente referência de pesquisador e pessoa que tive como orientador nesses dois anos.

À professora Dra. Andreia Turchetto-Zolet, pelo suporte teórico e discussões que contribuíram para a realização desse trabalho, essenciais para a realização das análises evolutivas.

Aos membros da comissão de acompanhamento, Dr. Arthur Fett Neto e Dr. Felipe Ricachenevsky, pelas conversas e discussões. Agradeço também aos membros da banca, Dr. Carlos Priminho Pirovani, Dr. Flávio Henrique Silva e novamente Dr. Felipe Ricachenevsky, e ao revisor interno Dr. Yugo Lima Melo, por analisarem e contribuírem para esse trabalho.

Aos professores e funcionários do PPGBCM da UFRGS, sempre muito dispostos e eficientes. À UFRGS, CNPq e CAPES, pela estrutura e apoio financeiro que oportunizaram a realização desse trabalho.

Aos meus amigos e colegas do LGPP, Nureyev, Isabel, Érika, Igor, Pabulo, Cordenonsi e Frank. Agradeço pela disposição em compartilhar conhecimentos, pelo apoio e camaradagem que tornaram minha trajetória mais leve. À *meine Freundin* Pia Engel, cuja passagem pelo laboratório foi breve mas suficiente para construirmos uma amizade que cruza barreiras geográficas.

Aos meus pais, Neura e Moacyr, que dedicam as suas vidas pelo sucesso e felicidade dos filhos, por sempre acreditarem em mim e apoiarem minhas escolhas. Agradeço aos meus irmãos Marcos e Carolina, pelo suporte incondicional que me dão durante toda minha vida.

À minha grande amiga Mariane, que acompanhou de perto minha jornada e foi uma das facilitadoras da minha vinda à Porto Alegre. Obrigada por sempre estar ao meu lado, seja para celebrar as pequenas alegrias da vida ou afogar as mágoas com um clericot ao som de “o Brasil que deu quase certo”.

Ao meu parceiro, melhor amigo e companheiro de vida, Davi, com quem compartilho essa jornada diariamente e cujo apoio foi essencial para que eu chegasse até aqui. Te amo, Roma.

Por fim, agradeço a todos que contribuíram direta ou indiretamente para a realização desse trabalho.

SUMÁRIO

LISTA DE ABREVIATURAS	9
LISTA DE FIGURAS E TABELAS	10
RESUMO	11
ABSTRACT	12
1. INTRODUÇÃO.....	13
1.1 BASES MOLECULARES DA EVOLUÇÃO.....	13
1.2 EVOLUÇÃO EM VIRIDIPLANTAE	16
1.3 CISTATINAS	20
1.4 FITOCISTATINAS.....	21
2. OBJETIVOS	25
2.1 OBJETIVO GERAL	25
2.2 OBJETIVOS ESPECÍFICOS	25
3. MATERIAIS E MÉTODOS.....	26
3.1 LEVANTAMENTO EM BANCO DE DADOS	26
3.2 ALINHAMENTO E ANÁLISE FILOGENÉTICA.....	26
3.3 CONSERVAÇÃO DE SEQUÊNCIA.....	28
3.4 ANÁLISES DA PRESSÃO DE SELEÇÃO POSITIVA EM FITOCISTATINAS DO TIPO II.....	28
4. RESULTADOS E DISCUSSÃO	31
4.1 PRESENÇA DE FITOCISTATINAS EM VIRIDIPLANTAE E A FILOGENIA DE FITOCISTATINAS.....	31
4.2 FILOGENIA DE FITOCISTATINAS TIPO II.....	39
4.3 ANÁLISE DE PRESSÃO DE SELEÇÃO.....	41
4.4 FILOGENIA DAS FITOCISTATINAS	43
5. CONSIDERAÇÕES FINAIS E PERSPECTIVAS.....	45
APÊNDICE 1 – Bancos de dados utilizados para a busca de sequências de fitocistatinas.....	52
APÊNDICE 2 – Relação das 188 espécies cujas sequências de fitocistatinas foram utilizadas neste estudo.	53
APÊNDICE 3 – Sequências aminoacídicas das fitocistatinas presentes em arroz, usadas como isca para as buscas em bancos de dados.	60

APÊNDICE 4 – Parâmetros utilizados para a inferência filogenética bayesiana.	61
APÊNDICE 5 – Alinhamento das fitocistatinas do tipo I, com resíduos coloridos de acordo com a conservação.	62
APÊNDICE 6 – Alinhamento das fitocistatinas do tipo II, com resíduos coloridos de acordo com a conservação.	69
APÊNDICE 7 – Árvore filogenética não-enraizada das fitocistatinas no tipo II utilizada para os testes de seleção positiva.	80
CURRICULUM VITÆ	81

LISTA DE ABREVIATURAS

ABA – ácido abscísico, do inglês *abscisic acid*

CDS – sequência codificadora, do inglês *coding sequence*

dN – taxa de substituições não-sinônimas

dS – taxa de substituições sinônimas

lnL – log-likelihood

MRCA – ancestral comum mais recente, do inglês *most recent common ancestor*

PhyCys-I – fitocistatina do tipo I, do inglês *phytocystatin*

PhyCys-II – fitocistatina do tipo II

WGD – duplicação completa do genoma, do inglês *whole genome duplication*

LRTs – teste da razão de verossimilhança, do inglês *likelihood ratio test*

ω – razão entre a taxa de substituições não-sinônimas e substituições sinônimas (dN/dS)

MCMC – algoritmos de Monte Carlo acoplados a cadeias de Markov, do inglês *Markov chain Monte Carlo*

ESS – tamanho de amostra efetivo, do inglês *effective sample size*

OTUs – unidade taxonômica operacional, do inglês *operational taxonomic unit*

LISTA DE FIGURAS E TABELAS

FIGURA 1 – Mutações a nível de nucleotídeo.

FIGURA 2 – Resumo da filogenia de Viridiplantae.

FIGURA 3 – Mecanismos de duplicação gênica.

FIGURA 4 – Modelos estruturais da cistatina de arroz.

FIGURA 5 – Modelos estruturais dos diferentes tipos de fitocistatinas

FIGURA 6 – Cladograma exibindo relações de Viridiplantae e o número correspondente de fitocistatinas tipo I e tipo II por espécie.

FIGURA 7 – Representação gráfica das relações entre sequências de fitocistatinas resultante de uma reconstrução filogenética bayesiana não-enraizada.

FIGURA 8 – Reconstrução filogenética bayesiana não-enraizada da subfamília das fitocistatinas.

FIGURA 9 – Nível de conservação de sequência das fitocistatinas do tipo I.

FIGURA 10 – Nível de conservação de sequência das fitocistatinas do tipo II.

FIGURA 11 – Reconstrução filogenética bayesiana de fitocistatinas do tipo II.

FIGURA 12 – Logo de sequências de fitocistatinas do tipo II dos dois ramos de duplicação de Eudicots, D1 e D2 e estrutura 3D indicando posições com alterações de carga significativas.

FIGURA 13 – Resumo da história evolutiva das fitocistatinas.

TABELA 1 – Modelos de códons empregados para o teste de seleção positiva nos sítios.

TABELA 2 – Comparações de modelos aninhados para testes de hipótese de seleção positiva.

TABELA 3 – Resultados do *codeml* para a filogenia completa das fitocistatinas do tipo II.

TABELA 4 – Resultados do *codeml* para a filogenia das fitocistatinas do tipo II de eudicotiledôneas.

RESUMO

As fitocistatinas englobam uma família de inibidores de proteases cisteínicas envolvidos em processos biológicos como regulação da degradação de proteínas durante a germinação, desenvolvimento da planta e morte celular, além de respostas a estresses abióticos e bióticos. Com expressão ubíqua, as fitocistatinas são classificadas em subfamílias de acordo com seu peso molecular e especificidade inibitória. Os genomas e transcriptomas de plantas disponíveis em nossa era de sequenciamento de alto rendimento permitiram a identificação de genes desta família em diversas espécies e uma reconstrução filogenômica a fim de compreender melhor a evolução molecular das cistatinas vegetais. Usando sequências de fitocistatinas dos tipos I e II, reconstruímos a filogenia da subfamília das fitocistatinas. Observamos dois grupos: um com fitocistatinas do tipo I (PhyCys-I) e fitocistatinas do tipo II (PhyCys-II) de Viridiplantae que possuem íntrons em sua estrutura gênica, e outro com PhyCys-I de angiospermas que não possuem íntrons. Nossos resultados sugerem que as PhyCys-I com íntrons se originaram a partir de múltiplos eventos de perda da porção carboxi-extendida de PhyCys-II. As PhyCys-I sem íntrons, por sua vez, teriam se originado a partir de um evento de retroduplicação compartilhado por todas as plantas com flores, seguido de uma duplicação gênica que levou à formação de dois subgrupos. Dois eventos principais de duplicação gênica também foram identificados em PhyCys-II: um em gimnospermas e outro em eudicotiledôneas. Apesar da elevada conservação a nível de sequência, substituições de aminoácidos com características químicas distintas podem ser identificadas entre as duas formas duplicadas em angiospermas. Nossas análises não identificaram nenhum sítio sob seleção positiva na sequência de PhyCys-II.

ABSTRACT

Phytocystatins encompass a family of inhibitors of cysteine proteases involved in biological processes such as regulation of protein turnover during germination, development and cell death, in addition to responses to abiotic and biotic stresses. With ubiquitous expression, phytocystatins are classified into subfamilies according to their molecular weight and inhibitory specificity. Plant genomes and transcriptomes available in our high-throughput sequencing era enabled the identification of genes from this subfamily in several species and allowed a phylogenomic reconstruction to better understand the molecular evolution of plant cystatins. In the present work, we explored the evolution of the phytocystatin gene subfamily, with a special focus on bifunctional phytocystatins. Using sequences of types I and II phytocystatins, we reconstructed the phylogeny of the phytocystatin subfamily. We observed two groups: one containing type-I phytocystatins (PhyCys-I) and type-II phytocystatins (PhyCys-II) from Viridiplantae that have introns in their gene structure, and another with intronless PhyCys-I from angiosperms. Our results suggest that PhyCys-I with introns originated from multiple events of loss of the carboxy-extended portion of PhyCys-II. In turn, intronless PhyCys-I would have originated from a retroduplication event shared by all flowering plants, followed by a gene duplication that led to the formation of two subgroups. Two major gene duplication events were also identified in PhyCys-II: one in gymnosperms and another in Eudicots. Despite the high conservation at the sequence level, substitutions of amino acids with different chemical properties can be identified between the two duplicated forms in angiosperms. Our analyses could not identify any sites under positive selection in the sequence of PhyCys-II.

1. INTRODUÇÃO

1.1 BASES MOLECULARES DA EVOLUÇÃO

O grande motor da evolução em nível molecular é o acúmulo de mutações nas sequências de DNA ao longo das gerações. As mudanças nas sequências nucleotídicas ocorrem através de substituições, recombinações, deleções, inserções ou inversões (Figura 1a). Tais mutações são a principal fonte de variabilidade e novidades na evolução dos organismos ao se propagarem pelas populações e eventualmente se fixarem em uma espécie (NEI; KUMAR, 2000).

As substituições podem ser classificadas como transições – mudanças entre purinas (A e G) ou pirimidinas (C e T) – ou transversões – mudanças entre uma purina e uma pirimidina (Figura 1b). Quando ocorrem em regiões codificadoras, as substituições podem ainda ser classificadas de acordo com seu efeito na sequência de aminoácidos produto da tradução. Devido à característica degenerada do código genético, um mesmo resíduo de aminoácido pode ser codificado por diferentes códons. Deste modo, uma mutação que altera o códon mas não causa mudança no aminoácido codificado é dita como sinônima. Por outro lado, uma mudança na sequência nucleotídica que altere o aminoácido codificado é dita como não-sinônima. Ainda existe um terceiro tipo considerado como mutação conservativa, na qual o aminoácido codificado é diferente, mas mantém as características bioquímicas do aminoácido ancestral.

A extensão, a frequência e o efeito das mutações dependem da região do DNA em que ela se encontra e das diversas forças de seleção nela atuando. De acordo com a síntese evolutiva moderna, a seleção natural possui papel preponderante para moldar a composição genética de populações e o processo de substituição de genes. Dentro de uma perspectiva seletivista do processo evolutivo, substituições de sequências gênicas são uma consequência de um processo adaptativo positivo, no qual um novo alelo predomina em gerações futuras de uma população apenas se ele aumenta a capacidade adaptativa dos organismos. Polimorfismos, por outro lado, são mantidos por seleção balanceadora, mantidos

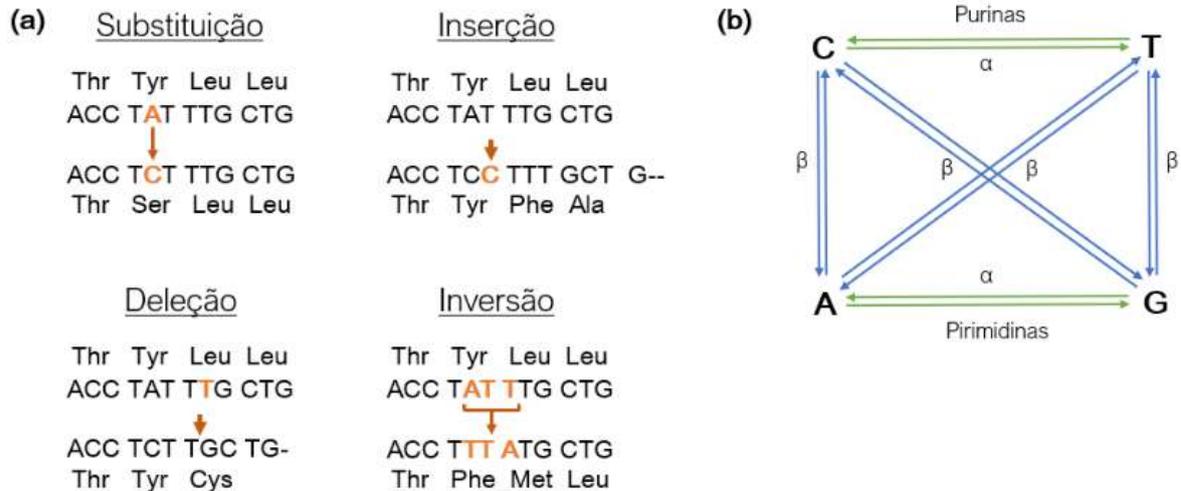


FIGURA 1 – Mutações a nível de nucleotídeo. (a) Tipos básicos de mutações ao nível de nucleotídeo e seu efeito na sequência aminoacídica. (b) Substituições nucleotídicas do tipo transições (verde) e transversões (azul). α e β indicam as taxas de transições e transversões, respectivamente. Figura adaptada de Nei e Kumar (2000).

quando a coexistência de dois ou mais alelos em um *locus* é vantajoso para o organismo ou para a população (GRAUR; LI, 2000).

Alternativamente, a teoria neutralista da evolução molecular postula que a maior parte das mudanças evolutivas no nível molecular são causadas pela fixação aleatória de mutantes seletivamente neutros, não gerando consequências adaptativas grandiosas aos organismos. Nessa perspectiva, a maioria das mudanças evolutivas e muito da variabilidade genética de uma espécie não são causadas nem por seleção positiva de alelos vantajosos, nem pela seleção balanceadora, mas sim pela deriva genética de alelos mutantes que são seletivamente neutros. Assim, substituições e polimorfismos são duas facetas do mesmo fenômeno, uma vez que *loci* polimórficos consistem de alelos que estão a caminho da fixação ou a caminho da extinção, já que o que visualizamos é uma etapa transiente de um processo dinâmico em curso.

A teoria neutralista não impede a adaptação. De acordo com Kimura (1983), mudanças ambientais podem atuar contra alelos que anteriormente eram considerados neutros, tornando-os deletérios e sujeitos à seleção purificadora. Após a remoção desses alelos, a população se tornará mais adaptada a novas

circunstâncias, fazendo com que a evolução adaptativa ocorra na ausência de seleção positiva.

Ambas teorias concordam que a maioria das novas mutações que emergem são deletérias e, portanto, são removidas rapidamente da população (algumas poucas gerações) e não contribuem nem para a taxa de substituição nem para a quantidade de polimorfismos em uma população. As diferenças entre as teorias são sobre a proporção relativa de mutações neutras conjuntamente as mutações não-deletérias. Seleccionistas afirmam que poucas mutações são seletivamente neutras, enquanto neutralistas afirmam que a maioria das mutações não-deletérias são efetivamente neutras (GRAUR; LI, 2000).

Kimura e Ohta (1974) postularam cinco princípios que governam a evolução molecular: (i) para cada proteína, a taxa de evolução, em termos de substituição de aminoácidos é aproximadamente constante por sítio e por ano para várias linhagens, desde que a função e a estrutura terciária da molécula permaneçam essencialmente inalteradas; (ii) moléculas funcionalmente menos importantes ou partes de uma molécula evoluem (em termos de substituições mutantes) mais rápido que as mais importantes; (iii) as substituições mutantes que perturbam menos a estrutura e função existentes de uma molécula (substituições conservativas) ocorrem com maior frequência na evolução que as mais disruptivas; (iv) duplicações gênicas devem sempre preceder a emergência de um gene tendo uma nova função; (v) a eliminação seletiva de mutantes deletérios e fixação aleatória de mutantes neutros ou pouco deletérios ocorre com maior frequência na evolução do que a seleção Darwiniana positiva de mutantes garantidamente vantajosos.

Em termos práticos, a principal ferramenta utilizada na detecção de seleção em dados comparativos de sequências homólogas é a razão da taxa de mutações não-sinônimas, dN , e da taxa de mutações sinônimas, dS , ($dN/dS = \omega$) (HUGHES; NEI, 1988). Na ausência de seleção, substituições sinônimas e não-sinônimas devem ocorrer a uma mesma proporção, o que resultaria em $\omega = 1$. Se houver seleção negativa, a taxa de mutações sinônimas deve superar a de mutações não-sinônimas ($\omega < 1$), enquanto se houver seleção positiva a taxa de mutações não-sinônimas supera a de mutações sinônimas ($\omega > 1$).

1.2 EVOLUÇÃO EM VIRIDIPLANTAE

Viridiplantae, grupo também conhecido como “plantas verdes”, é um clado que engloba cerca de 500 mil espécies (LUGHADHA et al., 2016) e engloba uma enorme diversidade de formas de vida presente nos mais variados ambientes. Evidências fósseis sugerem que o clado tenha pelo menos 750 milhões de anos, enquanto que as estimativas de tempo de divergência geradas a partir de dados moleculares sugerem que possa ultrapassar um bilhão de anos (GITZENDANNER et al., 2018). Durante esse período, diversos mecanismos ecológicos impuseram pressões seletivas que impulsionaram a diversidade de inovações morfológicas, fisiológicas e de desenvolvimento que permitiram a colonização e exploração de novos e emergentes habitats. Assim, a evolução das plantas influenciou os ambientes da Terra e criou uma cascata de diversidade em outras linhagens da árvore da vida.

A maioria das análises filogenéticas de Viridiplantae indicam que se trata de um grupo monofilético (Figura 2) que teria se originado de um evento endossimbionte, no qual uma célula eucariótica heterotrófica engolfou uma forma de vida procariota fotossintética, semelhante às atuais cianobactérias. A partir disso, essas duas formas de vida ficaram estavelmente integradas, o que eventualmente evoluiu a uma organela envolta por membranas, o cloroplasto (GOULD; WALLER; MCFADDEN, 2008). Esse evento único marcou a origem da fotossíntese em eucariotos e a três linhagens autotróficas com plastídios primários: as plantas verdes (Viridiplantae), as algas vermelhas (Rhodophyta) e as algas verde-azuladas (Glaucophyta). A partir desse ponto inicial, a fotossíntese se espalhou amplamente entre eucariotos através de eventos endossimbióticos secundários e terciários, resultando em uma complexa história de aquisição de plastídeos (GOULD; WALLER; MCFADDEN, 2008; KEELING, 2004, 2010).

Uma divergência inicial na evolução de Viridiplantae deu origem a duas linhagens principais, Chlorophyta e Streptophyta. Chlorophyta se diversificou como plâncton nos oceanos (Prasinophytes) e ambientes marinhos costeiros e de água doce (*core* Chlorophyta), e manifesta uma enorme variedade de formas corporais,

características e estratégias ecofisiológicas. Streptophyta evoluiu em água doce e habitats terrestres úmidos e colonizou o ambiente terrestre há aproximadamente 476-432 milhões de anos, dando origem às plantas terrestres. Atualmente, esse clado engloba algas de água doce (Carophyta) e as plantas terrestres (Embryophyta) (MCCOURT; DELWICHE; KAROL, 2004; RUHFEL et al., 2014).

As plantas terrestres podem ser divididas em duas principais linhagens, briófitas e traqueófitas. Essas, por sua vez, podem ser divididas em três linhagens monofiléticas, sendo Bryophyta (musgos), Marchantiophyta (hepáticas) e Anthocerotophyta (hornworts), e Lycopodiophyta (licófitas), Monilophytas (samambaias) e Spermatophyta (plantas com sementes), respectivamente. Dentre as Spermatophytas, vale destacar as linhagens de gimnospermas e angiospermas, sendo que esta última é o maior clado de plantas terrestres, composto por mais de 370 mil espécies (LUGHADHA et al., 2016).

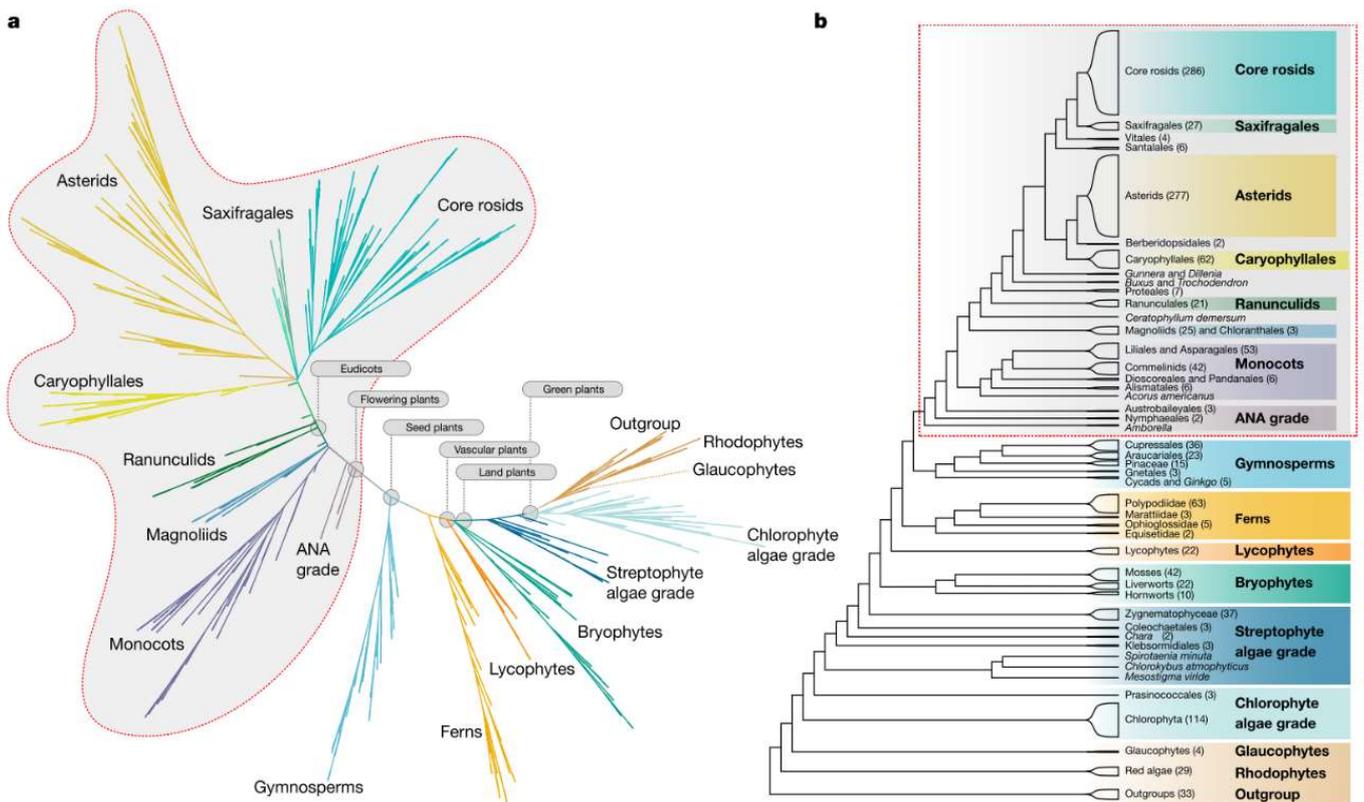


FIGURA 2 – Resumo da filogenia de Viridiplantae baseada inferida a partir de 410 famílias gênicas nucleares de cópia única 1153 taxons, incluindo 1090 espécies de Viridiplantae. (a) Filograma mostrando os comprimentos dos ramos internos proporcionais às unidades coalescentes entre eventos de ramificação. (b) Relacionamentos entre os principais clados, com a caixa vermelha evidenciando o clado das plantas com flores (Angiospermas). Figura retidada de Leebens-Mack et al. (2019).

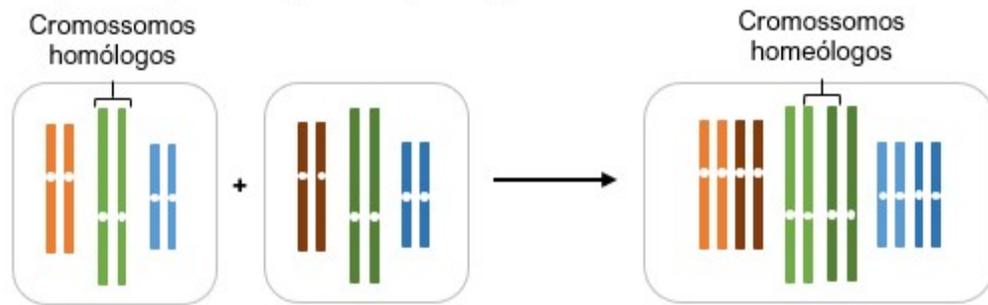
Os genomas das plantas modernas derivam de processos evolutivos complexos, envolvendo repetidos episódios de duplicação completa do genoma (*whole genome duplication*, WGD) e proliferação de elementos de transposição, levando à uma enorme diferença de tamanho entre genomas, mesmo de plantas evolutivamente próximas. Como consequência, as plantas possuem uma grande quantidade de genes duplicados, sendo a expansão de famílias gênicas um fator relevante no surgimento de inovações importantes (VAN DE PEER et al., 2009; WENDEL et al., 2016).

Eventos de WGD ocorreram múltiplas vezes, especialmente ao longo dos últimos 200 milhões de anos de evolução das angiospermas (VAN DE PEER; MAERE; MEYER, 2009). Embora seja um mecanismo extremo de duplicação gênica que leva ao aumento repentino do tamanho do genoma e no conjunto de genes, WGD não é o único mecanismo que dá origem a genes duplicados.

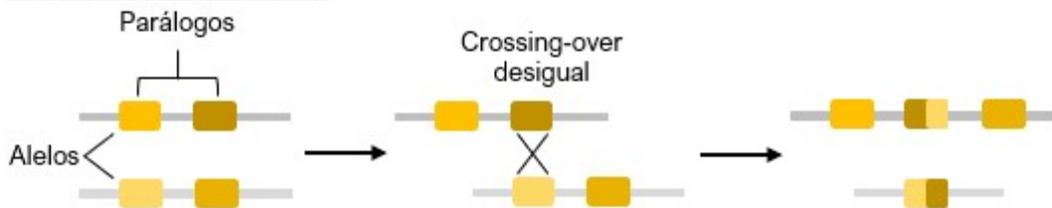
Genes e genomas duplicados podem fornecer “matéria-prima” para a diversificação evolutiva e a divergência funcional de genes duplicados pode oferecer vantagens adaptativas a poliploides. Em teoria, a duplicação gênica dá origem a duas cópias e permite que uma ou ambas evoluam sob pressão de seleção reduzida e, em alguns casos, adquiram novas funções que contribuam para a adaptação (Figura 3). Porém, o destino de genes duplicados é bastante diverso e imprevisível. Genes duplicados podem ser pseudogenizados ou perdidos, mantidos por seleção das funções já existentes ou pela seleção de novas funções. (PANCHY; LEHTI-SHIU; SHIU, 2016; QIAO et al., 2018).

A expansão e diversificação funcional de famílias gênicas seguidas de um evento de poliploidização é um grande fator influenciando a origem e rápida diversificação de linhagens de angiospermas. Diversos autores demonstram evidências de que a retenção de genes após a duplicação é enviesada não apenas de acordo com a função dos genes, mas também com o tempo e modo dos eventos de duplicação (BLANC; WOLFE, 2004; GUO, 2013; SEOIGHE; GEHRING, 2004).

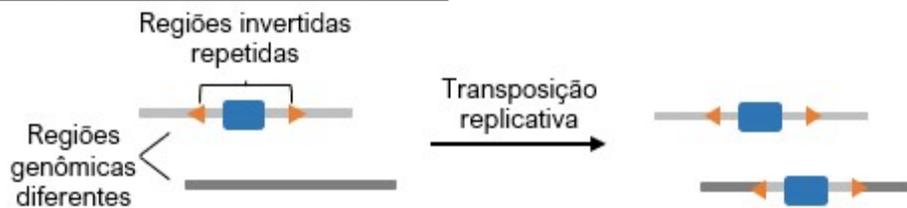
(a) Duplicação completa do genoma (WGD)



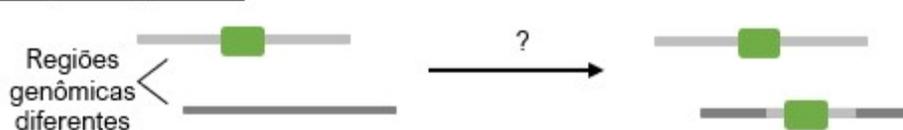
(b) Duplicação em tandem



(c) Duplicação mediada por transposon



(d) Duplicação segmental



(e) Retroduplicação

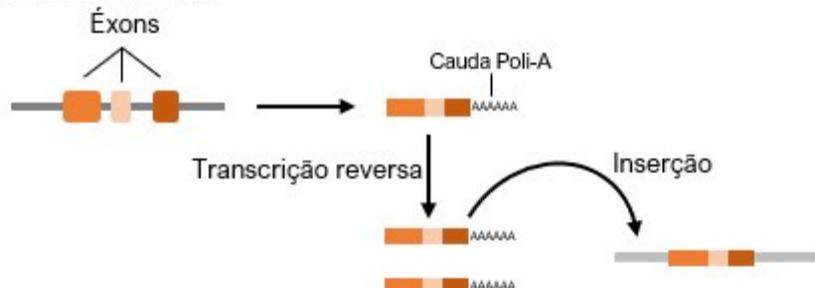


FIGURA 3 – Mecanismos de duplicação gênica. (a) Duplicação completa do genoma (WGD). (b) Duplicação em tandem. (c) Duplicação mediada por transposon. (d) Duplicação segmental. (e) Retroduplicação, ou duplicação de um gene via transcrição reversa do mRNA processado. Figura adaptada de Panchy, Lehti-Shiu & Shiu (2016).

1.3 CISTATINAS

As cistatinas são inibidores de proteases cisteínicas, função que dá origem ao nome cistatina. Elas compõem a superfamília I25 (MEROPS) e estão amplamente distribuídas em eucariotos. Os membros desta superfamília atuam como inibidores enzimáticos competitivos, coordenando a atividade de proteases do tipo papaína (família C1, MEROPS). Algumas cistatinas também possuem atividade inibitória contra proteases do tipo legumaina (família C13, MEROPS) por meio de um segundo sítio reativo.

As cistatinas compartilham uma conformação tridimensional conservada composta de cinco fitas beta antiparalelas sobrepostas a uma alfa-hélice central (Figura 4). As alças resultantes das estruturas de folhas-pregueadas antiparalelas contém o domínio cistatina, com atividade inibitória contra proteases do tipo papaína e formado por três motivos conservados: um resíduo de glicina (G) na região amino-terminal, um motivo Q-x-V-x-G e um motivo [A,P]-W presentes nas duas alças paralelas. A interação com o sítio ativo das proteases alvo se dá de forma direta e reversível, bloqueando temporariamente o acesso ao substrato (BODE et al., 1988). As cistatinas que também inibem proteases semelhantes a legumaina possuem um segundo sítio inibitório conservado, caracterizado pela sequência S-N-S-[L,I]. Na cistatina E/M humana, esse segundo sítio antecede a alfa-hélice, enquanto que nas cistatinas de plantas está presente em uma porção carboxi-terminal estendida

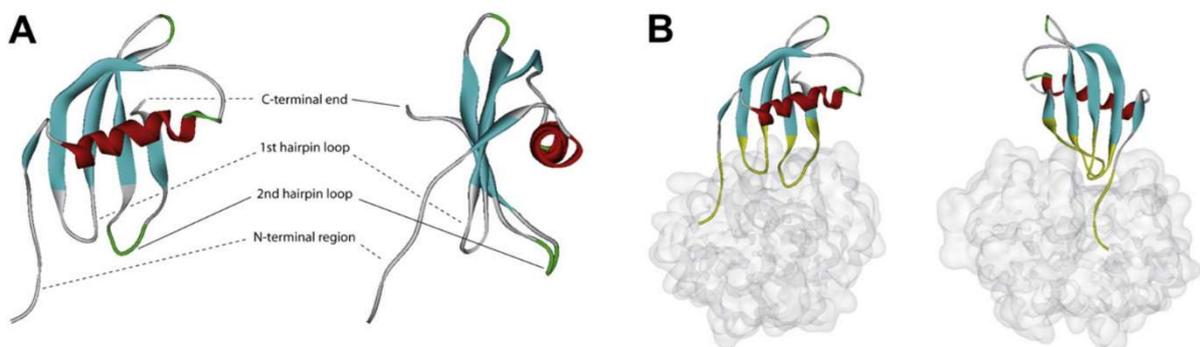


FIGURA 4 – Modelos estruturais da cistatina de arroz. (a) Vistas lateral e frontal da orizacistatina, evidenciando as folhas beta antiparalelas (azul), a alfa-hélice (em vermelho) e as duas alças inibitórias (em verde). (b) Vistas lateral e posterior de um modelo de complexação da orizacistatina com a papaína (cinza claro). Figura retirada de Benchabane et al. (2010)

(ALVAREZ-FERNANDEZ et al., 1999; CHRISTOFF et al., 2016; MARTINEZ et al., 2007).

Desde a primeira caracterização da cistatina na clara do ovo de galinha (FOSSUM; WHITAKER, 1968), diversos membros da superfamília foram descritos em bactérias, oomicetos, vírus, animais e plantas. Abe et al. (1987) identificaram a primeira cistatina vegetal a em sementes de arroz e, desde então, os membros dessa família de genes foram caracterizados em muitas espécies de plantas (ABRAHAM et al., 2006; CARDOSO et al., 2015; CHRISTOFF; MARGIS, 2014; MARTINEZ et al., 2005; PIROVANI et al., 2010; REIS; MARGIS, 2001; SCHNEIDER et al., 2020; SHIBAO et al., 2021; SUBBURAJ et al., 2017; ZHAO et al., 2014). Análises comparativas das características estruturais das cistatinas vegetais identificaram o motivo [L,V,I]-[A,G,T]-[R,K,E]-[F,Y]-[A,S]-[V,I]-[E,D,Q,V]-[H,Y,F,Q]-N na porção amino-terminal da alfa hélice, que é exclusivo das cistatinas de plantas. A diversidade intrínseca desses inibidores em plantas e o agrupamento das cistatinas vegetais em um ramo evolutivo independente das cistatinas e outros eucariotos levaram ao estabelecimento da família das fitocistatinas (MARGIS; REIS; VILLERET, 1998; MARTINEZ; DIAZ, 2008).

1.4 FITOCISTATINAS

As fitocistatinas são comumente subdivididas em três tipos de acordo com seu peso molecular (WANG et al., 2008). O tipo I (PhyCys-I, Figura 5A) caracteriza inibidores com peso molecular de 12-16 kDa e apenas um domínio inibitório para papaínas, que compreende a maioria das fitocistatinas. O tipo II (PhyCys-II, Figura 5B) tem um peso molecular de aproximadamente 23 kDa. Além do domínio inibitório para papaínas na região amino-terminal, eles possuem uma extensão carboxi-terminal com o motivo característico para a inibição de legumaínas. Essa estrutura carboxi-estendida característica das PhyCys-II é exclusiva das plantas. As fitocistatinas do tipo III (PhyCys-III, Figura 5C) também são conhecidas como multicistatinas devido à presença de múltiplos domínios cistatina. As proteínas classificadas como fitocistatinas do tipo III têm mais de 80 kDa e foram descritas em

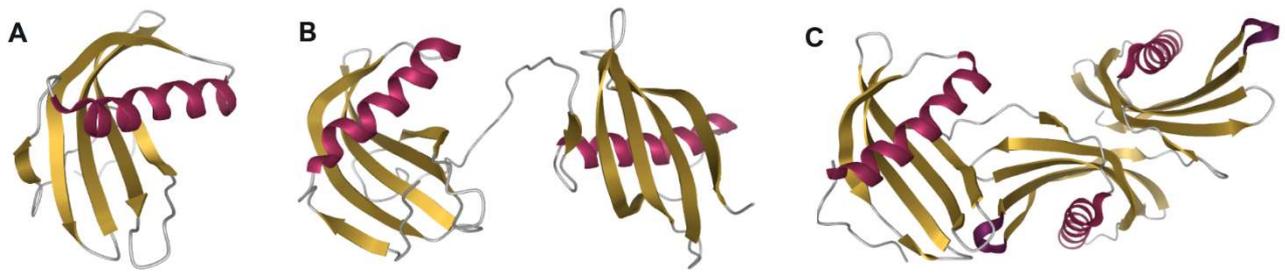


FIGURA 5 – Modelos estruturais dos diferentes tipos de fitocistatinas. (a) Fitocistatina do tipo I de *Oryza sativa* (1EQK). (b) Fitocistatina do tipo II de *Sesamum indicum* (2MZV). (c) Fitocistatina do tipo II de *Solanum tuberosum* (4LZI).

feijão-de-corda (DIOP et al., 2004), tomate (GOULET et al., 2008), cacau (PIROVANI et al., 2010), girassol (GHOLIZADEH; KOHNEHROUZ, 2011) e batata (GREEN et al., 2013).

Nas últimas décadas, avanços significativos sobre o papel das fitocistatinas nos processos celulares e fisiológicos das plantas contribuíram para o melhor entendimento dos seus efeitos *in planta*. Sabe-se que as fitocistatinas atuam no controle da proteólise em diversos estágios de desenvolvimento, tais como na maturação de sementes, germinação e crescimento de plântulas (SONG et al., 2017; ZHAO et al., 2014). Algumas fitocistatinas apresentam aumento nos níveis de expressão em resposta a ácido abscísico (ABA), um fitormônio envolvido na manutenção da dormência de sementes e capaz de retardar a germinação (HWANG et al., 2009; SONG et al., 2017; SUBBURAJ et al., 2017; ZHAO et al., 2014). Tal observação vai de encontro com o papel das fitocistatinas na regulação do *turnover* proteico através da inibição da atividade de proteases cisteínicas.

A proteólise é uma resposta essencial em condições de estresses bióticos e abióticos. A fim de evitar a degradação desenfreada de proteínas essenciais, é necessária a modulação da expressão de inibidores de proteases. Diversos trabalhos reportam o aumento da expressão de cistatinas em resposta à seca, salinidade, estresse oxidativo e estresse térmico (DIOP et al., 2004; MASSONNEAU et al., 2005; SUBBURAJ et al., 2017; VELASCO-ARROYO et al., 2018). Tan et al. (2017) reportaram que a superexpressão da cistatina *MpCYS4* da macieira aumentou a tolerância de plantas transgênicas à seca, melhorando o fechamento dos estômatos e aumentando os níveis transcricionais de genes relacionados à

resposta a seca e ABA. Resultados semelhantes foram encontrados com a superexpressão da orizacistatina I (Ocl) de arroz em soja e *Arabidopsis*, com plantas transgênicas menos suscetíveis à seca e indução da expressão de genes codificantes para enzimas envolvidas na síntese de estrigolactona (QUAIN et al., 2014).

Além do papel na resposta a estresses abióticos, as cistatinas também atuam na inibição de proteases exógenas de fitófagos e patógenos, contribuindo para a resposta da planta a ataques e infecções (GOULET et al., 2008; LIMA et al., 2020; PIROVANI et al., 2010; SANTANA et al., 2014; YANG; YEH, 2005). Plantas de tomate expressando uma cistatina de trigo, *TaMDC1*, apresentaram elevada resistência contra infecções bacterianas (*Pseudomonas syringae*), fúngicas (*Botrytis cinerea* e *Alternaria alternata*) e ataques de insetos herbívoros (besouro-da-batata, *Leptinotarsa decemlineata*) (CHRISTOVA et al., 2018). Em trigo, a expressão de *TaMDC1* é aumentada em resposta a metil-jasmonato e ácido salicílico, sugerindo a atividade dessa cistatina na resposta a estresse biótico mediada por tais hormônios vegetais.

Devido ao papel importante das fitocistatinas no desenvolvimento vegetal e frente a diversos estresses, sua utilização no melhoramento vegetal é uma possibilidade para o desenvolvimento de plantas mais resistentes. Diversos trabalhos exploram seus potenciais biotecnológicos, incluindo aplicações biomédicas e o mutações direcionadas para melhorar a performance contra patógenos (CHRISTOVA et al., 2018; LEGUIZAMÓN et al., 2019; MELO et al., 2019; ROBERTS, 2015). Em um trabalho inovador, Santiago et al. (2017) reportaram que uma cistatina de cana-de-açúcar, *CaneCPI-5*, se liga ao esmalte dentário e confere proteção ao reduzir significativamente a erosão inicial do mesmo, abrindo possibilidades para o uso de uma fitocistatina em formulações de cremes dentais.

Trabalhos anteriores com foco nas relações evolutivas entre as cistatinas de diversos organismos fornecem uma compreensão clara de uma origem comum para as cistatinas presentes em animais, bactérias e plantas (KORDIŠ; TURK, 2009). As características únicas e a diversidade das fitocistatinas apoiam fortemente a ideia de evolução paralela das cistatinas animais e vegetais (MARGIS-PINHEIRO et al.,

2008; MARTINEZ; DIAZ, 2008). As fitocistatinas parecem ter experimentado um longo e contínuo processo de diversificação, envolvendo duplicações e diversificações gênicas que levam à diversidade atual de inibidores desta família.

A disponibilidade de um número cada vez maior de genomas de plantas sequenciados permite expandir o conhecimento atual sobre a evolução das fitocistatinas, incluindo muitas linhagens de Viridiplantae pouco exploradas anteriormente. Devido à distribuição limitada de multicistatinas entre as espécies de plantas e sua descrição em outros trabalhos, nos concentramos na origem e evolução de PhyCys-I e PhyCys-II, buscando revelar uma visão geral sobre a diversidade atual de fitocistatinas e os processos evolutivos que levaram a elas.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Esse trabalho busca contribuir para o entendimento dos processos de emergência e diversificação dos membros da subfamília das fitocistatinas, com enfoque nas fitocistatinas do tipo II.

2.2 OBJETIVOS ESPECÍFICOS

2.2.1 Reconstruir as relações filogenéticas entre fitocistatinas do tipo I e II pertencentes a Viridiplantae;

2.2.2 Reconstruir as relações filogenéticas das fitocistatinas do tipo II, expandindo o conjunto de dados ao analisar sequências de um maior número de espécies;

2.2.3 Verificar a ocorrência de seleção positiva em fitocistatinas do tipo II e sua localização na cadeia polipeptídica.

3. MATERIAIS E MÉTODOS

3.1 LEVANTAMENTO EM BANCO DE DADOS

Inicialmente, realizou-se uma busca por sequências gênicas pertencentes à subfamília das fitocistatinas em genomas e transcriptomas de 188 espécies de Viridiplantae disponíveis em bancos de dados públicos e dados próprios do LGPP (Apêndices 1 e 2). A busca foi feita por *blastp* utilizando como isca as sequências proteicas das fitocistatinas de arroz (Ocl a OcXII), cuja caracterização e análise do padrão de expressão já foram feitos anteriormente (CHRISTOFF; MARGIS, 2014). Para a busca por sequências homólogas de fitocistatinas do tipo II, uma porção de 58 aminoácidos correspondente ao domínio inibitório para legumainas foi usada como isca (Apêndice 3). Essa abordagem possibilitou a obtenção de um maior número de sequências de espécies pertencentes a clados filogeneticamente distantes de *Oryza sativa*, uma vez que o uso da sequência completa restringiu os resultados da busca. Para todas as buscas, um ponto de corte de 10^{-5} para o valor de E foi estabelecido. Além das sequências de aminoácidos, foram obtidas as sequências codificantes (CDS) correspondentes ao transcrito primário para o *locus* equivalente à sequência de aminoácidos homóloga e informações quanto à organização gênica da sequência genômica.

As sequências resultantes foram filtradas para eliminar sequências redundantes de uma mesma espécie e sequências incompletas que não possuísem os domínios inibitórios característicos das fitocistatinas, que poderiam corresponder a pseudogenes. Multicistatinas não foram incluídas nas análises por terem sua distribuição restrita a poucas espécies e por terem sido caracterizadas em outros estudos.

3.2 ALINHAMENTO E ANÁLISE FILOGENÉTICA

Alinhamentos múltiplos de sequências foram realizados para dois conjunto de dados distintos: PhyCys-II, para compreender as relações filogenéticas entre as fitocistatinas com capacidade inibitória de dois tipos de proteases; PhyCys-I

juntamente com a porção amino-terminal das PhyCys-II, contendo os domínios inibitórios para proteases do tipo papaína. Esta segunda abordagem busca entender como ocorreram os processos evolutivos na subfamília das fitocistatinas como um todo.

Os alinhamentos foram feitos usando o programa MUSCLE (EDGAR, 2004) implementado no software MEGA7 (KUMAR; STECHER; TAMURA, 2016) de acordo com as sequências de amino ácidos, também sendo gerado um alinhamento de nucleotídeos correspondente. Ajustes manuais foram feitos aos alinhamentos para remover porções ricas em *gaps*, mantendo porções importantes para a estrutura tridimensional já conhecida das proteínas.

Os modelos de substituição de nucleotídeos que melhor representam os conjuntos de dados foram calculados com o programa ModelTest-NG (DARRIBA et al., 2020). Ferramentas como a utilizada calculam a verossimilhança de modelos candidatos, estimam os parâmetros de um dado alinhamento e geram informações sobre os parâmetros a serem utilizados para as reconstruções filogenéticas.

Para estimar a história evolutiva da família, uma abordagem de inferência filogenética bayesiana foi realizada com o programa BEAST 2.6 (BOUCKAERT et al., 2019), utilizando os parâmetros definidos no apêndice 4. Este tipo de análise se fundamenta em informações *a priori* sobre o conjunto de dados para identificar a árvore filogenética com maior probabilidade *a posteriori* dentre as árvores possíveis. As distribuições posteriores dos parâmetros foram estimadas por meio de amostragem de Cadeias Markovianas acopladas ao método Monte Carlo (*Markov chain Monte Carlo* - MCMC), com duas execuções independentes para cada conjunto de dados, as quais foram analisadas quanto à convergência e combinadas. O software Tracer 1.7.1 (RAMBAUT et al., 2018) foi usado para avaliar a convergência das execuções e certificar que os tamanhos efetivos de amostra (*Effective Sample Sizes* - ESS) para todos os parâmetros foram suficientes (> 200). As árvores das execuções independentes foram combinadas com a ferramenta LogCombiner, descartando 20% das árvores iniciais, e uma árvore de credibilidade de clado máxima foi gerada com TreeAnnotator. As filogenias resultantes foram visualizadas e formatadas usando a ferramenta FigTree 1.4.4 (RAMBAUT, 2014), e

posteriormente editadas com o software Adobe Illustrator CC 2018 para a confecção das figuras.

3.3 CONSERVAÇÃO DE SEQUÊNCIA

O nível de conservação da posição de um aminoácido na sequência de uma proteína é fortemente dependente da sua importância funcional ou estrutural. Dessa forma, analisar o nível de conservação dos resíduos de aminoácidos que compõem as sequências de PhyCys pode revelar a importância de posições específicas. Para estimar a conservação ao longo da sequência de PhyCys-I e PhyCys-II foi utilizada a ferramenta ConSurf (ASHKENAZY et al., 2016). A partir de um alinhamento de sequências homólogas, *scores* de conservação são computados usando algoritmos de máxima verossimilhança e *empirical Bayes*. Os *scores* de conservação são divididos em uma escala de coloração das posições mais variáveis às mais conservadas, cujas cores são projetadas no alinhamento múltiplo de sequências e na estrutura tridimensional da proteína.

3.4 ANÁLISES DA PRESSÃO DE SELEÇÃO POSITIVA EM FITOCISTATINAS DO TIPO II

Diferentes posições da sequência de aminoácidos apresentam frequências de mutações não-sinônimas e sinônimas distintas, cuja razão de mutações não sinônimas e sinônimas é um indicativo de presença ou não de seleção. Um método para testar por seleção é estimando a razão de substituições não-sinônimas e sinônimas (dN/dS , ou ω). Sob a hipótese de evolução neutra, é esperado que as taxas de mutações sinônimas e não-sinônimas sejam equivalentes e, dessa forma, é esperado que ω seja igual a 1. Seleção positiva e purificadora são indicadas quando $\omega > 1$ e $\omega < 1$, respectivamente.

Para determinar a pressão de seleção sobre os genes que codificam para fitocistatinas do tipo II, utilizou-se o programa *codeml* do pacote PAML (YANG, 2007), implementado na ferramenta ETE Toolkit (HUERTA-CEPAS; SERRA; BORK, 2016). Essa metodologia utiliza cálculos de máxima verossimilhança para estimar parâmetros de modelos de substituição de códons (Tabela 1) utilizando como base

o alinhamento de nucleotídeos de PhyCys-II e a topologia da árvore obtida com a análise bayesiana.

Site models (“modelos de sítios”, numa tradução literal) que permitem que os valores de ω variem entre os códons foram utilizados, além de *branch-site models* (“modelos de ramos e sítios”, numa tradução literal), que permitem que ω varie tanto entre sítios na proteína como entre ramos da árvore e buscam detectar seleção positiva afetando alguns sítios em linhagens específicas pré-definidas.

Os valores de verossimilhança obtidos para cada modelo foram utilizados para verificar a presença de seleção positiva por meio de testes da razão de verossimilhança (*likelihood ratio test*, $LRT = 2(\ln L_1 - \ln L_0)$). Foram realizadas comparações entre modelos aninhados, isto é, modelos com variáveis independentes hierarquizadas de acordo com o indicado na Tabela 2. As comparações foram feitas através de teste qui-quadrado considerando um $p \leq 0,05$ para rejeição da hipótese nula.

TABELA 1 – Modelos de códons empregados para o teste de seleção positiva nos sítios. P indica o número de parâmetros livres na distribuição ω para cada modelo.

Modelo	Parâmetros	Descrição
M0 <i>one-ratio</i>	ω	Assume uma taxa ω para todos os sítios
M1a <i>neutral</i>	p_0 ($p_1 = 1 - p_0$) $\omega_0 < 1$, $\omega_1 = 1$	Assume duas classes de sítios, conservados (ω_0) e neutros (ω_1), com frequências distintas (p_0 e p_1)
M2a <i>positive selection</i>	p_0, p_1 ($p_2 = 1 - p_0 - p_1$) $\omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$	Assume três classes de sítios: conservados (ω_0), neutros (ω_1) e positivamente selecionados (ω_2), com frequências distintas (p_0, p_1, p_2)
M3 <i>discrete</i>	p_0, p_1 ($p_2 = 1 - p_0 - p_1$) $\omega_0, \omega_1, \omega_2$	Assume uma distribuição discreta para modelar ω heterogêneos entre os sítios
M7 β	p, q	Assume uma distribuição $\beta(p,q)$ limitada num intervalo (0,1)
M8 $\beta\&\omega$	p_0 ($p_1 = 1 - p_0$) $p, q, \omega_s > 1$	Assume uma classe a mais que M7 amostrada da distribuição β , com ω constante com frequência p_0

TABELA 2 – Comparações de modelos aninhados para testes de hipótese de seleção positiva.

Modelo alternativo	Modelo nulo	Descrição
M3	M0	Testa para variabilidade entre os sítios do alinhamento
M2a	M1a	Testa para seleção positiva nos sítios (sujeito a não detectar alguns sítios)
M8	M7	Testa para seleção positiva nos sítios e levando em conta a variação em ω
bsA	bsA1	Testa para seleção positiva em sítios de um ramo específico da filogenia

4. RESULTADOS E DISCUSSÃO

4.1 PRESENÇA DE FITOCISTATINAS EM VIRIDIPLANTAE E A FILOGENIA DE FITOCISTATINAS

A busca por sequências de genes da família das fitocistatinas resultou em 298 fitocistatinas do tipo I pertencentes a 54 espécies das principais linhagens de plantas. Essas, unidas a 78 sequências de fitocistatinas do tipo II das mesmas espécies compuseram o conjunto de dados para avaliar a história evolutiva da subfamília como um todo. Para explorar a evolução específica das fitocistatinas do tipo II, a análise foi expandida para 187 espécies de plantas, resultando em 290 sequências.

Os dados evidenciam um número maior de cópias por espécie de fitocistatinas do tipo I em comparação com o tipo II em Viridiplantae, com o tipo I variando de 0 a 19 e o tipo II variando de 0 a 4 (Figura 7). Além disso, verificou-se uma tendência para o aumento no número de cópias de genes codificantes para fitocistatinas em Gimnopermas e Angiospermas. Essa tendência já havia sido descrita por Martinez e Diaz (2008) e especula-se que esteja correlacionada com o aumento do número de cópias das proteases cisteínicas alvo nos genomas das plantas, embora adaptações envolvendo interações planta-patógeno e plantas-predadores também possam estar envolvidas.

As buscas realizadas não permitiram identificar nenhum homólogo de cistatina em espécies de algas próximas a Viridiplantae, como Rhodophyta e Glaucophyta. Quanto aos principais clados de algas pertencentes à Viridiplantae, a presença de fitocistatinas tipo I e tipo II varia de acordo com a classe a que pertencem. Em Chlorophyta, as espécies pertencentes à Chlamydomonadales possuem apenas fitocistatinas do tipo I, enquanto Trebuxiophyceae e Sphaeropleales possuem apenas fitocistatinas do tipo II. As espécies de Charophyta, o clado de algas que compartilha um ancestral comum com as plantas terrestres, apresentam os dois tipos de fitocistatinas (Klebsormidiales e Charales) ou apenas fitocistatinas do tipo II (Zygnematales e Spirogloaeales). A ausência de

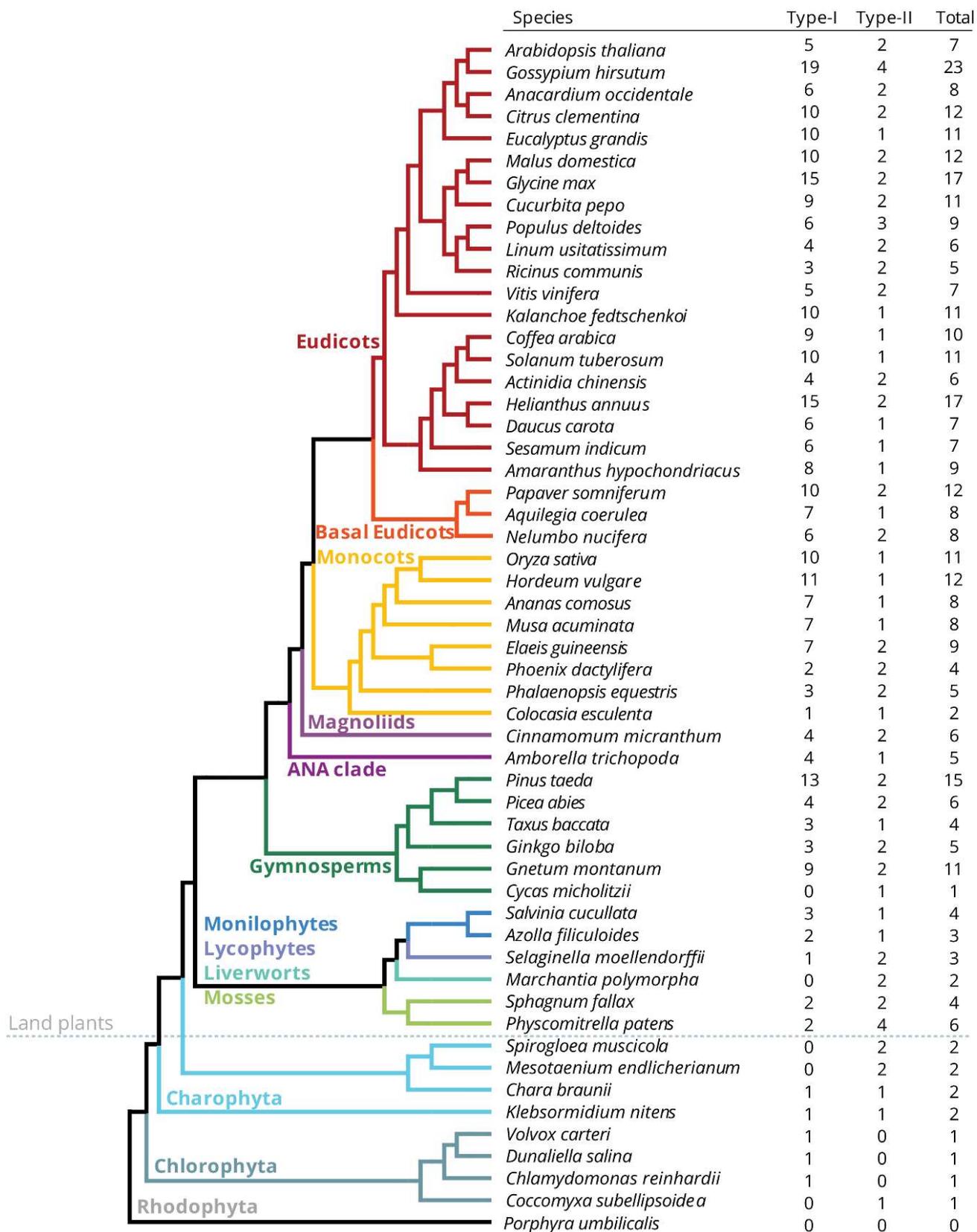


FIGURA 6 - Cladograma exibindo relações de Viridiplantae e o número correspondente de fitocistatinas tipo I e tipo II por espécie. Os ramos estão coloridos de acordo com o clado a que as espécies pertencem.

grupos externos adequados impossibilita uma reconstrução filogenética enraizada.

A similaridade estrutural entre os dois domínios de fitocistatinas do tipo II é uma forte evidência de que esta forma carboxi-estendida se originou de uma duplicação em tandem de um único domínio fitocistatina (MARGIS; REIS; VILLERET, 1998; MARTINEZ; DIAZ, 2008). A partir da análise filogenética da subfamília, podemos rastrear esse evento até o ancestral comum mais recente (MRCA, do inglês *most recent common ancestor*) de Viridiplantae, uma vez que as fitocistatinas que possuem uma extensão carboxi-terminal estão presentes tanto em Chlorophyta quanto em Carophyta. O agrupamento de fitocistatinas tipo I e tipo II de *Coccomyxa subellipsoidea* e *Klebsormidium nitens* sugere que ambas as formas carboxi-estendidas se originaram de um único evento ancestral de duplicação em tandem (Figuras 7 e 8). Nesse sentido, a ausência de fitocistatinas do tipo II em Chlamydomonadales, mas a presença em Sphaeropleales, pode ser devido à perda da fitocistatina carboxi-estendida no primeiro clado. Além disso, o agrupamento de fitocistatinas do tipo I de Charophyta com fitocistatinas tipo II do mesmo clado e de Chlorophyta sugere que essas fitocistatinas tipo I podem ter surgido da perda da porção carboxi-terminal correspondente ao segundo domínio fitocistatina. O mesmo padrão é visível em outros cladogramas, como o ramo bem definido de Monilophyta e musgos (Funariales e Sphagnales). Nas gimnospermas, poucas sequências de *Ginkgo biloba* e *Pinus taeda* se agrupam fora do clado das gimnospermas, indicando um possível processo de diversificação dessas sequências que as separa de outras espécies de gimnospermas. É interessante pontuar que genomas de gimnospermas passaram por processos de expansão.

O número de cópias de fitocistatinas do tipo II em musgos, hepáticas e licófitas é igual ou maior do que as fitocistatinas do tipo I. *Marchantia polymorpha*, por exemplo, possui apenas fitocistatinas do tipo II. Em monilófitas, gimnospermas e angiospermas, o número de cópias das fitocistatinas do tipo I supera o tipo II. Uma expansão acentuada da família de genes é visível em gimnospermas e angiospermas, principalmente devido a um aumento no número de fitocistatinas do tipo I. *Gossypium hirsutum* é a espécie com maior número de cópias de fitocistatinas, com 19 cópias de fitocistatinas do tipo I e 4 cópias de fitocistatinas do

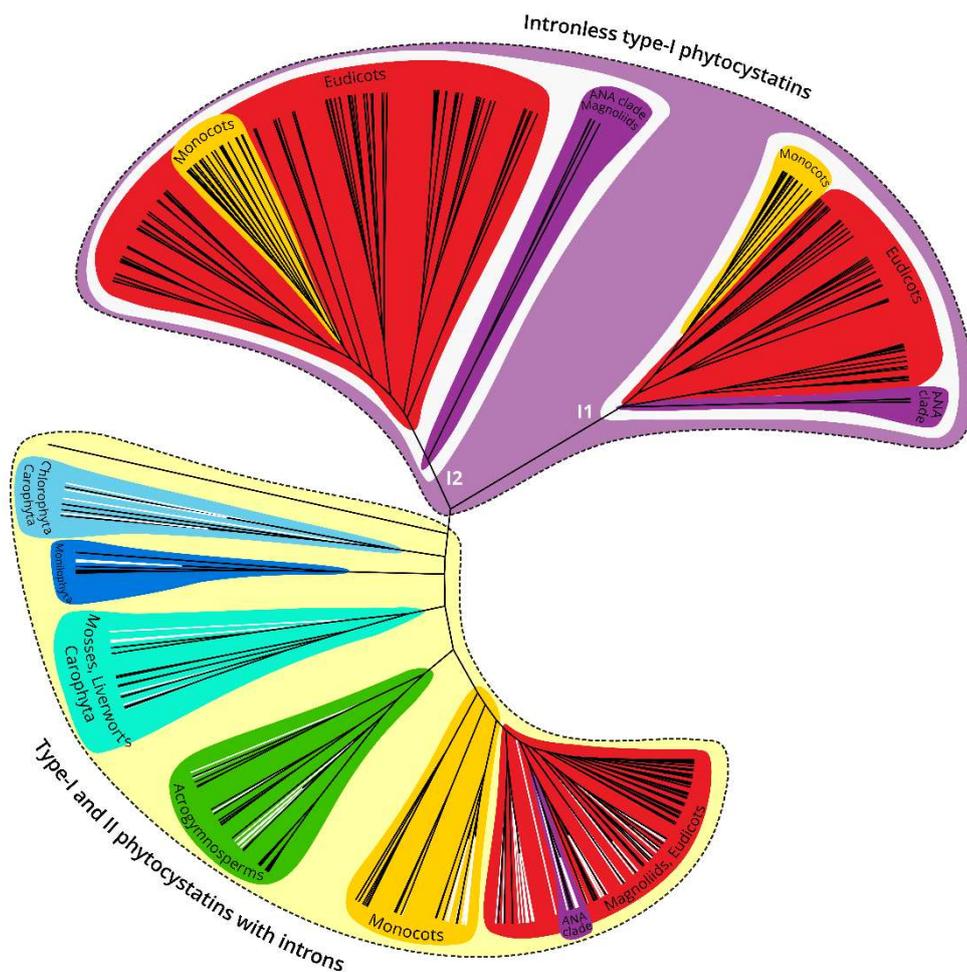


FIGURA 7 - Representação gráfica das relações entre seqüências nucleotídicas de fitocistatinas resultante de uma reconstrução filogenética bayesiana não-enraizada (em detalhes na figura 3). Análise feita com base em seqüências completas de fitocistatinas do tipo I e a região amino-terminal de fitocistatinas do tipo II. Os ramos brancos correspondem às fitocistatinas do tipo II. I1 e I2 indicam grupos distintos de fitocistatinas do tipo I sem íntrons.

tipo II, o que pode estar relacionado à sua natureza alotetraploide. Essa expansão da família gênica nas angiospermas tem sido correlacionada ao aumento do número de cópias de proteases semelhantes à papaína nas angiospermas (MARTINEZ; DIAZ, 2008), sugerindo um processo de coevolução de protease alvo e inibidor.

Conforme descrito anteriormente em caracterizações da família gênica em diferentes espécies (CHRISTOFF; MARGIS, 2014; MARGIS-PINHEIRO et al., 2008), a organização gênica dos membros da família da fitocistatina é bastante conservada, com poucas variações em número e tamanho de éxons e íntrons. Na presente análise demonstra-se que as fitocistatinas do tipo II quase sempre mantêm uma

organização gênica composta por quatro éxons e três íntrons. Os dois primeiros éxons geralmente correspondem à porção amino-terminal contendo o domínio inibitório para papaínas e os dois últimos correspondem à porção carboxi-terminal contendo o domínio inibitório para legumainas. Já as fitocistatinas do tipo I podem ser sem íntron (predominantemente presentes em plantas com flores) ou ter dois ou mais exons, com a estrutura composta de dois éxons os mais comuns em Viridiplantae.

Ao analisar a história evolutiva de toda a subfamília das fitocistatinas, observa-se a formação de dois grupos principais: um composto por fitocistatinas tipo I e tipo II de espécies de Viridiplantae e outro composto por fitocistatinas tipo I de angiospermas (Figuras 7 e 8). Uma análise mais detalhada da filogenia revela que o grupo contendo fitocistatinas do tipo II também é composto por fitocistatinas do tipo I, mas somente por aquelas que possui dois ou mais éxons. Em contraste, o segundo grupo é composto principalmente de genes de espécies de angiospermas compostos apenas de um éxon (sem íntrons), com poucas sequências que possuem um íntron e nenhuma que tenha mais de um íntron. O referido grupo de fitocistatinas do tipo I sem íntrons ainda pode ser subdividido em dois agrupamentos, I1 e I2 (do inglês, *intronless* 1 e 2).

Diversas espécies de plantas com flores estão presentes em ambos os agrupamentos, como *Arabidopsis thaliana*, *Oryza sativa*, *Aquilegia coerulea* e *Amborella trichopoda*. Devido à distribuição limitada de PhyCys-I sem íntrons em Viridiplantae e o agrupamento obtido na filogenia, sugere-se que as PhyCys-I sem íntrons tenham se originado no ancestral comum mais recente (MRCA) das plantas com flores, a partir de um gene ancestral com pelo menos um íntron, possivelmente a partir de um mecanismo de retroduplicação. A formação de dois grupos distintos sugere ainda que, após a retroduplicação, um novo evento de duplicação deu origem a duas cópias da forma sem íntrons, uma vez que os clados e espécies representados em I1 também estão representados em I2. É possível que esses eventos de duplicação estejam associados ao evento de WGD ϵ , datado durante o Permiano e compartilhado por todas as plantas com flores (CLARK; DONOGHUE, 2018). Posteriormente, essas sequências teriam passado por um processo de

diversificação e seleção que levou à sua manutenção nas espécies e à formação de dois grupos distintos na filogenia.

O grupo contendo sequências de PhyCys-I e PhyCysII com íntrons exibe uma organização de sequências de acordo com a relação filogenética das espécies de plantas, de um modo geral. Algumas sequências de espécies de Chlorophyta,

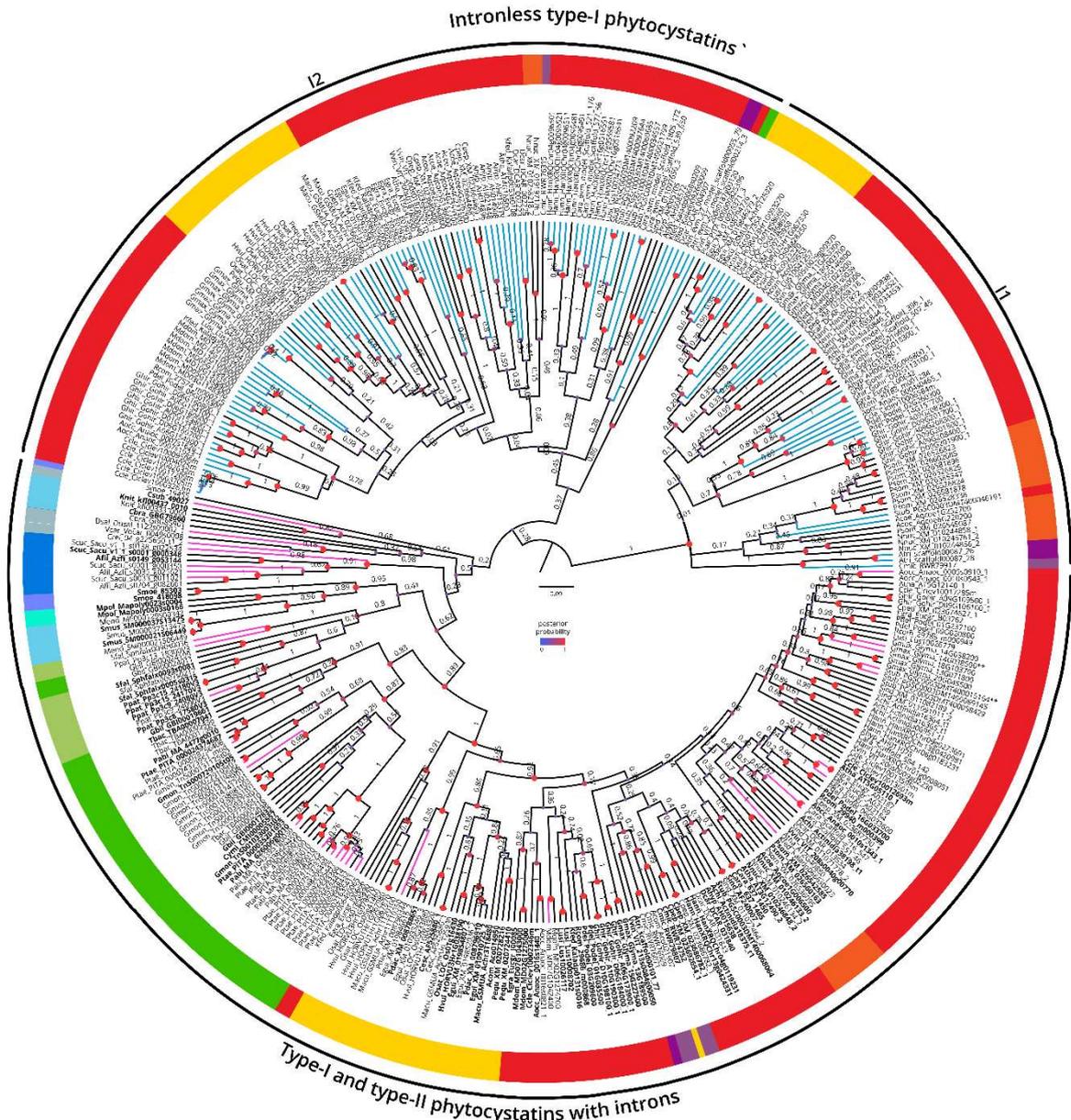


FIGURA 8 - Reconstrução filogenética bayesiana não-enraizada das relações da subfamília de fitocistatina com base em sequências completas do tipo I e a região amino-terminal das fitocistatinas do tipo II (OTUs em negrito), com base em sequências nucleotídicas. O círculo externo é colorido de acordo com o clado. Os ramos são coloridos de acordo com o número de éxons que compõe a porção utilizada do gene como segue: azul, um éxon; preto, dois éxons; rosa, três ou mais éxons. I1 e I2 indicam grupos distintos de fitocistatinas do tipo I sem íntrons.

gimnospermas e monocotiledôneas não agrupam com outras sequências do mesmo clado. Curiosamente, a região amino-terminal das PhyCys-II não forma um grupo isolado das demais, como se esperaria se tivessem se originado de um único evento de duplicação e se diversificado desde então. Uma vez que sequências de PhyCys-II se encontram num mesmo grupo que PhyCys-I de espécies do mesmo clado, pode ser argumentado que as PhyCys-I com íntrons tiveram sua origem nas PhyCys-II, a partir de eventos de perda da extensão carboxi-terminal ao longo da evolução das plantas. Teriam ocorrido pelo menos um evento compartilhado por angiospermas, um por gimnospermas, um por Lycophytas, além de outros eventos independentes em espécies de Charophyta.

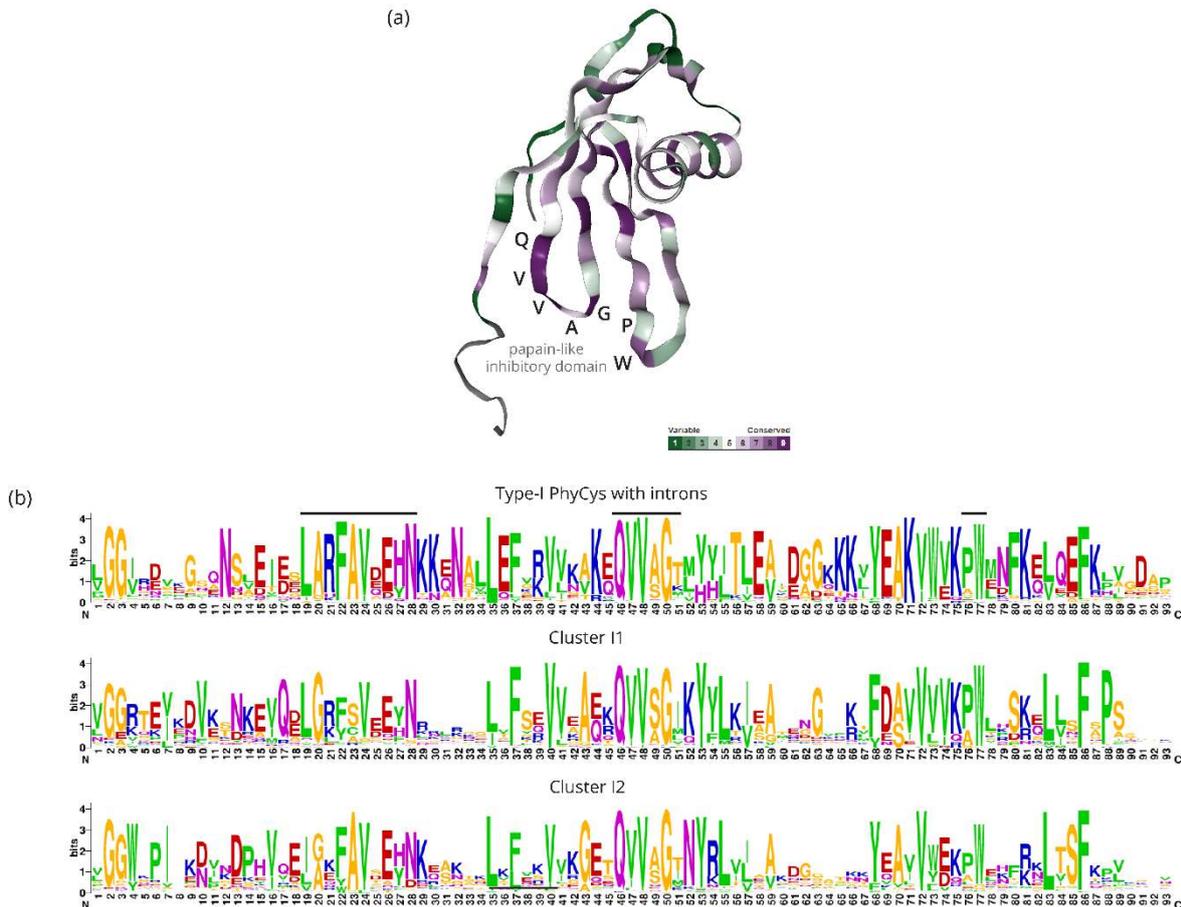


FIGURA 9 - Nível de conservação de sequência das fitocistatinas do tipo I. (a) Fitocistatina do tipo I de *Oryza sativa* (1EQK) com resíduos coloridos de acordo com o score de conservação em fitocistatinas tipo I (b) Sequências logo de fitocistatinas tipo I dos grupos distintos: fitocistatinas tipo I com íntrons, clusters sem íntron I1 e I2. Os resíduos estão coloridos em cinco grupos: pequenos não polares (laranja), hidrofóbicos (verde), polares (magenta), carregados negativamente (vermelho) e positivamente (azul)

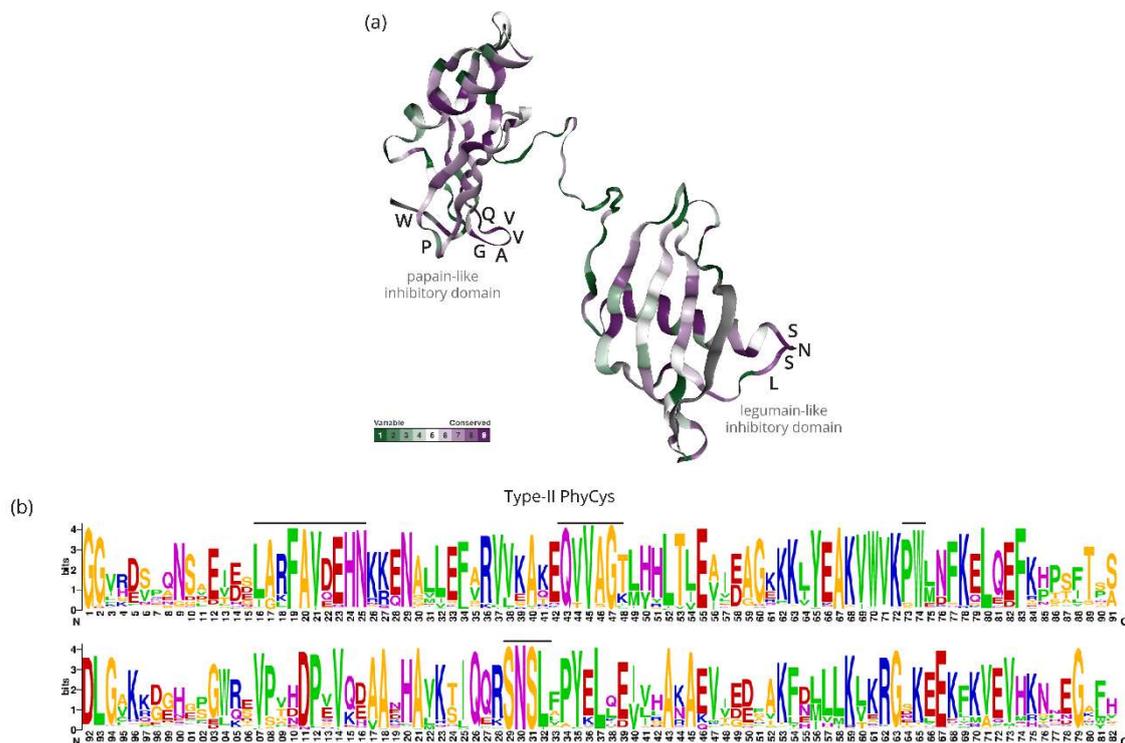


FIGURA 10 - Nível de conservação de sequência das fitocistatinas do tipo II. (a) Fitocistatina do tipo II de *Sesamum indicum* tipo II (2MZV) com resíduos coloridos de acordo com o score de conservação em fitocistatinas do tipo II. (d) Sequência logo de fitocistatinas do tipo II em Viridiplantae. Os resíduos estão coloridos em cinco grupos: pequenos não polares (laranja), hidrofóbicos (verde), polares (magenta), carregados negativamente (vermelho) e positivamente (azul)

Ao analisar a conservação da sequência de PhyCys-I e PhyCys-II, os resíduos que apresentam maior taxa de conservação estão localizados nos pontos de contato da alfa-hélice e das folhas beta, essenciais para a manutenção da dobra protéica característica, e nos sítios reativos envolvidos na inibição de proteases cisteínicas (Figuras 9 e 10). Os resíduos localizados na alfa-hélice e nas folhas beta são altamente conservados ou exibem substituições que não alteram significativamente a carga e/ou o tamanho do aminoácido. Curiosamente, o motivo LARFAVDEHN característico de cistatinas vegetais é altamente conservado em fitocistatinas tipo I com íntrons e na região amino-terminal de fitocistatinas tipo II de plantas com flores, mas não é mais conservado em fitocistatinas tipo I sem intron.

4.2 FILOGENIA DE FITOCISTATINAS TIPO II

A árvore obtida com PhyCys-II assemelha-se às relações evolutivas entre as espécies de plantas, de algas a angiospermas, com fitocistatinas tipo II presentes em muitas clorófitas (Figura 11). Além das sequências de duas espécies de

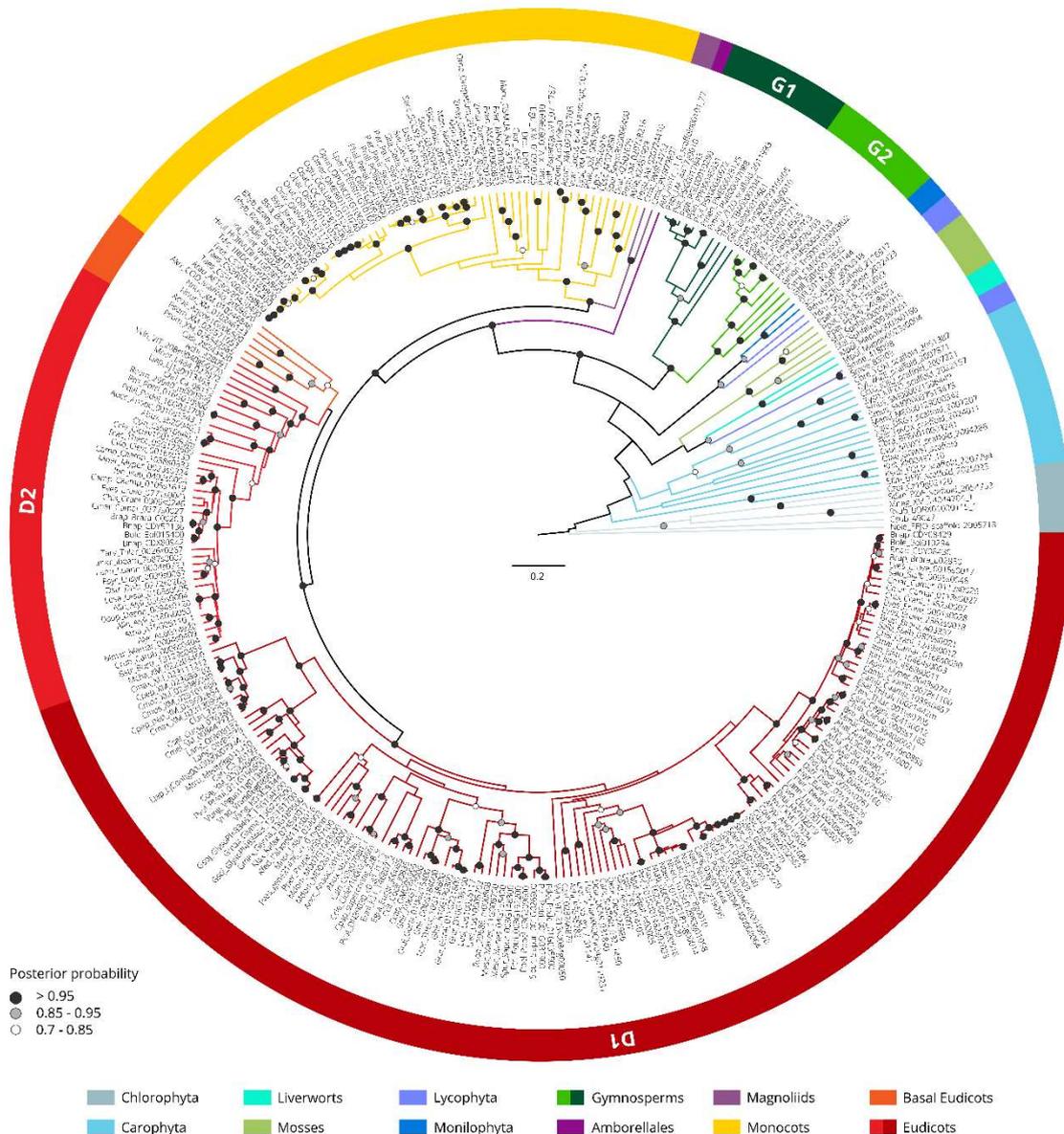


FIGURA 11 – Reconstrução filogenética bayesiana de fitocistatinas do tipo II. Os ramos estão coloridos de acordo com o clado que as OTUs pertencem (vide legenda). Probabilidades posteriores (PP) superiores a 0,7 estão indicadas como círculos no ramo correspondente, preenchidos de acordo com os valores de PP. D1 e D2 indicam as duplicações principais em eudicotiledôneas; G1 e G2 indicam as duplicações principais em gimnospermas.

Lycophyta que se agrupam com Monilophyta, todas as fitocistatinas do tipo II se agrupam com sequências dentro do mesmo clado.

Dois eventos de duplicação principais foram identificados: um nas Gimnospermas (G1 e G2) e outro nos Eudicots (D1 e D2). Uma hipótese proposta é que a duplicação em Gimnospermas ocorreu no ancestral comum das Gingkófitas, Pinófitas e Gnetófitas, uma vez que espécies desses clados estão presentes em ambos os ramos da filogenia. A análise mostra que as sequências de *Cycas micholitzii*, *Taxus baccata*, *Pinus sylvestris* e *Pinus taeda* estão presentes em apenas um ramo da filogenia. Isso pode ser um indicador de que uma segunda cópia, se presente nessas espécies, passou por um processo de pseudofuncionalização gênica ou perda gênica. No entanto, os resultados obtidos podem ser enviesados pelo uso de dados transcriptômicos e pela filtragem de sequências incompletas, e por isso não podemos excluir a possibilidade de que outra cópia da fitocistatina tipo II esteja presente nos genomas dessas espécies.

O ramo de duplicação de Eudicots é composto principalmente de espécies de rosídeas, mas também tem sequências de duas espécies de Ericales (asterídeas) e Vitales. Pode ser proposto que uma duplicação pode ter ocorrido no ancestral comum de rosídeas e Vitales, sendo posteriormente perdida em Myrtales e Cucurbitales. Outro evento de duplicação pode ter ocorrido em Ericales e os processos de diversificação e seleção atuando nessas proteínas levaram à convergência para similaridades na sequência e, possivelmente, na função.

Além disso, outros eventos de duplicação de fitocistatinas do tipo II ocorreram mais recentemente nos níveis de gênero e espécie e podem estar associados a duplicações completas do genoma (como *G. hirsutum*, um alotetraploide) ou rearranjos cromossômicos (como os cromossomos 2 e 7 de *Malus domestica*).

Quando se examina mais de perto as diferenças nas sequências que formam dois ramos na filogenia das angiospermas, observa-se algumas substituições de aminoácidos que podem impactar a complexação de cistatinas com suas proteases alvo ou outras proteínas interatoras (Figura 10). Essas mudanças impactam a carga das cadeias laterais em posições específicas e podem

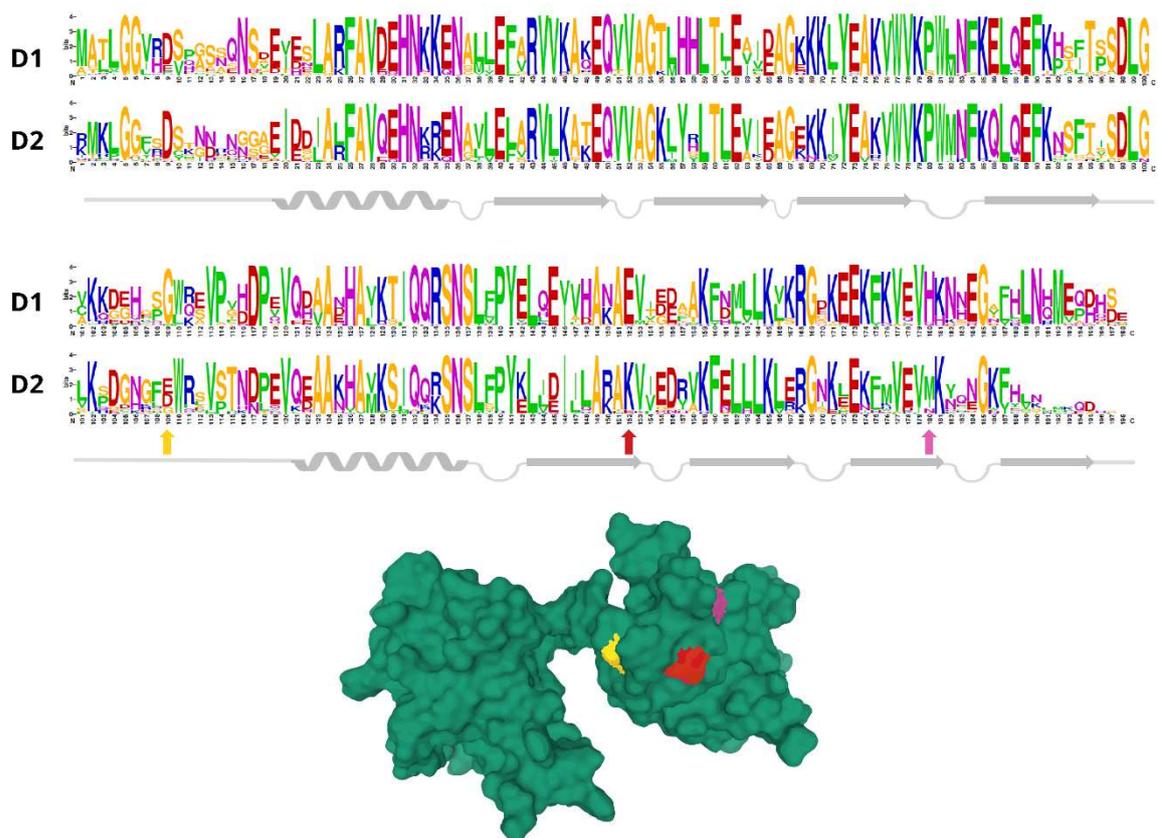


FIGURA 12 - Logo de seqüências de fitocistatinas do tipo II dos dois ramos de duplicação de Eudicots, D1 e D2. Os resíduos estão coloridos em cinco grupos: pequenos não-polares (laranja), hidrofóbicos (verde), polares (magenta), carregados negativamente (vermelho) e positivamente (azul). As setas indicam resíduos com mudanças de carga em ambas as seqüências. Na parte inferior, estrutura 3D da fitocistatina *Sesamum indicum* (2MZV), indicando a posição das alterações de carga visíveis no alinhamento.

afetar a carga líquida da proteína, como a substituição de uma glicina por resíduos eletricamente carregados de ácido (Figura 12, seta amarela), um glutamato por uma lisina (Figura 12, seta vermelha) e uma histidina para uma metionina e isoleucina (Figura 12, seta rosa).

4.3 ANÁLISE DE PRESSÃO DE SELEÇÃO

A filogenia obtida para as fitocistatinas do tipo II foi utilizada para os testes de modelos de seleção positiva e subsequentes testes da razão de verossimilhança (LRTs). Utilizando a filogenia completa das fitocistatinas do tipo II, a comparação entre modelos M3 vs M0 indica a existência de variabilidade entre as taxas de ω entre os sítios do alinhamento (Tabela 3). Entretanto, as comparações de modelos

indicam ausência de seleção positiva: a hipótese nula foi rejeitada no teste M2a vs M1a, mas não no teste M8 vs M7. Embora haja significância estatística, nenhum sítio foi identificado pelas análises de NEB e BEB com probabilidades posteriores superiores a 0,9. Esse resultado sugere que a diferença estatística encontrada é resultante de um falso positivo. Além disso, por conter sequências de organismos evolutivamente muito distantes, desde Chlorophyta até eudicotiledôneas, os sinais de seleção presentes podem ter sido amenizados e não detectados.

TABELA 3 - Resultados do *codeml* para a filogenia completa das fitocistatinas do tipo II. LnL = *log-likelihood*; LRT = *likelihood ratio test*; gl = graus de liberdade; BEB = *Bayes Empirical Bayes*.

Modelo	LnL	LRT (gl)	p	ω	Estimativas do modelo	Sítios sob seleção positiva (BEB>0,90)
M0	-49236.64	2538,9954	0	0,173	n/a	n/a
M3	-47967.14	(4)		0,193	$p_0=0,3017$; $p_1=0,4887$ $p_2= 0,2096$ $\omega_0=0,0440$; $\omega_1=0,1567$; $\omega_2=0,4911$	n/a
M1a	-48531,77	0	1,00	0,284	$p_0=0,8405$; $p_1=0,15945$ $\omega_0=0,1485$; $\omega_1=1,0000$	n/a
M2a	-48531,77	(2)		0,284	$p_0=0,8405$; $p_1=0,0801$; $p_2=0,0793$ $\omega_0=0,1485$; $\omega_1=1,0000$; $\omega_2=1,0000$	n/a
M7	-47843,98	23,3937	8^{-6}	0,246	$p=0,7928$; $q=2,3842$	n/a
M8	-47832,28	(2)		0,230	$p_0= 0,9737$; ($p_1=0,0263$) $p=0,9236$; $q=3,4180$; $\omega=1,0000$	n/a

A fim de reduzir a possível interferência de sequências muito distantes, uma nova filogenia apenas com as sequências de fitocistatinas do tipo II de eudicotiledôneas foi inferida (Apêndice 7) e utilizada para os testes de seleção positiva. Similarmente, nenhum sítio sob seleção positiva foi identificado, apesar da diferença estatística entre os modelos M7 e M8 (Tabela 4). Além disso, a busca por identificação de sítios sob seleção positiva no ramo correspondente à duplicação em eudicotiledôneas (D2) também não teve resultados estatisticamente significativos. Numa outra perspectiva, a elevada conservação a nível de sequência pode ser um indicativo de insuficiência de variação para que os métodos estatísticos empregados tenham poder de detecção de seleção, uma vez que uma elevada similaridade aumentaria a estimativa de dN em comparação a dS .

TABELA 4 - Resultados do *codeml* para a filogenia das fitocistatinas do tipo II de eudicotiledôneas. LnL = *log-likelihood*; LRT = *likelihood ratio test*; gl = graus de liberdade; BEB = *Bayes Empirical Bayes*.

Model	LnL	LRT (gl)	p	ω	Model estimates	Sítios sob seleção positiva (BEB>0,90)
M1a	-24024.30	2E-06	0,99	0,281	$p_0=0,82721$ $p_1=0,17279$ $\omega_0=0,13114$ $\omega_1=1,0000$	n/a
M2a	-24024.30			0,281	$p=0,82722$ $p_1=0,04963$ $p_2=0,12315$ $\omega_0=0,13114$ $\omega_1=1,0000$ $\omega_2=1,0000$	n/a
M7	-23734.15	31,2458	1E-07	0,236	$p=0,61719$ $q=1,96166$	n/a
M8	-23718.53			0,210	$p_0=0,94925$ ($p_1=0,05075$) $p=0,85412$ $q=4,12285$ $\omega=1,0000$	n/a
bsA	-21916,85	0 (2)	1,00		$\omega_0=0,11906$ (76,8%) $\omega_1=1,00000$ (11,7%) $\omega_{2a}=1,00000$ (10,0%) $\omega_{2b}=1,00000$ (1,5%)	n/a
bsA1	-21916,85					

4.4 FILOGENIA DAS FITOCISTATINAS

Diversos mecanismos e eventos levaram à diversidade de fitocistatinas presentes atualmente. Tomando o conjunto de resultados aqui apresentado, pode-se sumarizar a evolução das fitocistatinas em Viridiplanteae conforme representado na Figura 13. Inicialmente, um gene ancestral semelhante à cistatina presente no ancestral de Viridiplantae sofreu uma duplicação em tandem seguida de um processo de neofuncionalização, dando origem a uma forma duplicada de cistatina apenas presente nas plantas e com atividades inibitórias contra proteases dos tipos papaína e legumaína. Ao longo do processo evolutivo, diversos eventos de duplicação de genes levaram à expansão do número de cópias por espécie. Múltiplas perdas do segundo domínio de uma fitocistatina carboxi-extendida levaram à formação de fitocistatinas de domínio único com íntrons. Um processo de retroduplicação seguido de uma nova duplicação presente no MRCA de plantas com flores levou à formação da forma sem íntron de fitocistatinas e subsequente diversificação.

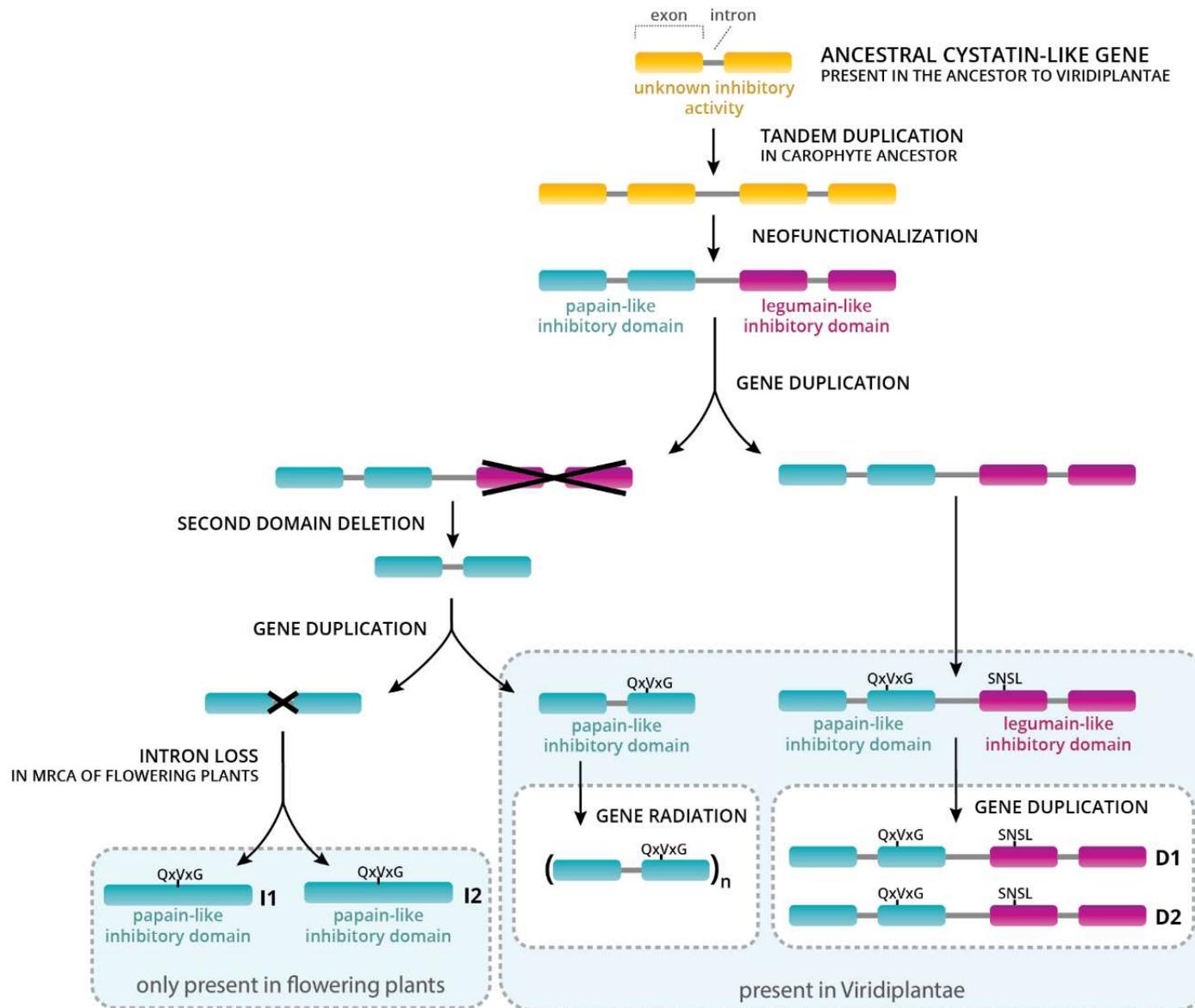


FIGURA 13 - Resumo da história evolutiva das fitocistatinas. Uma duplicação em tandem no gene ancestral semelhante à cistatina seguida pela neofuncionalização deu origem às fitocistatinas do tipo II. As duplicações gênicas seguidas pela deleção do domínio inibitório do tipo legumaina originaram as fitocistatinas do tipo I. Fitocistatinas tipo I e tipo II com íntrons estão presentes em Viridiplantae. No MRCA de plantas com flores, uma forma sem íntron de fitocistatinas tipo I originou dois grupos de fitocistatinas existentes, I1 e I2.

5. CONSIDERAÇÕES FINAIS E PERSPECTIVAS

O presente trabalho traz uma abordagem filogenômica para compreender as relações evolutivas de fitocistatinas, ampliando a compreensão dos processos envolvidos na emergência e evolução dos membros dessa família de genes em Viridiplantae. Através da reconstrução da filogenia da subfamília das fitocistatinas, identificamos diversos eventos de duplicação e diversificação que levaram à variedade de fitocistatinas presentes atualmente. Esses resultados servem como base para futuros estudos funcionais buscando elucidar a manutenção de múltiplas cópias de PhyCys-I e PhyCys-II e seus efeitos fisiológicos e alvos *in planta*, através de mutantes *knockout* para cada uma das cópias e metodologias proteômicas para a identificação de alvos específicos para cada inibidor.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABE, K. et al. Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds. **The Journal of biological chemistry**, [s. l.], v. 262, n. 35, p. 16793–16797, 1987.
- ABRAHAM, Z. et al. Structural and functional diversity within the cystatin gene family of *Hordeum vulgare*. **Journal of Experimental Botany**, [s. l.], v. 57, n. 15, p. 4245–4255, 2006.
- ALVAREZ-FERNANDEZ, M. et al. Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. **Journal of Biological Chemistry**, [s. l.], v. 274, n. 27, p. 19195–19203, 1999.
- ASHKENAZY, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic acids research**, [s. l.], v. 44, n. W1, p. W344–W350, 2016.
- BENCHABANE, M. et al. Plant cystatins. **Biochimie**, [s. l.], v. 92, n. 11, p. 1657–1666, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.biochi.2010.06.006>>
- BLANC, G.; WOLFE, K. H. Functional divergence of duplicated genes formed by polyploidy during arabidopsis evolution. **Plant Cell**, [s. l.], v. 16, n. 7, p. 1679–1691, 2004.
- BODE, W. et al. The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. **The EMBO journal**, [s. l.], v. 7, n. 8, p. 2593–2599, 1988.
- BOUCKAERT, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. **PLoS Computational Biology**, [s. l.], v. 15, n. 4, p. 1–28, 2019.
- CARDOSO, T. H. S. et al. TcCYPR04, a cacao papain-like cysteine-protease detected in senescent and necrotic tissues interacts with a cystatin TcCYS4. **PLoS ONE**, [s. l.], v. 10, n. 12, p. 1–20, 2015.
- CHRISTOFF, A. P. et al. Rice bifunctional phytocystatin is a dual modulator of legumain and papain-like proteases. **Plant Molecular Biology**, [s. l.], v. 92, n. 1–2, p. 193–207, 2016. Disponível em: <<http://dx.doi.org/10.1007/s11103-016-0504-5>>
- CHRISTOFF, A. P.; MARGIS, R. The diversity of rice phytocystatins. **Molecular Genetics and Genomics**, [s. l.], p. 1321–1330, 2014.
- CHRISTOVA, P. K. et al. The wheat multidomain cystatin TaMDC1 displays antifungal, antibacterial, and insecticidal activities in planta. **Plant Cell Reports**, [s. l.], v. 37, n. 6, p. 923–932, 2018.
- CHU, M. H. et al. Crystal structure of tarocystatin-papain complex: Implications for the inhibition property of group-2 phytocystatins. **Planta**, [s. l.], v. 234, n. 2, p. 243–

254, 2011.

CLARK, J. W.; DONOGHUE, P. C. J. Whole-Genome Duplication and Plant Macroevolution. **Trends in Plant Science**, [s. l.], v. 23, n. 10, p. 933–945, 2018. Disponível em: <<https://doi.org/10.1016/j.tplants.2018.07.006>>

DARRIBA, Di. et al. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. **Molecular Biology and Evolution**, [s. l.], v. 37, n. 1, p. 291–294, 2020.

DIOP, N. N. et al. A multicystatin is induced by drought-stress in cowpea (*Vigna unguiculata* (L.) Walp.) leaves. **FEBS Letters**, [s. l.], v. 577, n. 3, p. 545–550, 2004.

DRUMMOND, A. J. et al. Relaxed phylogenetics and dating with confidence. **PLoS Biology**, [s. l.], v. 4, n. 5, p. 699–710, 2006.

EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, [s. l.], v. 32, n. 5, p. 1792–1797, 2004.

FOSSUM, K.; WHITAKER, J. R. Ficin and papain inhibitor from chicken egg white. **Archives of Biochemistry and Biophysics**, [s. l.], v. 125, n. 1, p. 367–375, 1968.

GHOLIZADEH, A.; KOHNEHROUZ, B. **Functional fusion expression of sunflower multicystatin in E. coli and its comparison with a single domain cystatin** *Indian Journal of Biochemistry & Biophysics*. [s.l: s.n.].

GITZENDANNER, M. A. et al. **Plastome Phylogenetics: 30 Years of Inferences Into Plant Evolution**. 1. ed. [s.l.] : Elsevier Ltd., 2018. v. 85

GOULD, S. B.; WALLER, R. F.; MCFADDEN, G. I. Plastid evolution. **Annual Review of Plant Biology**, [s. l.], v. 59, p. 491–517, 2008.

GOULET, M. C. et al. Tailoring the specificity of a plant cystatin toward herbivorous insect digestive cysteine proteases by single mutations at positively selected amino acid sites. **Plant Physiology**, [s. l.], v. 146, n. 3, p. 1010–1019, 2008.

GRAUR, D.; LI, W.-H. **Fundamentals of Molecular Evolution**. 2nd. ed. [s.l: s.n.].

GREEN, A. R. et al. Characterization of *Solanum tuberosum* multicystatin and the significance of core domains. **Plant Cell**, [s. l.], v. 25, n. 12, p. 5043–5052, 2013.

GUO, Y. L. Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. **Plant Journal**, [s. l.], v. 73, n. 6, p. 941–951, 2013.

HU, Y. J. et al. Resonance assignments and secondary structure of a phytocystatin from *Sesamum indicum*. **Biomolecular NMR Assignments**, [s. l.], v. 9, n. 2, p. 309–311, 2015. Disponível em: <<http://dx.doi.org/10.1007/s12104-015-9598-y>>

HUERTA-CEPAS, J.; SERRA, F.; BORK, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. **Molecular Biology and Evolution**, [s. l.], v. 33, n. 6, p. 1635–1638, 2016.

HUGHES, A. I.; NEI, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. **Nature**, [s. l.], v. 335, n. 6186, p. 167–170, 1988.

HWANG, J. E. et al. Regulation of seed germination and seedling growth by an Arabidopsis phytocystatin isoform, AtCYS6. **Plant Cell Reports**, [s. l.], v. 28, n. 11, p. 1623–1632, 2009.

KEELING, P. J. Diversity and evolutionary history of plastids and their hosts. **American Journal of Botany**, [s. l.], v. 91, n. 10, p. 1481–1493, 2004.

KEELING, P. J. The endosymbiotic origin, diversification and fate of plastids. **Philosophical Transactions of the Royal Society B: Biological Sciences**, [s. l.], v. 365, n. 1541, p. 729–748, 2010.

KIMURA, M. **The neutral theory of molecular evolution**. [s.l: s.n.]. v. 241

KIMURA, M.; OTA, T. On some principles governing molecular evolution. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 71, n. 7, p. 2848–2852, 1974.

KORDIŠ, D.; TURK, V. Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. **BMC Evolutionary Biology**, [s. l.], v. 9, n. 1, p. 1–22, 2009.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Molecular biology and evolution**, [s. l.], v. 33, n. 7, p. 1870–1874, 2016.

LEEBENS-MACK, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. **Nature**, [s. l.], v. 574, n. 7780, p. 679–685, 2019.

LEGUIZAMÓN, N. D. P. et al. In vivo and in vitro anti-inflammatory and pro-osteogenic effects of citrus cystatin CsinCPI-2. **Cytokine**, [s. l.], v. 123, 2019.

LIMA, A. M. et al. A new *Piper nigrum* cysteine proteinase inhibitor, PnCPI, with antifungal activity: molecular cloning, recombinant expression, functional analyses and molecular modeling. **Planta**, [s. l.], v. 252, n. 2, p. 1–15, 2020.

LUGHADHA, E. N. et al. Counting counts: Revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. **Phytotaxa**, [s. l.], v. 272, n. 1, p. 82–88, 2016.

MARGIS-PINHEIRO, M. et al. Molecular evolution and diversification of plant cysteine proteinase inhibitors: New insights after the poplar genome. **Molecular Phylogenetics and Evolution**, [s. l.], v. 49, n. 1, p. 349–355, 2008. Disponível em: <<http://dx.doi.org/10.1016/j.ympev.2008.04.025>>

MARGIS, R.; REIS, E. M.; VILLERET, V. Structural and phylogenetic relationships among plant and animal cystatins. **Archives of Biochemistry and Biophysics**, [s.

l.], v. 359, n. 1, p. 24–30, 1998.

MARTINEZ, M. et al. The strawberry gene Cyf1 encodes a phytocystatin with antifungal properties. **Journal of Experimental Botany**, [s. l.], v. 56, n. 417, p. 1821–1829, 2005.

MARTINEZ, M. et al. Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases. **FEBS Letters**, [s. l.], v. 581, n. 16, p. 2914–2918, 2007.

MARTINEZ, M.; DIAZ, I. The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. **BMC Evolutionary Biology**, [s. l.], v. 8, n. 1, p. 1–12, 2008.

MASSONNEAU, A. et al. Maize cystatins respond to developmental cues, cold stress and drought. **Biochimica et Biophysica Acta - Gene Structure and Expression**, [s. l.], v. 1729, n. 3, p. 186–199, 2005.

MCCOURT, R. M.; DELWICHE, C. F.; KAROL, K. G. Charophyte algae and land plant origins. **Trends in Ecology and Evolution**, [s. l.], v. 19, n. 12, p. 661–666, 2004.

MELO, I. R. S. et al. CICPI, a cysteine protease inhibitor purified from *Cassia leiandra* seeds has antifungal activity against *Candida tropicalis* by inducing disruption of the cell surface. **International Journal of Biological Macromolecules**, [s. l.], v. 133, p. 1115–1124, 2019.

MONTEIRO JÚNIOR, J. E. et al. Expression in *Escherichia coli* of cysteine protease inhibitors from cowpea (*Vigna unguiculata*): The crystal structure of a single-domain cystatin gives insights on its thermal and pH stability. **International Journal of Biological Macromolecules**, [s. l.], v. 102, p. 29–41, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.ijbiomac.2017.04.008>>

MOURA, G. T. De et al. Crystal structure and physicochemical characterization of a phytocystatin from *Humulus lupulus*: Insights into its domain-swapped dimer. **Biochimica et Biophysica Acta - Proteins and Proteomics**, [s. l.], v. 1869, n. 1, 2021.

NEI, M.; KUMAR, S. **Molecular Evolution and Phylogenetics**, 2000.

PANCHY, N.; LEHTI-SHIU, M.; SHIU, S. H. Evolution of gene duplication in plants. **Plant Physiology**, [s. l.], v. 171, n. 4, p. 2294–2316, 2016.

PIROVANI, C. P. et al. *Theobroma cacao* cystatins impair *moniliophthora perniciosa* mycelial growth and are involved in postponing cell death symptoms. **Planta**, [s. l.], v. 232, n. 6, p. 1485–1497, 2010.

QIAO, X. et al. Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*). **Frontiers in Plant Science**, [s. l.], v. 9, n. February, p. 1–16, 2018.

QUAIN, M. D. et al. Ectopic phytocystatin expression leads to enhanced drought stress tolerance in soybean (*Glycine max*) and *Arabidopsis thaliana* through effects on strigolactone pathways and can also result in improved seed traits. **Plant Biotechnology Journal**, [s. l.], v. 12, n. 7, p. 903–913, 2014.

RAMBAUT, A. **FigTree - a graphical viewer of phylogenetic trees and a program for producing publication-ready figures**. 2014. Disponível em: <<http://tree.bio.ed.ac.uk/software/figtree/>>.

RAMBAUT, A. et al. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. **Systematic Biology**, [s. l.], v. 67, n. 5, p. 901–904, 2018.

REIS, E. M.; MARGIS, R. Sugarcane phytocystatins: Identification, classification and expression pattern analysis. **Genetics and Molecular Biology**, [s. l.], v. 24, n. 1–4, p. 291–296, 2001.

ROBERTS, T. C. **microRNA: Basic Science**. Cham: Springer International Publishing, 2015. v. 887 Disponível em: <<http://link.springer.com/10.1007/978-3-319-22380-3>>

RUHFEL, B. R. et al. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. **BMC Evolutionary Biology**, [s. l.], v. 14, n. 1, p. 23, 2014. Disponível em: <<http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-14-23>>

SANTANA, L. S. et al. TcCYS4, A cystatin from cocoa, Reduces necrosis triggered by MpNEP2 in tobacco plants. **Genetics and Molecular Research**, [s. l.], v. 13, n. 3, p. 7636–7648, 2014.

SANTIAGO, A. C. et al. A New Sugarcane Cystatin Strongly Binds to Dental Enamel and Reduces Erosion. **Journal of Dental Research**, [s. l.], v. 96, n. 9, p. 1051–1057, 2017.

SCHNEIDER, V. K. et al. Recombinant expression, characterization and phylogenetic studies of novel cystatins-like proteins of sweet orange (*Citrus sinensis*) and clementine (*Citrus clementina*). **International Journal of Biological Macromolecules**, [s. l.], v. 152, p. 546–553, 2020.

SEOIGHE, C.; GEHRING, C. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. **Trends in Genetics**, [s. l.], v. 20, n. 10, p. 461–464, 2004.

SHIBAO, P. Y. T. et al. Sugarcane cystatins: From discovery to biotechnological applications. **International Journal of Biological Macromolecules**, [s. l.], v. 167, p. 676–686, 2021.

SONG, C. et al. The *Arabidopsis* phytocystatin AtCYS5 enhances seed germination and seedling growth under heat stress conditions. **Molecules and Cells**, [s. l.], v. 40, n. 8, p. 577–586, 2017.

SUBBURAJ, S. et al. Molecular characterization and expression profiling of

Brachypodium distachyon L. Cystatin genes reveal high evolutionary conservation and functional divergence in response to abiotic stress. **Frontiers in Plant Science**, [s. l.], v. 8, n. May, 2017.

TAN, Y. et al. Overexpression of MpCYS4, a phytocystatin gene from malus prunifolia (Willd.) Borkh., enhances stomatal closure to confer drought tolerance in transgenic arabidopsis and apple. **Frontiers in Plant Science**, [s. l.], v. 8, 2017.

VALADARES, N. F. et al. X-ray crystallography and NMR studies of domain-swapped canecystatin-1. **FEBS Journal**, [s. l.], v. 280, n. 4, p. 1028–1038, 2013.

VAN DE PEER, Y. et al. The flowering world: a tale of duplications. **Trends in Plant Science**, [s. l.], v. 14, n. 12, p. 680–688, 2009.

VAN DE PEER, Y.; MAERE, S.; MEYER, A. The evolutionary significance of ancient genome duplications. **Nature Reviews Genetics**, [s. l.], v. 10, n. 10, p. 725–732, 2009.

VELASCO-ARROYO, B. et al. Silencing barley cystatins HvCPI-2 and HvCPI-4 specifically modifies leaf responses to drought stress. **Plant Cell and Environment**, [s. l.], v. 41, n. 8, p. 1776–1790, 2018.

WANG, K. M. et al. Characterization of inhibitory mechanism and antifungal activity between group-1 and group-2 phytocystatins from taro (*Colocasia esculenta*). **FEBS Journal**, [s. l.], v. 275, n. 20, p. 4980–4989, 2008.

WENDEL, J. F. et al. Evolution of plant genome architecture. **Genome Biology**, [s. l.], v. 17, n. 1, p. 1–14, 2016. Disponível em: <<http://dx.doi.org/10.1186/s13059-016-0908-1>>

YANG, A. H.; YEH, K. W. Molecular cloning, recombinant gene expression, and antifungal activity of cystatin from taro (*Colocasia esculenta* cv. Kaosiung no. 1). **Planta**, [s. l.], v. 221, n. 4, p. 493–501, 2005.

YANG, Z. PAML 4: Phylogenetic analysis by maximum likelihood. **Molecular Biology and Evolution**, [s. l.], v. 24, n. 8, p. 1586–1591, 2007.

ZHAO, P. et al. Comprehensive analysis of cystatin family genes suggests their putative functions in sexual reproduction, embryogenesis, and seed formation. **Journal of Experimental Botany**, [s. l.], v. 65, n. 17, p. 5093–5107, 2014.

APÊNDICES

APÊNDICE 1 - Bancos de dados utilizados para a busca de sequências de fitocistatinas.

Banco de dados	Link de acesso
1,000 Plants (1KP) Project	https://db.cngb.org/onekp/
Cucurbit Genomics Database (CuGenDB)	http://cucurbitgenomics.org/
Ensembl Plants	http://plants.ensembl.org/index.html
FernBase	https://fernbase.org/
Kiwifruit Genome Database	http://bioinfo.bti.cornell.edu/cgi-bin/kiwi/home.cgi
Klebsormidium Genome Project	http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/
Lotus Base	https://lotus.au.dk/
Mulberry Genome Database (MorusDB)	https://morus.swu.edu.cn/
National Center for Biotechnology Information (NCBI)	https://www.ncbi.nlm.nih.gov/
PhycoCosm	https://phycocosm.jgi.doe.gov/phycocosm/home
Phytozome	https://phytozome.jgi.doe.gov/pz/portal.html https://phytozome-next.jgi.doe.gov/
Plaza Gymnosperms	https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/
Solanaceae Genomics Network (Sol Genomics)	https://solgenomics.net/

APÊNDICE 2 - Relação das 188 espécies cujas sequências de fitocistatinas foram utilizadas neste estudo.
A tabela está ordenada de acordo com acrônimo utilizado para cada espécie.

Acrônimo	Espécie	Família	Ordem	Classe	Filo	Fonte	Tipo de dado
Achi	<i>Actinidia chinensis</i>	Actinidiaceae	Ericales	Magnoliopsida (Asterídeas)	Streptophyta	Kiwifruit DB	Genoma
Acoe	<i>Aquilegia coerulea</i>	Ranunculaceae	Ranunculales	Magnoliopsida (Eudicots basais)	Streptophyta	Phytozome	Genoma
Acom	<i>Ananas comosus</i>	Bromeliaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Afil	<i>Azolla filiculoides</i>	Salviniaceae	Salviniales	Polypodiopsida	Streptophyta	FernBase	Genoma
Ahal	<i>Arabidopsis halleri</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Ahyp	<i>Amaranthus hypochondriacus</i>	Amaranthaceae	Caryophyllales	Magnoliopsida	Streptophyta	Phytozome	Genoma
Aktr	<i>Akebia trifoliata</i>	Lardizabalaceae	Ranunculales	Magnoliopsida (Eudicots basais)	Streptophyta	1KP	Transcriptoma
Alin	<i>Alyssum linifolium</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Alyr	<i>Arabidopsis lyrata</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Aocc	<i>Anacardium occidentale</i>	Anacardiaceae	Sapindales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Aoff	<i>Asparagus officinalis</i>	Asparagaceae	Asparagales	Liliopsida	Streptophyta	Phytozome	Genoma
Arul	<i>Araucaria rulei</i>	Araucariaceae	Araucariales	Pinopsida	Streptophyta	1KP	Transcriptoma
Ashe	<i>Apostasia shenzhenica</i>	Orchidaceae	Asparagales	Liliopsida	Streptophyta	NCBI	Genoma
Atau	<i>Aegilops tauschii</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Atha	<i>Arabidopsis thaliana</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Atri	<i>Amborella trichopoda</i>	Amborellaceae	Amborellales	Magnoliopsida - ANA clade	Streptophyta	Phytozome	Genoma
Bdis	<i>Brachypodium distachyon</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Bhyb	<i>Brachypodium hybridum</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Bnap	<i>Brassica napus</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Ensembl Plants	Genoma
Bole	<i>Brassica oleracea capitata</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Brap	<i>Brassica rapa</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Bsta	<i>Brachypodium stacei</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Bstr	<i>Boechera stricta</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Bsyl	<i>Brachypodium sylvaticum</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma

Bvul	<i>Beta vulgaris</i>	Amaranthaceae	Caryophyllales	Magnoliopsida	Streptophyta	Ensembl Plants	Genoma
Camp	<i>Caulanthus amplexicaulis</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Cang	<i>Capsicum annuum glabriusculum</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Sol Genomics	Genoma
Cara	<i>Coffea arabica</i>	Rubiaceae	Gentianales	Magnoliopsida (Asteridaeas)	Streptophyta	Phytozome	Genoma
Cari	<i>Cicer arietinum</i>	Fabaceae	Fabales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Casi	<i>Camelia sinensis</i>	Theaceae	Ericales	Magnoliopsida (Asteridaeas)	Streptophyta	NCBI	Genoma
Cbra	<i>Chara braunii</i>	Characeae	Charales	Coleochaetophyceae	Charophyta	Ensembl Plants	Genoma
Cbre	<i>Cylindrocystis brebissonii</i>	Mesotaeniaceae	Zygnematales	Zygnematophyceae	Charophyta	1KP	Transcriptoma
Ccaj	<i>Cajanus cajan</i>	Fabaceae	Fabales	Magnoliopsida (Rosidaeas)	Streptophyta	NCBI	Genoma
Ccap	<i>Corchorus capsularis</i>	Malvaceae	Malvales	Magnoliopsida (Rosidaeas)	Streptophyta	Ensembl Plants	Genoma
Ccit	<i>Corymbia citriodora</i>	Myrtaceae	Myrtales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Ccle	<i>Citrus clementina</i>	Rutaceae	Sapindales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Ccus	<i>Cylindrocystis cushleckae</i>	Mesotaeniaceae	Zygnematales	Zygnematophyceae	Charophyta	1KP	Transcriptoma
Cesc	<i>Colocasia esculenta</i>	Araceae	Alismatales	Liliopsida	Streptophyta	NCBI	Genoma
Cglo	<i>Chaetosphaeridium globosum</i>	Chlorarachniaceae	Chaetosphaeriales	Coleochaetophyceae	Charophyta	1KP	Transcriptoma
Cgra	<i>Capsella grandiflora</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Chis	<i>Crambe hispanica</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Clan	<i>Citrullus lanatus</i>	Cucurbitaceae	Cucurbitales	Magnoliopsida (Rosidaeas)	Streptophyta	CuGenDB	Genoma
Clon	<i>Curcuma longa</i>	Zingiberaceae	Zingiberales	Liliopsida	Streptophyta	NCBI	Genoma
Cmar	<i>Cakile maritima</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Cmax	<i>Cucurbita maxima</i>	Cucurbitaceae	Cucurbitales	Magnoliopsida (Rosidaeas)	Streptophyta	NCBI	Genoma
Cmel	<i>Cucumis melo</i>	Cucurbitaceae	Cucurbitales	Magnoliopsida (Rosidaeas)	Streptophyta	NCBI	Genoma
Cmic	<i>Cinnamomum micranthum f. kanehirae</i>	Lauraceae	Lurales	Magnoliopsida - magnoliids	Streptophyta	NCBI	Genoma
Cmos	<i>Cucurbita moschata</i>	Cucurbitaceae	Cucurbitales	Magnoliopsida (Rosidaeas)	Streptophyta	NCBI	Genoma
Cpap	<i>Carica papaya</i>	Caricaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Cpep	<i>Cucurbita pepo subsp. pepo</i>	Cucurbitaceae	Cucurbitales	Magnoliopsida (Rosidaeas)	Streptophyta	NCBI	Genoma
Cqui	<i>Chenopodium quinoa</i>	Chenopodiaceae	Caryophyllales	Magnoliopsida	Streptophyta	Ensembl Plants	Genoma
Crei	<i>Chlamydomonas reinhardtii</i>	Chlamydomonadaceae	Chlamydomonadales	Chlorophyceae	Chlorophyta	Phytozome	Genoma
Crub	<i>Capsella rubella</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma

Csat	<i>Cucumis sativus</i>	Cucurbitaceae	Cucurbitales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Csin	<i>Citrus sinensis</i>	Rutaceae	Sapindales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Csub	<i>Coccomyxa subellipsoidea</i>	Coccomyxaceae	Trebouxiophyceae <i>ordo incertae sedis</i>	Trebouxiophyceae	Chlorophyta	Phytozome	Genoma
Cvio	<i>Cleome violacea</i>	Cleomaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Cvul	<i>Chara vulgaris</i>	Characeae	Charales	Coleochaetophyceae	Charophyta	1KP	Transcriptoma
Cyl	<i>Cylindrocystis sp.</i>	Mesotaeniaceae	Zygnematales	Zygnematophyceae	Charophyta	1KP	Transcriptoma
Cymi	<i>Cycas micholitzii</i>	Cycadaceae	Cycadales	Cycadopsida	Streptophyta	1KP	Transcriptoma
Czof	<i>Chromochloris zofingiensis</i>	Chromochloridaceae	Sphaeropleales	Chlorophyceae	Chlorophyta	1KP	Transcriptoma
Dcar	<i>Daucus carota</i>	Apiaceae	Apiales	Magnoliopsida (Asterideas)	Streptophyta	Phytozome	Genoma
Dcat	<i>Dendrobium catenatum</i>	Orchidaceae	Asparagales	Liliopsida	Streptophyta	NCBI	Genoma
Doli	<i>Dichanthelium oligosanthes</i>	Poaceae	Poales	Liliopsida	Streptophyta	NCBI	Genoma
Drot	<i>Dioscorea rotundata</i>	Dioscoreaceae	Dioscoreales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Dsal	<i>Dunaliella salina</i>	Dunaliellaceae	Chlamydomonadales	Chlorophyceae	Chlorophyta	Phytozome	Genoma
Dsop	<i>Descurainia sophioides</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Dstr	<i>Diptychocarpus strictus</i>	Dioscoreaceae	Dioscoreales	Liliopsida	Streptophyta	Phytozome	Genoma
Efim	<i>Entransia fimbriata</i>	Klebsormidiaceae	Klebsormidiales	Klebsormidiophyceae	Charophyta	1KP	Transcriptoma
Egra	<i>Eucalyptus grandis</i>	Myrtaceae	Myrtales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Egui	<i>Elaeis guineensis</i>	Arecaceae	Arecales	Liliopsida	Streptophyta	NCBI	Genoma
Esal	<i>Eutrema salsugineum</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Esyr	<i>Euclidium syriacum</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Euni	<i>Eugenia uniflora</i>	Myrtaceae	Myrtales	Magnoliopsida (Rosideas)	Streptophyta	LGPP own data	Transcriptoma
Even	<i>Ensete ventricosum</i>	Musaceae	Zingiberales	Liliopsida	Streptophyta	NCBI	Genoma
Eves	<i>Eruca vesicaria</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	1KP	Transcriptoma
Fves	<i>Fragaria vesca</i>	Rosaceae	Rosales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Gbil	<i>Ginkgo biloba</i>	Ginkgoaceae	Ginkgoales	Ginkgoopsida	Streptophyta	Plaza Gymno	Transcriptoma
Ghir	<i>Gossypium hirsutum</i>	Malvaceae	Malvales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Gmax	<i>Glycine max</i>	Fabaceae	Fabales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Gmon	<i>Gnetum montanum</i>	Gnetaceae	Gnetales	Gnetopsida	Streptophyta	Plaza Gymno	Transcriptoma

Grai	<i>Gossypium raimondii</i>	Malvaceae	Malvales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Gsoj	<i>Glycine soja</i>	Fabaceae	Fabales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Hann	<i>Helianthus annuus</i>	Asteraceae	Asterales	Magnoliopsida (Asterideas)	Streptophyta	Phytozome	Genoma
Hvul	<i>Hordeum vulgare</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Imar	<i>Iberis amara</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Itin	<i>Isatis tinctoria</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Kfed	<i>Kalanchoe fedtschenkoi</i>	Crassulaceae	Saxifragales	Magnoliopsida	Streptophyta	Phytozome	Genoma
Klax	<i>Kalanchoe laxiflora</i>	Crassulaceae	Saxifragales	Magnoliopsida	Streptophyta	Phytozome	Genoma
Knit	<i>Klebsormidium nitens</i>	Klebsormidiaceae	Klebsormidiales	Klebsormidiophyceae	Charophyta	Klebsormidium Genoma Project	Genoma
Ksub	<i>Klebsormidium subtile</i>	Klebsormidiaceae	Klebsormidiales	Klebsormidiophyceae	Charophyta	1KP	Transcriptoma
Lang	<i>Lupinus angustifolius</i>	Fabaceae	Fabales	Magnoliopsida (Rosideas)	Streptophyta	Ensembl Plants	Genoma
Lann	<i>Lunaria annua</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	NCBI	Genoma
Lesa	<i>Lepidium sativum</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	NCBI	Genoma
Ljap	<i>Lotus japonicus</i>	Fabaceae	Fabales	Magnoliopsida (Rosideas)	Streptophyta	Lotus Base	Genoma
Lper	<i>Leersia perrieri</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Lsat	<i>Lactuca sativa</i>	Asteraceae	Asterales	Magnoliopsida (Asterideas)	Streptophyta	Phytozome	Genoma
Lusi	<i>Linum usitatissimum</i>	Linaceae	Malpighiales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Macu	<i>Musa acuminata</i>	Musaceae	Zingiberales	Liliopsida	Streptophyta	Phytozome	Genoma
Mcha	<i>Momordica charantia</i>	Cucurbitaceae	Cucurbitales	Magnoliopsida (Rosideas)	Streptophyta	NCBI	Genoma
Mdom	<i>Malus domestica</i>	Rosaceae	Rosales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Mend	<i>Mesotaenium endlicherianum</i>	Mesotaeniaceae	Zygnematales	Zygnematophyceae	Charophyta	PhycoCosm	Genoma
Mesc	<i>Manihot esculenta</i>	Euphorbiaceae	Malpighiales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Mgut	<i>Mimulus guttatus</i>	Scrophulariaceae	Lamiales	Magnoliopsida (Asterideas)	Streptophyta	Phytozome	Genoma
Mmar	<i>Malcolmia maritima</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Ensembl Plants	Genoma
Mneg	<i>Monoraphidium neglectum</i> SAG 48.87	Selenastraceae	Sphaeropleales	Chlorophyceae	Chlorophyta	NCBI	Genoma
Mnot	<i>Morus notabilis</i>	Moraceae	Rosales	Magnoliopsida (Rosideas)	Streptophyta	MorusDB	Genoma
Mper	<i>Myagrum perfoliatum</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosideas)	Streptophyta	Phytozome	Genoma
Mpol	<i>Marchantia polymorpha</i>	Marchantiaceae	Marchantiales	Marchantiopsida	Streptophyta	Phytozome	Genoma

Msin	<i>Miscanthus sinensis</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Mtru	<i>Medicago truncatula</i>	Fabaceae	Fabales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Natt	<i>Nicotiana attenuata</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Ensembl Plants	Genoma
Nben	<i>Nicotiana benthamiana</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Sol Genomics	Genoma
Nnuc	<i>Nelumbo nucifera</i>	Nelumbonaceae	Proteales	Magnoliopsida (Eudicots basais)	Streptophyta	NCBI	Genoma
Nole	<i>Neochloris oleoabundans</i>	Neochloridaceae	Sphaeropleales	Chlorophyceae	Chlorophyta	1KP	Transcriptoma
Ntab	<i>Nicotiana tabacum</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Sol Genomics	Genoma
Obar	<i>Oryza barthii</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Obra	<i>Oryza brachyantha</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Oeur	<i>Olea europaea var. sylvestris</i>	Oleaceae	Lamiales	Magnoliopsida (Asteridaeas)	Streptophyta	Phytozome	Genoma
Oglu	<i>Oryza glumipatula</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Olae	<i>Onychonema laeve</i>	Desmidiaceae	Desmiales	Zygnematophyceae	Charophyta	Ensembl Plants	Genoma
Omer	<i>Oryza meridionalis</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Oniv	<i>Oryza nivara</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Opun	<i>Oryza punctata</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Oruf	<i>Oryza rufipogon</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Osat	<i>Oryza sativa</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Otho	<i>Oropetium thomaeum</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Pabi	<i>Picea abies</i>	Pinaceae	Pinales	Pinopsida	Streptophyta	Plaza Gymno	Genoma
Paxi	<i>Petunia axillaris</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Sol Genomics	Genoma
Pcat	<i>Psidium cattleianum</i>	Myrtaceae	Myrtales	Magnoliopsida (Rosidaeas)	Streptophyta	LGPP own data	Transcriptoma
Pdac	<i>Phoenix dactylifera</i>	Arecaceae	Arecales	Liliopsida	Streptophyta	NCBI	Genoma
Pdel	<i>Populus deltoides</i>	Salicaceae	Malpighiales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Pdru	<i>Phylloglossum drummondii</i>	Lycopodiaceae	Lycopodiales	Lycopodiopsida	Streptophyta	1KP	Transcriptoma
Pequ	<i>Phalaenopsis equestris</i>	Orchidaceae	Asparagales	Liliopsida	Streptophyta	NCBI	Genoma
Pgla	<i>Picea glauca</i>	Pinaceae	Pinales	Pinopsida	Streptophyta	Plaza Gymno	Transcriptoma
Phal	<i>Panicum hallii</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Pinf	<i>Petunia inflata</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Sol Genomics	Genoma

Pmen	<i>Pseudotsuga menziesii</i>	Pinaceae	Pinales	Pinopsida	Streptophyta	Plaza Gymno	Transcriptoma
Pmil	<i>Panicum miliaceum</i>	Poaceae	Poales	Liliopsida	Streptophyta	NCBI	Genoma
Ppat	<i>Physcomitrella patens</i>	Funariaceae	Funariales	Bryopsida	Streptophyta	Phytozome	Genoma
Pper	<i>Prunus persica</i>	Rosaceae	Rosales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Ppin	<i>Pinus pinaster</i>	Pinaceae	Pinales	Pinopsida	Streptophyta	Plaza Gymno	Transcriptoma
Psit	<i>Picea sitchensis</i>	Pinaceae	Pinales	Pinopsida	Streptophyta	Plaza Gymno	Transcriptoma
Psom	<i>Papaver somniferum</i>	Papaveraceae	Ranunculales	Magnoliopsida (Eudicots basais)	Streptophyta	NCBI	Genoma
Psyl	<i>Pinus sylvestris</i>	Pinaceae	Pinales	Pinopsida	Streptophyta	Plaza Gymno	Transcriptoma
Ptae	<i>Pinus taeda</i>	Pinaceae	Pinales	Pinopsida	Streptophyta	Plaza Gymno	Genoma
Ptri	<i>Populus trichocarpa</i>	Salicaceae	Malpighiales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Pvir	<i>Panicum virgatum</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Pvul	<i>Phaseolus vulgaris</i>	Fabaceae	Fabales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Rcom	<i>Ricinus communis</i>	Euphorbiaceae	Malpighiales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Risl	<i>Rorippa islandica</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Rsub	<i>Raphidocelis subcapitata</i>	Selenastraceae	Sphaeropleales	Chlorophyceae	Chlorophyta	PhycoCosm	Genoma
Sacc	<i>Saccharum</i>	Poaceae	Poales	Liliopsida	Streptophyta	SUCEST	ESTs
Salb	<i>Sinapis alba</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosidaeas)	Streptophyta	Phytozome	Genoma
Sbic	<i>Sorghum bicolor</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Scuc	<i>Salvinia cucullata</i>	Salviniaceae	Salviniales	Polypodiopsida	Streptophyta	FernBase	Genoma
Sdim	<i>Scenedesmus dimorphus</i>	Scenedesmaceae	Sphaeropleales	Chlorophyceae	Chlorophyta	1KP	Transcriptoma
Sfal	<i>Sphagnum fallax</i>	Sphagnaceae	Sphagnales	Sphagnophytina	Streptophyta	Phytozome	Genoma
Sind	<i>Sesamum indicum</i>	Pedaliaceae	Lamiales	Magnoliopsida (Asteridaeas)	Streptophyta	NCBI	Genoma
Sita	<i>Setaria italica</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Slyc	<i>Solanum lycopersicum</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Phytozome	Genoma
Slyd	<i>Solanum lycopersicoides</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Sol Genomics	Genoma
Smoe	<i>Selaginella moellendorffii</i>	Selaginellaceae	Lycopodiales	Lycopodiopsida	Streptophyta	Phytozome	Genoma
Smus	<i>Spirogloea muscicola</i>	Spirogloaceae	Spirogloaeales	Zygnematophyceae	Charophyta	NCBI	Genoma
Spenn	<i>Solanum pennellii</i>	Solanaceae	Solanales	Magnoliopsida (Asteridaeas)	Streptophyta	Sol Genomics	Genoma

Sper	<i>Solanum peruvianum</i>	Solanaceae	Solanales	Magnoliopsida (Asterídeas)	Streptophyta	Sol Genomics	Genoma
Spim	<i>Solanum pimpinellifolium</i>	Solanaceae	Solanales	Magnoliopsida (Asterídeas)	Streptophyta	Sol Genomics	Genoma
Spol	<i>Spirodela polyrhiza</i>	Araceae	Alismatales	Liliopsida	Streptophyta	NCBI	Genoma
Spur	<i>Salix purpurea</i>	Salicaceae	Malpighiales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Sspo	<i>Saccharum spontaneum</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Stub	<i>Solanum tuberosum</i>	Solanaceae	Solanales	Magnoliopsida (Asterídeas)	Streptophyta	Phytozome	Genoma
Svir	<i>Setaria viridis</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Taes	<i>Triticum aestivum</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Tarv	<i>Thlaspi arvense</i>	Brassicaceae	Brassicales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Tbac	<i>Taxus baccata</i>	Taxaceae	Cupressales	Pinopsida	Streptophyta	Plaza Gymno	Transcriptoma
Tcac	<i>Theobroma cacao</i>	Malvaceae	Malvales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Tdic	<i>Triticum dicoccoides</i>	Poaceae	Poales	Liliopsida	Streptophyta	Ensembl Plants	Genoma
Tint	<i>Thinopyrum intermedium</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Ugib	<i>Utricularia gibba</i>	Lentibulariaceae	Lamiales	Magnoliopsida (Asterídeas)	Streptophyta	NCBI	Genoma
Vang	<i>Vigna angularis</i>	Fabaceae	Fabales	Magnoliopsida (Rosídeas)	Streptophyta	Ensembl Plants	Genoma
Vcar	<i>Volvox carteri</i>	Volvocaceae	Chlamydomonadales	Chlorophyceae	Chlorophyta	Phytozome	Genoma
Vcar	<i>Vriesea carinata</i>	Bromeliaceae	Poales	Liliopsida	Streptophyta	LGPP own data	Transcriptoma
Vrad	<i>Vigna radiata</i>	Fabaceae	Fabales	Magnoliopsida (Rosídeas)	Streptophyta	Ensembl Plants	Genoma
Vung	<i>Vigna unguiculata</i>	Fabaceae	Fabales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Vvin	<i>Vitis vinifera</i>	Vitaceae	Vitales	Magnoliopsida (Rosídeas)	Streptophyta	Phytozome	Genoma
Zmar	<i>Zostera marina</i>	Zosteraceae	Alismatales	Liliopsida	Streptophyta	Phytozome	Genoma
Zmay	<i>Zea mays</i>	Poaceae	Poales	Liliopsida	Streptophyta	Phytozome	Genoma
Zygn	<i>Zygnemopsis sp.</i>	Zygnemataceae	Zygnematales	Zygnematophyceae	Charophyta	1KP	Transcriptoma

APÊNDICE 3 - Sequências aminoacídicas das fitocistatinas presentes em arroz, usadas como isca para as buscas em bancos de dados.

Ao fim da figura, em roxo e sublinhado, está indicada a porção de 58 aminoácidos utilizada para as buscas de fitocistatinas do tipo II, correspondente ao domínio inibitório para proteases do tipo legumína.

```
>Osa_OCI_LOC_Os01g58890
MRKYRVAGLVAALLVLHSLATPSAQAEAHRRAGGEGEEKMSDDGGPVLGGVEFVGNENDLHLVLDLARFAVTEHNKKANSLLFEKLVSVKQQVAVAGTLYYFTIEVKEG
DAKKLYEAKVWEKPMDFKELQEFKPVDSANA

>Osa_OCII_LOC_Os05g41460
MRASSLFAESVFTTSAAAGRRCPRLAAPVTLFFSTGRGSPAMAEAAQPRGVKGGIHDAAPAGRENDLTVELARFAVAEHNKANAMLELERVVKRQQVVGGF
MHYLTVEVKEPGGANKLYEAKVWERAWENFKQLQDFKPLDDATA

>Osa_OCIII_LOC_Os05g33880
MLRRRGFCCCSGAPAAAAAALLLAVAAAAAPRAAGFHLGGDESVLVVRGMLAAIRREQAEAEADAARFAVAEYKNKQGALEFARIVKAKRQVVTGLHDLMLLEVDSG
KKSLSYSAKVVVKPWLDKAVVEFRHVGDSSQSATAADDNAGQVVPDPTADPTVASRNDLHNTENNKVSVVLSTFSQTYSV

>Osa_OCIV_LOC_Os01g68660
MAARCPVGVASVLLLVIVLVTVAASAGSARSGGGGGGIRELRRGGGAGRRVGRTEVRDVEGDREVQELGRFSVEEHNRRRRSRDCGDVRLVFRVVAQRQVVSGLK
YYLRVAAAEEGAAGQNGGEPVFDVAVVVVKPWLESRTLLTFAPAADSFNES

>Osa_OCV_LOC_Os01g68670
MASKLYYAVAPLVLLVLLAPLSARLAAAAAADDGQWPAGGGRGRKVGGRDVEDVEGNREVQELGLFCVVEHNRRRGGSATRGRGLVFSRVVAAQTQVVSIGIKYY
LRIAAQEADDELVFDVAVVVKAWVPSREMVSVVPAALPGY

>Osa_OCVI_LOC_Os03g11180
MAMTTRTLLLAAVCAAAALFRGWSPIKNIDDPHIQELGRWAITENNRVSPSDELTFHRVTGGEQQVVS GMNYRLEIEAASGGGDVTGSYGAVVFEQEWSENTRKLISF
DKNHNF

>Osa_OCVII_LOC_Os03g11170
MTMRTSSLLLAAVAVVAIVAGATAATVGSWEFVDINDPHVQELGRWAVAEEDRVAAGGLTFERVTGEGKQVAVGNRYRLTLEASSSGAKDGRYEAVVVEQDPRSNA
RKLVSFEPIH

>Osa_OCVIII_LOC_Os03g31510
MARIPLLLALLLAVSAAAAAQVGNRGGHPLVGGWSPITDVGDPHIQELGGWAVERHASLSSDGLRFRRTS GEGQQVVS GMNYRLVVSASDPAGATASYVAVVVEQS
WTNTRQLTSFKPAAAH

>Osa_OCIX_LOC_Os03g11160
MRTSSLVLFVAAVVFGAACTAAAGDESWKTI DANDRHVQDVALWAVAETD WASATGGLTLNTVDGAEKRF EAGVNYRYRLTLEASSRVVAKYLRFAQVVEEGDEHKL
VSFVPIH

>Osa_OCX_LOC_Os04g28250
MATSPMLFLVSLLLVVAATGDEASPSNAAAAPAVLVGGRT EIRDVGSNKAVQSLGRFAVAEHNRRLRHGGSGGPADFPVKLAFARVVEAQKQVVS GVAAYLKV
AASARDPRGAAAGDRVFDVAVVVKAWLKS KELVSTPASSTK

>Osa_OCXI_LOC_Os09g08100
MARHPGLLLILLAAVAAVATTSRAQWVGGWNVIEDVAGNNQIQRVGAWAVGKHNLGTNDRLQFVRVVAEEQVVQGSNYLVVIDAASSRKKTRELYAVVADLVGA
TTYQLSSFKLATK

>Osa_OCXII_LOC_Os01g16430
MRVAATTRPASSSAAAPLPLFLLAVAAAAAALFLVGSASLAMAGHLVGGAHDA PSAANSVETDALARFAVDEHNKRENALLEFVRVVEAKEQVAVAGTLHHLTLEAL
EAGRKKVVEAKVVKPWLDKELQEFRN TGDATTF TNADL GAKKGGHEPGWRDVPVHPVVKDAADHAVKSIQQRNSLFPYELLEIVRAKAEVVEDFAKFDILMKL
KRGNKEEFKAEVHKNLEGAFVLNQMQEHDSSSQ
```

APÊNDICE 4 - Parâmetros utilizados para a inferência filogenética bayesiana.

Conjunto de dados	Tipo de dado	Parâmetros utilizados	
Fitocistatinas do tipo I + porção N-terminal de fitocistatinas do tipo II	Nucleotídeos	<i>Site model</i>	GTR+I+G4
		<i>Clock model</i>	Uncorrelated relaxed lognormal (DRUMMOND et al., 2006)
		<i>Tree model</i>	Birth-death
		<i>Chain length</i>	10 ⁸
		<i>Sample every</i>	5.000 trees
	Aminoácidos	<i>Site model</i>	WAG+G4
		<i>Clock model</i>	Uncorrelated relaxed lognormal (DRUMMOND et al., 2006)
		<i>Tree model</i>	Birth-death
		<i>Chain length</i>	10 ⁸
		<i>Sample every</i>	5.000 trees
Fitocistatinas do tipo II	Nucleotídeos	<i>Site model</i>	GTR+I+G4
		<i>Clock model</i>	Uncorrelated relaxed lognormal (DRUMMOND et al., 2006)
		<i>Tree model</i>	Birth-death
		<i>Chain length</i>	10 ⁸
		<i>Sample every</i>	1.000 trees
	Aminoácidos	<i>Site model</i>	JTT+I+G4
		<i>Clock model</i>	Uncorrelated relaxed lognormal (DRUMMOND et al., 2006)
		<i>Tree model</i>	Birth-death
		<i>Chain length</i>	10 ⁸
		<i>Sample every</i>	1.000 trees

186 Achi_Actinidia06951.t1
187 Hann_HanXRQChr04g0096521
188 Sind_XM_011097650.2
189 Atha_AT2931980_1
190 Kfed_Kaladp0490s0006
191 Fabi_MA_88557g0010
192 Rcom_29785_m000949
193 Ghir_Gohir.A09g109500.1
194 Ccle_Ciclev10017285m
195 Ghir_Gohir.D06g160800.1
196 Nnuc_XM_010244856.2
197 Accc_Anaoc_0007s1623_1
198 Accc_Aco022953
199 Cara_evm.model.Scaffold_584.142
200 Afil_Azfi_s003s.g025621
21 Mend_SM00021S06449
202 Kfed_Kaladp0053s0108
203 Ptae_PITA_000092205447
204 Gmon_Tns001038981t01
205 Ghir_Gohir.A09g208400.1
206 Gbil_GBI00024038
207 Cpep_XM_023680408.1
208 Sind_XM_011085844.2
209 Mdom_MD10G133500
210 Accc_Aco022220
211 Gmax_Glyma_11g253300
212 Pdel_Podel_09g020800
213 Gmax_Glyma_09g110600
214 Gmon_Tns000729105t07
215 Fabi_MA_43044g0010
216 Vvin_VIT_200s0187g00040
217 Vvin_VIT_200s0187g00110
218 Gmax_Glyma_18g103700
219 Pdac_XM_008790344
220 Ptae_PITA_000028756408
221 Cpep_XM_023665046.1
222 Sfal_Sphfalx0090s0036
223 Cara_evm.model.Scaffold_1805.172
224 Egra_Eucgr_E03919
225 Ahyp_AH011078
226 Accc_Aco021907
227 Cesc_AF525880.1
228 Macu_GSMUA_AchrUn_randomT25930
229 Egra_Eucgr_E03920
230 Mdom_MD12G1005200
231 Accc_Anaoc_0004s0465_1
232 Nnuc_XM_010245761.2
233 Hvul_HORVU1Hr16063320
234 Gmax_Glyma_20G045500
235 Ptae_PITA_00000825396
236 Ccle_Ciclev10003263m
237 Sfal_Sphfalx0007s0217
238 Gmon_Tns000131125t01
239 Hann_HanXRQChr06g0183231
240 Pdel_Podel_01G237100
241 Ahyp_AH005705
242 Kfed_Kaladp0068s0209
243 Mdom_MD14G1003700
244 Osat_LOC_Os03g11170
245 Cmic_RWR79917
246 Nnuc_XM_010258218.2
247 Egui_XM_019846279.2
248 Accc_Anaoc_0007s1626_1
249 Ghir_Gohir.A03g095800.1
250 Gmax_Glyma_13g209000
251 Kfed_Kaladp0926s0009
252 Pdel_Podel_14G157700
253 Ppat_Pp3c9_24090V3
254 Fequ_XM_020734796
255 Gmax_Glyma_14g038200
256 Hvul_HORVU2Hr1G124720
257 Rcom_30183_m001284
258 Cara_evm.model.Scaffold_396.1
259 Ptae_PITA_000089688432
260 Atri_evm_27_model_scaffold00055_79
261 Acce_Aqcoe4g276700
262 Hann_HanXRQChr16g0516541
263 Psom_XM_026536095
264 Ghir_Gohir.A06g153800.1
265 Cara_evm.model.Scaffold_521.176
266 Gmon_Tns000277535t10
267 Psom_XM_026526825
268 Ptae_PITA_000038429459
269 Kfed_Kaladp0666s0023
270 Vcar_Vocar_0049s0008
271 Tbac_TBA00005436
272 Gmax_Glyma_15g115300
273 Ghir_Gohir.A09g208200.1
274 Tbac_TBA00017565
275 Cpep_XM_023686782.1
276 Gmax_Glyma_07g266000
277 Atha_AT5647550_1
278 Fequ_XM_020719077
279 Sind_XM_011071778.2
280 Hann_HanXRQChr11g0344521

MARMGGIIEVKEIGNSEEMDSLARFAIEKYNKKNALLEFEKVVVAKKQV
SALRGGWKFPIDV-TDFTVVSLGRFAIDEHDMKESHASLKFVKVVKGDQQL
GKIVGGWQFIMNP-KDFEVVALANFAVTEHDMKESHASLKFVKVVKGDQQL
SVVLGGKGGVFNIRTRNREIQQLRGFAVVEQFMKQWENAILQFDSRVVSAEKQV
NAQLGGLFEIDVEHNNKEVQDLGRYAVVEEHNKQKGLRLVFSRVVSAEKQV
STRIGGIADIPGFERHKEIQDLGRFAVMEYNNKQKQAGLQFDSRVVSAEKQV
MATAGGIIEVQGSANSVEIDSLARFAVDDYNNKQKGNALLEYKVVVSAEKQV
MATLGGIIEVQGSANSVEIDSLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
MATVGGVREVATANNQEIVNLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
EAPLGGWRFPIDT-KDPHVMEIARFAVVEEYNNKQKGNALLEYKVVVSAEKQV
--VGGREIEIDVKDNEVEVQDLGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
EALTGGWQFIDP-KDQHVVEIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
EPKVGGITKIKHVGSNKEVQDLGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
MAKVGGIIEVQGSANSVEIDSLARFAVDDYNNKQKGNALLEYKVVVSAEKQV
AVITGGRRQTIENPQSNKELLEIARFAVVEEYNNKQKGNALLEYKVVVSAEKQV
FAASQDNDKMDVPSDDAVADVAFVAVKQLQEQSNILAQATLVLSPNAQV
SGAPGGWFAGNLT-TDPHVIEVSKFAVDEHNNKQKGNALLEYKVVVSAEKQV
STRVGGIIRDIDVPLWAKESQDLGRFAVDEYNNKQKGNALLEYKVVVSAEKQV
SKIVGGREIDIPDFQHDKTVQDLARFAVKEEYNNKQKGNALLEYKVVVSAEKQV
RNIMGGRFVDEVDNPKAMQALGRFAVVEEYNNKQKGNALLEYKVVVSAEKQV
LPWVEG-ATCRDSSKNSEIQSLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
GVECGGWEIFVNI-MDFHVVEIARFAVMEYNNKQKGNALLEYKVVVSAEKQV
--VGGREIENKQKGNALLEYKVVVSAEKQV
GLITGGWQFIDP-KDQHVVEIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
EPKVGGITKIKHVGSNKEVQDLGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
ARMVGGIIEVQGSANSVEIDSLARFAVDEYNNKQKGNALLEYKVVVSAEKQV
MATVGGIIEVQGSANSVEIDSLARFAVDDYNNKQKGNALLEYKVVVSAEKQV
ESALGGWEIFKIDV-NDSHVAEIANIYAVSEYNNKQKGNALLEYKVVVSAEKQV
ANLGLGGKHFVHGSNSQEIQLGRFAVDEYNNKQKGNALLEYKVVVSAEKQV
STRIGGIIRDIDVPEHKKIQLGRFAVDEYNNKQKGNALLEYKVVVSAEKQV
EALAGGWRPFIKNI-SDFRVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
DALGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
MAMIGGIISQVGEQSNVATIDRLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
MAALGGWRDADKEIENSVEIDELARFAVVEEYNNKQKGNALLEYKVVVSAEKQV
MGTLLGLLDVDFQNSIERLNLGRFAIDEHNNKQKGNALLEYKVVVSAEKQV
--LVGGWEIFKIDV-KDPHVVEIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
RTMVGAPKNIENPNS-NSVEIRELNFVAVSQYNNKQKGNALLEYKVVVSAEKQV
DALVGGWEIFKIDV-KDPHVVEIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
EPLVGGWEIFKIDV-KDPHVVEIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
MATVGGIIEVQGSANSVEIDSLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
HPIMGGIIDVDEGAQNSAEVQDLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
MALMGGIIEVQGSANSVEIDSLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
MATVGGIIEVQGSANSVEIDSLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
-ALVGGWRPFIKNI-SDFRVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
PFLRGGWEIFIRNLT-SDFHLREIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
EGFVGGKKEVNRVNTNKEVQDLGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
RIMVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
GLVRRGMLAAV--ERAEAEADARFAVDEHNNKQKGNALLEYKVVVSAEKQV
MAALGGWRDADKEIENSVEIDELARFAVVEEYNNKQKGNALLEYKVVVSAEKQV
GGTGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
GALVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
--MLGKPKNVEIENSIDLVDEIARFAVTEHNNKQKGNALLEYKVVVSAEKQV
FETKGGRRQDIIPGFERHKEIQDLARFAVDEHNNKQKGNALLEYKVVVSAEKQV
MATLGGIIEVQGSANSVEIDSLARFAVDDYNNKQKGNALLEYKVVVSAEKQV
MNMVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
DTPRGGWEIFKIDV-TDFSIVAIGFAVTEHNNKQKGNALLEYKVVVSAEKQV
PFLRGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
AATVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
GAKVGGRRKIEIRDVKSNEVQDLGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
FPALGGWEIFKIDV-EDPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
--VGGRRVVEIENSIDLVDEIARFAVTEHNNKQKGNALLEYKVVVSAEKQV
AVALLGGWEIFKIDV-KDPHVVEIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
GRMVGGRRTEIDTVKTNNEVQDLGRFAVVEEYNNKQKGNALLEYKVVVSAEKQV
EIDPKGWIIFVKI-NDPFAIEIARFAVTEHNNKQKGNALLEYKVVVSAEKQV
GVFISEWFFLKKDI-CDPKIQAIAKFAVVEEYNNKQKGNALLEYKVVVSAEKQV
GGMVGGRRVSDVKTNRVQDELGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
--MLGKPKNVEIENSIDLVDEIARFAVTEHNNKQKGNALLEYKVVVSAEKQV
ELIDGGWEIFKIDV-NDSHVAEIANIYAVSEYNNKQKGNALLEYKVVVSAEKQV
MAALGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
TTIDNGWRPFIKNI-NDPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
GMMVGGRRVSDVKTNRVQDELGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
GRKVGARRKIEIRDVKSNEVQDLGRFVVEEYNNKQKGNALLEYKVVVSAEKQV
STRVGGIIRDIDVPEHKKIQLGRFAVDEHNNKQKGNALLEYKVVVSAEKQV
EAFVGGYQQVQA--NRIHQDLGRFAVLAAYNNKQKGNALLEYKVVVSAEKQV
GLVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
KQVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
ANVLGGWRDADKEIENSVEIDELARFAVDEHNNKQKGNALLEYKVVVSAEKQV
GHLVGGWEIFKIDV-KDPHVVEIARFAVSEYNNKQKGNALLEYKVVVSAEKQV
FSPADAFRALD--PNDFLVVTAARFAVDEHNNKQKGNALLEYKVVVSAEKQV
SQMVGGRRVSDVKTNRVQDELGRFAVTEHNNKQKGNALLEYKVVVSAEKQV
KLLVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
STRVGGIIRDIDVPEHKKIQLGRFAVDEHNNKQKGNALLEYKVVVSAEKQV
GPRVGGIIRDIDVPEHKKIQLGRFAVDEHNNKQKGNALLEYKVVVSAEKQV
HHTLGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
MAGLGGWRDADKEIENSVEIDELARFAVDEHNNKQKGNALLEYKVVVSAEKQV
ESALGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
GTVIGGRTPVEDVQKPKAMQALGRFAVVEEYNNKQKGNALLEYKVVVSAEKQV
SEVFGELRIFIEI-GESEITA-ARFAVDEHNNKQKGNALLEYKVVVSAEKQV
ASTLGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
AARVGGWRPFIKNI-NDPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
EAGVGGWEIFEDV-KNPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
SPMVGAWEIPIKNI-NDPHVQELGFAVTEHNNKQKGNALLEYKVVVSAEKQV
MNTLGLLNNA--KNFAIEIARFAVQNNKQKGNALLEYKVVVSAEKQV
WGMLGARARKVDHVKTNKMIEIARFAVDEHNNKQKGNALLEYKVVVSAEKQV

281 Hvu1_HORVU2Hr1G124660
 282 Tbac_TBA00005316
 283 Macu_GSMUA_Achr3T29820
 284 Atri_evm_27_model_scaffold00087_28
 285 Hann_HanXRQChr04g0096491
 286 Nnuc_XM_010246134.2
 287 Acoe_Aqcoe4G277100
 288 Atha_AT4G16500.1
 289 Stuh_PGSC0003DMF400097644
 290 Egra_Eucgr_L00642
 291 Cpep_XM_023658118.1
 292 Macu_GSMUA_Achr5T25320
 293 Osat_IOC_Os03g31510
 294 Egui_XM_019852036.2
 295 Gmax_Glyma_05G149800
 296 Hann_HanXRQChr02g0058381

T I E N G W E P I G N I - N D Q H I Q G G L G S W A V L E F S K R M N C V L K F S K V V S G R Q Q Q
 S R K V G G L D V F N F Q Q N K R I K S L A K M A V K E Y N K Q Q N A A L D L F S K V V K A G Q Q Q
 G K A L G G W F I K D V - N D P H V Q E I A Q F A V S E H M K Q D H H P Q L D L F G K V A K G S E Q Q
 F L R L G G R A E I F N V K S T E V Q D L G R F C V E T Y N Q D Q H H P Q L D L F G K V V K A G S E Q Q
 G T L T G G W K P I F D V - T D F T V V D I A K F A V D E H N K E A H A S L D K F V K V V K A G S E Q Q
 G V I L G G I H D S V G F E N S V E I E D L A R F A V D E H N K K Q N A L L K F V G R V V M A K M K V Q
 G A V V G G Y K I I K D L - K E F I Q D L S K F V V L Q Y N K Q E K S N L K Y V N V I K G Q M K V Q
 G G L G S R K P I K N V - S D F D V V A V A K V A I E E H N K E A K K R L D F V F K V I K G S E Q Q
 E A L A G D W E P I T N I - - T E E V T K I G R F A V D E H N K E A K K R L D F V F K V I K G S E Q Q
 G P R V G G R T A V F D V R S N E E V Q G D L G R Y S V E E Y N R R S R G E V K F L A V L A A E R Q V Q
 S G R I G G R R R I K D V R S N E E V Q G D L G R Y S V E E Y N R R S R G E V K F L A V L A A E R Q V Q
 G R M V G A R T E V Q D V E T N K E V Q D L G L F S V D E Y N R R L A L L T F S R V A A A Q R E Q V Q
 G P L V G G W S P I T D V - G D F H I Q E L G G A V E R H A S L S S D G L R F R R V T S A G E Q Q
 M A T L E G M R D S Q G N Q N I E I D E L G R F A V E R H N K K E N A L L E F G R V V K A K Q Q V Q
 G G L V G G R S K I F D V K A N K K V Q D L G R F S V E E H N R M L R R Q V T F V E V V E A Q Q Q V Q
 A I P G G R R K V K N V K T D T E I Q D L G S Y S V D E Y N R L Q R G D L K F S Q V I A A E T Q V Q

001 Lusi_Lus10026779
 002 Ghir_Gohir_A09G208300.1
 003 Hann_HanXRQChr16g0516551
 004 Afil_Azfi_s0704.g082661
 005 Scuc_Sacu_v1.1_s0001.g000350
 006 Cpep_XM_023664468.1
 007 Ghir_Gohir_D01G221400.1
 008 Atri_evm_27_model_scaffold00087_26
 009 Ptae_PITA_000028582390
 010 Hann_HanXRQChr01g0008051
 011 Acoe_Anaoc_00050910.1
 012 Lusi_Lus10034773
 013 Mdom_MD07G1042400
 014 Osat_IOC_Os01g68660
 015 Hvu1_HORVU4Hr1G069140
 016 Kmit_kf100333_0060
 017 Osat_IOC_Os03g11180
 018 Ghir_Gohir_D09G106100.1
 019 Ccle_Ciclev10017209M
 020 Ghir_Gohir_D02G113100.1
 021 Egra_Eucgr_B03923
 022 Pdac_XM_008794726
 023 Osat_IOC_Os04g28250
 024 Gbil_GBI00024011
 025 Psom_XM_026585347
 026 Acoe_Aqcoe4G277300
 027 Acom_Aco017354
 028 Ccle_Ciclev10013595m
 029 Acoe_Aqcoe1G252700
 030 Macu_GSMUA_AchrUn_randomT14080
 031 Gmon_TnS000131125t02
 032 Gmax_Glyma_11G064200
 033 Mdom_MD12G1006000
 034 Sind_XM_011082189.2
 035 Egra_Eucgr_A02187
 036 Atri_evm_27_model_scaffold00214_3
 037 Psom_XM_026545087
 038 Ccle_Ciclev10002632m
 039 Vvin_VIT_205s0029g00370
 040 Pdel_Podel_02G260300
 041 Ccle_Ciclev10013601m
 042 Ptae_PITA_00003584414
 043 Ghir_Gohir_D06G160600.1
 044 Egui_XM_019846616.2
 045 Cpep_XM_023664469.1
 046 Achi_Actinidia20846.t1
 047 Cara_evm.model.Scaffold_216.547
 048 Ghir_Gohir_D02G018300.1
 049 Hvu1_HORVU1Hr1G067980
 050 Hann_HanXRQChr04g0096501
 051 Dcar_DCAR_002223
 052 Rcom_27574_m000225
 053 Ghir_Gohir_A06G154000.1
 054 Mdom_MD05G1358700
 055 Pdel_Podel_06G016300
 056 Gmax_Glyma_09G010900
 057 Egui_XM_010933246.3
 058 Egra_Eucgr_B03762
 059 Ghir_Gohir_D09G201900.1
 060 Ccle_Ciclev10009506m
 061 Ptae_PITA_00003583366
 062 Dcar_DCAR_031843
 063 Stuh_PGSC0003DMF400069145
 064 Cpep_XM_023687730.1
 065 Psom_XM_026581878
 066 Gmon_TnS001038981t03
 067 Psom_XM_026581896
 068 Nnuc_XM_010244858.1
 069 Ccle_Ciclev10006584m
 070 Osat_IOC_Os01g58890
 071 Acom_Aco022509
 072 Cmic_RNR76071
 073 Acoe_Aqcoe2G264800
 074 Stuh_PGSC0003DMF400058429
 075 Hann_HanXRQChr17g0559581

V A G T I H H I T L E A T D G G K K K T Y E A K V V V K F W I K F K E V Q A S K L V G D A - - -
 V S G I K Y F L T I E G M K N G K K T F D S I V L R P W K K D K K L Q S E S R I Q - - - -
 V S G I N W R L T I E V D V R N V T K C E V A Q V Y E Q P W Q N F K K L L D N E K M V - - - -
 V S G V M Y L L H L Q R F R K G Q T C S Y E A Q V W V Q A W R N F K S L L Q N E F K V D E Q N A -
 V A G I M Y L L V L E A P N G G P S C F Y E A K V W R A W E N F K S L E G F E E I Q L P I L V
 V A G T N Y L R L T A L E G T S S R T Y -
 I S G I N Y K L V L Q A K D G T A D N S Y E A V V W E K A W L K F R K L T S E T L V K G - - - -
 V S G M K Y L L L E A K D G D V L K K F D A A K V V V Q A W R H S K Q M L S E S P S T D - - - -
 V S G L M Y L L R V E T D G G K T R L Y A A K I W V K P W E K F R Q L Q E E K F V Q T L H Q -
 V A G K M Y L L T L E A E D A A A T K V Y E A K V W V K P W M N F K S L L E F L M L L L S L G
 V S G T V Y L L T L E A S E G D K K N L Y E A K V W V K P G C N S R S C L M - E L M L L L S L G
 I Q G K N Y L R L L V F D A A V I S N Y V A V V W L F G V S P K K L M - E L C L G R - - - -
 V S G T M H Y L L K I E A S E G G K K V Y E A K V W V K P W E H F R Q V L Q T E K L V S D S Y R G
 V S G L K Y L L R V A A E E G A P R V F D A V V V K P W L E S R T L W S T E A F A A D S P N E
 V Q G V N Y R V F L H L V T G E K G Q F A G V Y E G A S S S A R K L V S P D F I I - - - -
 V N G L L F L L K L K - V S G G S A E S Y K A T V Y Q F F - G A A K T L D K S F R E V - - - -
 V S G M N Y R L E I E A A S G G G T G S Y G A V V F E E W S N T R K L L Q S E P D K N H N F - -
 V S G T M Y I T L E A M D G D K T K V Y E A K V W D K P W M N F K E L Q D S E F V I G D A P A G
 V S G T N Y R L I L V V K D G F S T K K F E A V V W E K P W E H F K S L T S E K F V A K - - - -
 V S G I K Y L L K I E A M E N K L F T F E S V V V K P W L H S K R L L H S F S P S A H R - - - -
 V A G M N Y R L L V E A K D G T N T K S F E A L V W D K P W E H S R R L S S F K A V K G K A - -
 V S G I K Y L L K I A A R D G R E - R S F D A V V V V K P W L K S R S L L S F A P S A H R - - -
 V S G V A Y L L K V A A S A R D F R V F D A V V V K A W L K S K E L L S F T P A S S T K - - -
 V A G T M H L L I E V S G T E R P R L Y E A K V L V K P W E N F K Q L V D S F K F V H - - - -
 V Q R F K Y M N V S A I E N G K L N L Y D A V V V Q A A I P S N E L I S S E N I F L V K - - - -
 V A G I N Y D L V I A A R K G D I R R N Y L A V V Y E R P G G K N K T L T S Y E R P A S L N G V -
 V S G I K Y Y L R I A F A D G G K R S F D A V V V K P W L P S R T L A V V S G P A V D R Y G - -
 V A G K L Y L L I Q V I N T G G K K I Y E A K I S V K S W N N F K Q L W E S F K H V K H G F S
 V S G M K Y Y L L K V N A T Q N G A S K M F D A V V V K P W D Q S K E L V Q S F S P Y A N - - - -
 V A G V N Y K L L L V A K N E K K S A G Y E A V V W E K E W E N F R K L L S E F R R N - - - -
 V A G V Y Y L L I Q S K D E G T T A S Y E A R V W T K P W Q D F K R L E Y I R E V P F F A L S D
 V A G T N Y R L L V L K N G S T T A S Y Q A T V L E K P W L H F R N L L S F K P L R S - - - -
 I A G I N Y K L V L E V T D S S S P I N Y E I I V F E D L - S R T T K L L E S E V R V Q N - - - -
 V A G T M Y I T L E A A E G G Q K K V Y E A K V W V K P W M N F K Q L Q E S F E P V A - - - -
 V A G T L Y H L T L E A V D A G K K K I Y E A K V W V K P W M N F K Q L Q E S F E P V A - - - -
 V A G I N Y R L L I A Q A F E F F V R Y Y K A I V W E K P W L K F M N L L S F E P V A - - - -
 V S G L K Y Y L L K I S A L T S V S E K K F D A V V V K P W I R S K V L L N S F A P S V N - - - -
 V S G T N Y R L I L V V K D G F S T K K F E A V V W E K P W E H F K S L T S E F K P V K - - - -
 V S G F K Y Y L L I A A T Q N G V R N T F D A V V V K P W L P S K Q L L H F S P S S I Y - - - -
 V S G I F Y Y L L K I E A T Q S E L M F D S V L L V K P W L R S K E L L H F E P S I G L R V - -
 V A G K L Y Y I T L Q V I D T G K K R Y A A K I C - - - W - Q A K E K R R R E A F - - - -
 V A G V N Y Y L T I E A K D G G K T R V Y A E A I W M K A W E N F K Q L L D F K P V Q T P H - -
 V A G I N Y R L L E A K N G S V A K K Y Q A V W E K S W L N F R N L T S E T L V K G - - - -
 V S G I I Y I T I E A D G G K K K Y E A K V W V K P W M K F K E L Q E E I P L - - - -
 V A G T N Y L R L T A L E G T S R G T Y G T V L V - - - - F - - - - - - - - - - - - - - - -
 V S G I K Y Y L L K I S A T R D G V P K T F D A V V V V K P W A H S K E L L N E S S A T K - - - -
 V A G K V Y H L T L E A I D A G K K Q I Y E A K V W V K P W M N F K K L E E E R H T R D I P S F
 V A G I K Y Y L L K I K A M Q G G V T K T F E S V V L V K P W V Q S K D L L N S E S P S T H - - - -
 V A G M H Y F T I E K E G G A K K Y E A K V W E K A W E N F K Q L Q E S F K P A - - - -
 V A G M N Y N L T I T A S D R G V R K N Y V A L V W D K P W L K F R K L L S F K G P V - - - -
 V A G T N F R L T I S V K M D V S I N Y Q A T V F D Q P W T H T R N L T S E F K P M - - - -
 V A G T N Y K L L V L A V S G E D S - K Y E A V V W E K P W E K F R E L T S E F K P V R G - - - -
 V S G I N Y K L L K A K D G S A V N T Y E A V W E K A W L H F R N L T S E F L V M D - - - -
 V A G T N Y R L L F I S A K N E S S G T N Y E V F V Y D R A W E H H K E L T S E H R S K K N - - -
 V S G T N Y R L L V L - A V K G R A N A K Y Q A V Y E K P W E N L K R L T S E Q F K G - - - -
 V G T N Y R L L V L K V K N G S T T S Y Q A T V L E K P W L H F R N L T S E F P L R S - - - -
 V S G I I Y Y T I E A I D G G K K K Y E A K V W V K P W M K F K E L Q E E F - - - -
 V S G I M Y Y T L E A D G G V K K Y D A K I W V K P W M N F K S L Q E E I V G D A P A E
 V S G I K Y Y L T I E G M E N G K K K T F D S V L I K P W K E D K E L I S E S P I Q - - - -
 V S G T N Y R L L V A A Q D G T A T N K Y Q A V V W E N A - K Q I K K L T S E H R L - - - -
 V A G T V Y Y L T I E A K D G A K T G V Y A E A I W K A W E N F K Q L Q E E F P A Q T P H - - -
 V A G T M H L L T V E V I D A G K K K Y E A K V W V K P W L N F K E L Q E E F K P A G D S A - -
 V A G I Y Y I T L V A I D A G K K K I Y E A K V W V K V E N F K K V E E K L V G D D S A K
 V A G K L Y H L T L E A S D G M K K Y E A K V W V K P W M N F K Q L Q E E F K P - - - -
 V A G I Q Y Y L L K V S A M E N G S P K F F D A V V L V K A W E P A N R L I S E N S P S T E N - -
 V A G M M Y L T I K V K D S G S N F L Y E A K I W K S W Q N F R R L E D S P V S T N T G G H
 V S G I Q Y Y L L H E A I Q N G K F K L Y D A V V V K A W E P S N S L V S D P S Y - - - -
 V S G I N Y Y L L H I E A L Q H G M T K R F D A V V A V K P W L P S K Q L L A S P - - - -
 V S G I K Y Y L T I E A T G E N I Q M F D S I V I R P W L H S K E L L R E A P S E - - - -
 V A G T L Y Y F T I E V K D G A K K L Y E A K V W E K P W M D F K E L Q E E F K P V D A S A N A
 V A G I N Y R L L V L Q A D A A S E C Y E A V V W E K A W Q G F R K L Q E E F R V V - - - -
 V A G T M H Y I T L E V E A G Q K I Y E A K V W V K P W E N F K E L Q E E F R V G D S S S T
 V G Y N Y R L D I Q T R N A D Q I N T Y A Q V Y E K L - D G Q M I L N S Q P L - - - -
 V A G T M Y I T L E A I E G K K K A Y E A K V W V K V Q N F K Q I E D R I I G D A A S A
 V G M N H R L T I E V E D H R S I K T C K V L V E Y Q F L Q N V K K L V S E R I V - - - -

076 Ghir_Gohir_D09G201700.1
077 Ghir_Gohir_A06G5154100.1
078 Osat_LOC_0s09g08100
079 Psom_XM_026528338
080 Dcar_DCAR_002225
081 Mend_ME000129S00342
082 Hvu1_HORVU7Hr1G081260
083 Osat_LOC_0s05g41460
084 Egui_XM_010912231.3
085 Ppat_Fp3c13_18300V3
086 Osat_LOC_0s03g11160
087 Vvin_VIT_200s0187g00120
088 Smoe_19455
089 Chra_GB885901
090 Hann_HanXRCChr11g0344531
091 Ccle_Ciclev10004025m
092 Ptae_PITA_000021635411
093 Cara_evm.model.Scaffold_551.46
094 Dcar_DCAR_013534
095 Pabi_MA_9875200g0010
096 Vvin_VIT_206s0009g02230
097 Hvu1_HORVU3Hr1G093270
098 Ahyp_AH006515
099 Acoe_Ac006220200
100 Gmon_TnS000729105t06
101 Kfed_Kaladp0060s0419
102 Egui_XM_029265818.1
103 Acoe_Anaoc_0016s0821_1
104 Osat_LOC_0s01g68670
105 Psom_XM_026526824
106 Pabi_MA_10433951g0030
107 Mdom_MD05g1358900
108 Ghir_Gohir_D09G201800.1
109 Achi_Actinidial6365.t1
110 Ahyp_AH013319
111 Egra_Eucgr_A02184
112 Ahyp_AH008572
113 Ptae_PITA_000075197375
114 Ptae_PITA_000024373951_mod
115 Stub_PGSC0003DMT400034557
116 Scuc_Sacu_v1.1_s0138.g022534
117 Psom_XM_026526823
118 Osat_LOC_0s05g33880
119 Gmon_TnS000147417t01
120 Cara_evm.model.Scaffold_577.56
121 Dcar_DCAR_004907
122 Scuc_Sacu_v1.1_s0033.g011021
123 Psom_XM_026602685
124 Kfed_Kaladp0081s0039
125 Gmax_Glyma_13G071800
126 Cmic_RWR76069
127 Hvu1_HORVU2Hr1G067330
128 Acom_Aco000064
129 Kfed_Kaladp0064s0065
130 Macu_GSMUA_Achr1T20100
131 Ahyp_AH021448
132 Acoe_Ac0055136000
133 Cara_evm.model.Scaffold_539.650
134 Stub_PGSC0003DMT400092209
135 Hann_HanXRCChr09g0273691
136 Atha_AT2G40880.1
137 Pequ_XM_020733958
138 Macu_GSMUA_Achr4T13950
139 Hvu1_HORVU1Hr1G067870
140 Crei_04_g225650_t1_2
141 Acoe_Anaoc_0010s0543_1
142 Pdel_Podel_069033600
143 Cara_evm.model.Scaffold_507.45
144 Gbil_GBI00024199
145 Acom_Aco020743
146 Egui_XM_010928371.3
147 Hann_HanXRCChr13g0420981
148 Gmax_Glyma_18G003700
149 Cpep_XM_023680405.1
150 Mdom_MD03g1148300
151 Stub_PGSC0003DMT400046191
152 Nnuc_XM_019197176.1
153 Ahyp_AH014908
154 Ghir_Gohir_D06G160700.1
155 Gmax_Glyma_04G096400
156 Hann_HanXRCChr04g0096481
157 Ptae_PITA_000074704573
158 Egra_Eucgr_E03921
159 Achi_Actinidial6365.t1
160 Egra_Eucgr_J00666
161 Lusi_Lus10041257
162 Atha_AT5G12140.1
163 Cmic_RWR76315
164 Hvu1_HORVU3Hr1G070020
165 Kfed_Kaladp0080s0094
166 Egra_Eucgr_A02189
167 Mdom_MD02G1274700
168 Cpep_XM_023674627.1
169 Ptae_PITA_000048208498
170 Kfed_Kaladp0926s0008

V S G I K Y I L T I E G H N G E K R M F N S V V L I K P W K G D K E L L S S P T Q
V A G T N Y R L V I L Q A K E S S R K R E L Y V A V V V A D L V G A T P A D R N L S S E N L V K G
V Q G S N Y F L V I S A A K S S R K R E L Y V A V V V A D L V G A T P A D R N L S S E N L V K G
V A G M Q Y F L V I S A A K S S R K R E L Y V A V V V A D L V G A T P A D R N L S S E N L V K G
V A G T N Y R L V I L Q A K E S S R K R E L Y V A V V V A D L V G A T P A D R N L S S E N L V K G
V D G T L Y T L A L K L E A R D G K K E H I E A V V V H H S A S G E W K K L S S E K F T K
V N G V N Y Q L V I D L R D G S H R T Y K A V V V F E K D S N K F K Q L S S E T F A N
V G G F M H Y L T V E A K N G G A N K L Y E A A K I V W E R A W E N F R K Q L S S E K F L D D A T A
V G G I N Y R L V I L Q A K E S S R V A K R F Q A V V V E E G - D E H K L L S S E Q P V Q S
V Q G K M Y F T I E A A E A N G V P K N Y D A A K V V W V K P W E G Y K K L S S E L F S A P S H Y D
V E G L N V R M V I Q A L E S S E L P K M Y K A V V V W E K P W E N F K N L S S E S F V L V
V S G V M Y L L T V E A R D S K N V C S Y E A V V V W V K E W E N F R R L S S E D R I
V A G T L Y Y L I K V R L R A K G G T P T K Y E A V V V W D Q P W V P H R R L S S E V R I P T K T I F
V S G M K Y Y M K I E L T K S G P K V F E A V V V V V K P W L R S K Q L S S E G P S Q G V L R P
V S G T N Y L L I L V K D G G P S T K K F E A V V V L E K P W E H F K S L S S E R H T R D I P S F
V A G M R Y Y L T I E A R D G G K T G V Y A A E L V V K A W E N F K Q L S S E K F V Q T P H
V A G K L Y H L T L E A I D S G K K K Q I Y E A K V W V K P W M N F K Q L S S E K S A N D V P V F
V A G T M Y H L T L E A I D S G K K K H M Y E A K V W V K P W M N F K Q L S S E K S A N D V P V F
V L G L V Y R L V I L Q A L E S S E T M D G S K T K L Y A A A K I V W V K P W E - - - - - K E
V A G T I Y Y I T L D A R D G G K K N K Y E A A K I V W V K P W L N F K E L Q S S E K P I G D B P A T
V S G V K M Y Y L T V E A A E A N G V P K N Y D A A K V V V V K P W L Q S R S L S S E A P A D A
V A G L K Y Y L T V K A E G E D H T K V F D A A V V V V V K P W V Q S K K L S S E G P S L E S N L N
V S G L K Y Y L L K V A V E N G I A D T V F A V V V V V K P W I R S K E L S S E A P S T K
V S G T V Y Y L T I E A R D G A G A K L Y E A A K V V W V K P W E N Y K S L S S E S S V I Q P E V K
V S G I N Y R L V I L Q A L E S S E Q D R D G G K G K Y E A A K V V W V K P W L R L K S L S S E Q P V
V E G I N Y R L V I L Q A L E S S E T K A D Y E A A K V V W V K P W E N F R K L S S E - - - - -
X E D F A F N M L L K V R R G D K E E K P V A V V V H K E T - E G T F H L H R M K Q D N S - - - - -
V S G I K Y Y L R I A A E A D D E L V F D A A V V V V V K A W V P S R E M L S S E V P S A E L P G Y
V A G I Q Y Y L K V S A M E N G S P K F F E A V V V L V K A W E P A N R L S S E R H T R D I P S F
V S G M V Y R L T M E A I D G H R T G L Y L A A K V V V V P W T N F T Q L S S E H Q L D A D K D G
V A G I N Y R L V I S A A K N K S V P A D Y E A K V W E K A W E H F K Q L S S E K R L S K P N - - - - -
V S G I K Y Y L T I E G M E N G K K K T F D S I V I L I K P W K K D K K L S S E S P I Q - - - - -
V A G T Y Y L T I E A T C G G Q T K V Y E A I V E K L W M N F K Q L S S E K L I S D A - - - - -
V S G I N Y R L I I S A N D - - - - - T L D Y E A I V - - - - - F - - - - - V P A S - - - - -
V A G T L Y H L T L E A V D S G K K K I Y E A A K V W L Q E F K Y A E D V P A S K A S G L G R G I
V A G M N Y R L D F E A I N I N F N K K Y Q A A I I Y E D L - K G N L E L S S E S P L L Q A N E K
V A G V R Y Y L T I E A R D G K K T G V Y A A A K I V V K A W E N F K Q L S S E K F V Q T P H - - - - -
V A G T V Y H L P I E A D D G E - P K L Y E A A K V V W V K P W E N F K R L S S E K P V Q P S I S
V A G T N Y R L V I N A K D G G H V R K Y L T V A A K V V V V D K P W E K I K K L S S E K Q M - - - - -
V S G L I Y R L I E V E S G G I K K L Y D A A K V L E K A W Q N F K S L S S E T P H Q Q G S - - - - -
V S G I K Y Y L K V S L W E G G K P H L F D A A K V V V V K A W E P S R S L S S E G P V D - - - - -
V T G T L H D L M L E V D S G K K S L Y S A A K V V W V K P W L D F Q A V V E R H V G D S Q S Q
V A G L Y Y L T I E V K D S S A S K L Y E A A K I V V K S W E K F Q S L S S E R H V P D S N G D Q
V G A K Y Y G L I K A A E N G T S N T Y A A I V L D S K - D S G K E L S S E R A K R I L A V - - - - -
V A G L N Y R L I E A R D G D E I Q H Y E A I V V V K A G K D S K L S S E A P I E E L A - - - - -
V S G I M Y Y L H L Q A S G A K T C G Y E A A K V V V Q A W K N F K S L S S E T L E Q C V - - - - -
V S G I K Y Y L K V S L W E G G K P H L F D A A K V V V V K A W E P S R S L S S E G P V D - - - - -
V A G T L Y Y L T V E A G G K K K I Y E A A K V V V K P W E N F K E L S S E K F V G D S S S T
V S G V Y Y L K V I A R A F A G D R P F D A A K V V V K A W A K S K E L S S E S P F K - - - - -
V A G I N Y R L V L Q - A E A D A S E C Y E A A K V W E K A W Q G F R K L S S E R F V - - - - -
V A G L H H I T M E V R D G V A N K L Y E A A K V W V K P W M N F K Q L S S E K P V D E - - - - -
V A G Q M Y Y L T L E A E A D G G K K L Y E A A K I V V K P W M D F K E L S S E K L D D S L A A
V A G F K N Y R L T I - A E D S S K - - K Y E V V V Y D Q P W Q R Y R K L S S E K L S S E K L
V A G Y N Y R L N I A A R D G G T L N T Y V A Q V Y E D I - D Q S L T L S S E R H L - - - - -
V A G T N Y K I V I K A L D G T A S N L Y E A I V W V K P W L K F K L S S E R K L P - - - - -
V E G I N Y R L V I E A K D G D S T H N Y L A E W D K P - D K S K S L S S E Q Q L L E A - - - - -
V Q G T M Y Y I K L E A N D G G E K K T Y A A K V W V K P W E N F K E L S S E Q R T K L V - - - - -
V A G T M Y H L T L E A E A D G Q T K N F E A A K V V V K P W M N F K Q L S S E K S S - - - - -
V S G I K Y R I R V A A V D R T R R F D A A K V V V R A W R G D R L S S E S P S T D - - - - -
V A G T N Y R L V L Q V K D A S G A R K Y Q A A K V W E K P W E N F R K Q L S S E K L V K S - - - - -
V A G C M H Y F T I E V E K G G A K K L Y E A A K V W E K A W E N F K Q L S S E K P A A - - - - -
V A G V N V H L E L A D D T G K Q T V T V E S T W S R P W L A S K N - D A A Q P A T Q I T A L
V S G V Y Y I T L E A S E G D K K N Y E A A K V W E K P W L Q F K E L S S E K L L D D A S A - - - - -
V A G K L Y H L T L E A N D A G N K K I Y E A A K V V V K P W M N F K Q L S S E K H V E G K T S S
V A G I K Y Y L K I K T S T G V P K Y E A I V V V R P W H K P R Q L L N E S P S P A T K - - - - -
V A G T M H L L V I E V S G T E P R L Y E A A K V V V K P W E N - - - - - - - - - - - - - - -
V A G I N Y R L T L K A S G Q V K Y E A T V W E K P - D G T R O L T S S E K Q V - - - - -
V S G I K Y Y L K I A A R D G R E - R T F D A A K V V V K P W L K S R S L S S E T P - - - - -
V Q G A M Y Y I K L E A N D G G E K K T Y E A A K V V V K P W E N F K E L S S E Q R T K L V - - - - -
V S G M K Y Y L K I S A R N G V H E M F N S A V V V V K P W L H S K Q L L S S E A P A S S S T T
V A G K Y R L V I T G V S N G Q F S Y E A V V L D I P W E H S W K L L S S E K P - - - - -
V A G I N Y R L V I S A N K S V T S D Y E G I V W E K A W E H F R K Q L S S E H Q L L K F N - - - - -
V S G V K Y Y L K I S A V K S P K L Y E A A K V V V K S W E K F E L S S E F S P S P A T K - - - - -
V A G T N Y R L I L A A K D R A S S K Y E A A K V W E K P W E S F R N L S S E K - - - - -
V A G V N Y R L I T A K R L S A H R Y E A L V Y E R P W E H F M N L S S E K M I D Q Q E I N
V S G I N Y R L V L K A K D G S A V T Y E A V W E K A W L H F R N L S S E T L V M D - - - - -
V A G T N Y R L V L K N G S T A S Y Q A T V L E K P W L H F R N L S S E K P L R S - - - - -
V A G M N Y R L T V T A A D G G V E S S Y A F W N D K P W Q K F R Q L V S S E K P V - - - - -
V A G V R Y Y L T I E A R D G G K T G V Y A A E I V V K A W E N F K Q L S S E K F V Q T P H - - - - -
V S G T K Y Y L T I E V K D G A R D T K S F E A V W D K S W E H S R R L S S E K A V R G K A - - - - -
V A G T A Y Y I T L E A N G G Q R K Y E A A K V V K P W L D F K E L S S E K F I A D A - - - - -
V A G T Y Y L T I E A R D G G K K L Y E A A K V V V K P W M N F K E Q E R L V E G A P P E - - - - -
V S G I K Y Y L I L A A P L H N Q L L Y D A A K V V V K P W L N H R S I S S S T A T A V - - - - -
V A G T M H L L T V E A D G E T N K Y E A A K V L E K A W E N L K Q L S S E N H L H D V - - - - -
V A G V N Y L L T V E A D G E V T K Y E A L V W E K D W E G F R K L S S E K L I K E R E K
V S G M N Y K L I F V A A D G S G R V T Y A A Q V Y E H - W S R T R R L S S E R F A G G - - - - -
V A G T L H L L T L E A V D A G G K K Y E A A K V V V K P W M N F K Q L S S E K Y A E D V P A S
V S G L Y H L L T L E A R D G G K K N Y E A A K V V V K P W E N F K Q L S S E K L S S E K - - - - -
V S G T M Y H I T L E A R D G G K K V Y E A A K V V V K W E N F K Q L S S E K - - - - -
V A G V R Y Y L T I E A R D G G S K T G V Y A A E I V V K A W E N F K Q L S S E K F V Q T P H - - - - -
V A G M N Y R L V L A A S E G K R G K Y E A A V V V R P W L N E K K L S S E R S V - - - - -

```

266 Gmon_TnS000277535t10
267 Psom_XM_026526825
268 Ptae_FITA_000038429459
269 Kfed_Kaladp0666s0023
270 Vcar_Vocar_0049s0008
271 Tbac_TBAA0005436
272 Gmax_Glyma_15G115300
273 Ghir_Gohir.A09G208200.1
274 Tbac_TBAA00017565
275 Cpep_XM_023686782.1
276 Gmax_Glyma_07G266000
277 Atha_At5G47550_1
278 Pequ_XM_020719077
279 Sind_XM_011071778.2
280 Hann_HanXRQChr11g0344521
281 Hvul_HORVU2Hr1G124660
282 Tbac_TBAA0005316
283 Macu_GSMUA_Achr3T29820
284 Atri_evm_27_model1_scaffold00087_28
285 Hann_HanXRQChr04g0096491
286 Nnuc_XM_010246134.2
287 Acoe_Acoecoe4G277100
288 Atha_At4G16500_1
289 Stub_PGSC0003DMT400097644
290 Egra_Eucgr_L00642
291 Cpep_XM_023658118.1
292 Macu_GSMUA_Achr5T25320
293 Osat_LOC_Os03g31510
294 Egui_XM_019852036.2
295 Gmax_Glyma_05G149800
296 Hann_HanXRQChr02g0058381

```

```

V A G V M Y Y L T I M V K D S G G N S L Y E A K I W V K S W E N F R R L E D S P V N D T G G H
V A G I Q Y F L K V S A Q N G K P K L Y D A V V V V R S W Q P S N L L V S E S P F S S N - - -
V S G L L Y F L T I K T K D G S K T K L Y E A K V W V R F W A K F K Q L L Q E S K F V Q T L Y - -
V A G M M Y Y L T L E A M D T G G K K L Y E A K V W V R F W M N F K Q L Q E S K H A C E G R G S
V A G I L Y K M E L L E D E R G Q Q V L F T C S V W N R F W N T G Q N G L G H S N H I T K F H Y Q
V S G T N Y R L V L K A K D G S A T A S Y E A I V W E K F W L H F M N L L S E S K F S S S H A T N
V S G I K Y I L T I E G M E N G E K R I F N S V V L I K F W K G D K E L L S E S F S Q - - - -
V A G Y M Y Y L T M E V K K K E D L K I C K A Q V W E R F W L K Q M Q L L S E S K F L N A E A A A
V A G T M H H L T L E V V D A G K K K L Y E A K V W V R S W M N F K E L L Q D S K H A - - - -
V A G L N Y R L S L T A S D - - - S N N Y Q A I V Y E K A W R H Y R N L L T S E T F L H A - - - -
V S G T N Y R L K V A A N D G D G S K N Y L A I V W D K P W M K F R N L L T S E F P A N N G R F L
V A G T K Y R L L I V E A G G G G P A K Y E A I V W D Q P W K K F R Q L L S E S H K V N Y - - - -
V A G K M Y H L T L E A V D G G E K K M Y G A K I W E K F W M N F K Q L Q E S K Q R S - - - -
V S G M K Y M K I E L T K S G D P K V F E S V V V V K F W L R S K Q L L V K E G F S Q G V L R P
V S G M N Y E L I I D V T H F E G E G K Y K A E V Y E Q E W T K K R Q L L S E S T K V N - - - -
V A G T R Y Y L T L A V M D G A Q S R L Y M A V I Y V R P W Q N F K Q L L Q E S T E V K E - - - -
V S G T N Y R L V L S A S D G S S G S A K Y E A V V W E K F W E K F R Q L L T S E S K L V N S - - -
V S G I K Y Y L L I E A K D G G V L K K F D A V V V Q A W R H S K Q M L L S E S E S T D - - - -
V S G K N Y N L T I M A A D G G A S K N Y V A L V W D K P W Q K F R K L V A F E S K G F V - - - -
V A G M M H H L T L E V I E A G K K L C E A K V W V K F W M N F K E L L H E S F R L V G D A P T V
V A G V N Y R L V I A A K R D G K L N I Y L A V V Y V R A G G N K K A L L T S E S Q S L L R G V - -
V S G T K Y D L K I A A K D G G G I K N Y E A V V V E K L W L H S K S L L S E S F K A L - - - -
V E G I N Y R L V I E A K D G E G T H N Y L A V V W D K F - D K S K S L L T S E S F K Q L L E A - -
V A G I K Y Y L R I E G M Q R G S A R K V F E S V V V V R P W V R S K E L L V D F E S S A S - - -
V A G T K Y Y L R I L G S E N G R K R I F E S V V V V K F W L R S K Q L L I D F E S A S - - -
V S G I K Y Y L Q V V A V E G T G H R T F D A V V I V K P W L S S R S L L V S F A P R R P A L N S
V S G M N Y R L V V S A S D F A G T A S Y V A V V Y E Q S W T N T R Q L L T S E S F A P A A H - -
V A G T L H H L T L E V M E A G K K K L Y E A K L W V K F W L K F K E L L Q E S F Q - - - -
V S G I K Y M K I S A T Q G G D S R I F E S V V V V K F W L R S K Q L L N F E S T Q Y T Q E
V A G T K Y Y L K I E I N K G G K M K V F D A E V V V Q S W K H S K K L L G F E K F A F V D K - -

```

```

1 2 3 4 5 6 7 8 9
Variable Average Conserved

```

x - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

APÊNDICE 6 – Alinhamento das fitocistatinas do tipo II, com resíduos coloridos de acordo com a conservação.

Type-II phytocystatins MSA colored by conservation

001 Eves_Eruve_0015s0017	G	V	R	D	V	Q	N	S	D	E	I	E	S	L	A	R	F	A	V	D	E	H	N	K	K	E	N	A	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H	
002 Cmax_XM_023111212	G	V	H	D	S	Q	N	S	A	D	V	D	D	L	A	R	F	A	V	D	E	H	N	K	K	E	N	A	L	L	E	L	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	M	H	
003 Atha_AT5G05110_1	G	F	S	D	S	K	N	G	K	E	I	D	D	I	A	L	A	V	Q	E	H	N	R	R	E	N	A	V	L	E	L	A	R	V	L	K	A	T	E	Q	V	V	A	G	K	L	Y		
004 Vcar_Locus_644_Transcript_10_14	G	V	R	E	S	E	N	S	V	E	V	E	E	L	A	R	A	V	E	H	N	K	K	E	N	A	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H			
005 Tcac_Thecc_10G138800	G	V	H	P	S	H	Q	N	S	A	E	L	E	N	L	A	R	A	V	D	E	H	N	K	K	E	N	A	M	V	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H	
006 Ppin_PPI00045831	G	G	L	R	D	V	K	Q	N	S	I	E	T	L	D	L	G	R	A	V	D	E	H	N	K	Q	Q	N	G	D	V	S	F	R	R	V	V	A	K	E	Q	V	V	A	G	T	M	Y	
007 Egui_XM_010910725	G	D	V	R	D	S	Q	N	S	V	E	I	E	E	L	A	R	A	V	D	E	H	N	K	K	E	N	A	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H	
008 Afil_s0149_g053144	G	D	H	H	E	S	A	S	H	D	L	H	D	L	G	K	A	I	D	Q	H	N	S	R	Q	N	S	D	L	E	F	L	G	I	V	S	A	N	H	H	V	S	G	L	M	Y			
009 PvuI_PhuVI_011G196600	G	V	H	D	S	Q	N	S	L	E	T	E	S	L	A	R	A	V	D	E	H	N	K	K	Q	N	S	L	L	E	F	A	R	V	V	R	A	Q	E	Q	V	V	A	G	T	L	H		
010 Cvio_Clevi_0016s0139	G	D	V	R	G	S	A	W	N	S	A	E	I	E	A	I	A	R	A	V	E	H	N	R	R	E	N	A	L	L	E	L	A	R	V	I	K	A	T	E	Q	V	V	A	G	K	L	Y	
011 Tdic_TRIDC3Ag025860_MOD	G	K	S	E	N	P	A	N	S	L	E	T	D	G	L	A	R	A	V	D	E	H	N	K	R	E	N	A	L	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	V	H	
012 Drot_Dr09110	G	V	K	E	S	Q	N	S	V	E	I	E	D	L	A	R	A	V	E	H	N	K	K	E	N	K	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	V	H			
013 Vang_KOM28344	G	V	R	D	S	Q	N	S	V	E	I	E	S	L	A	R	A	V	D	E	H	N	K	K	Q	N	S	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	V	H		
014 Spol_Spipo3G0066400	G	V	S	D	V	E	Q	N	S	A	E	V	E	S	L	A	R	A	V	E	H	N	K	K	E	N	A	V	L	L	E	F	S	R	V	V	K	A	K	E	Q	V	V	A	G	S	I	L	H
015 Alin_Alyli_0556s0004	G	F	S	D	S	K	N	G	K	E	I	D	D	I	A	L	A	V	Q	E	H	N	R	R	E	N	A	V	L	E	L	S	R	V	L	K	A	T	E	Q	V	V	A	G	K	L	Y		
016 Pvir_Pavir_5NG183500	G	V	K	E	N	P	A	N	S	A	E	S	D	G	L	G	R	A	V	E	H	N	K	R	Q	N	A	L	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
017 Nnuo_XM_010246133	G	I	R	E	S	V	E	N	S	V	E	I	E	S	L	A	R	A	V	D	E	H	N	K	K	Q	N	A	L	L	E	F	A	R	I	V	V	K	A	K	E	Q	V	V	A	G	T	L	H
018 Grai_Gorai_011G221800	G	V	H	P	S	Q	N	S	V	E	L	E	N	L	A	R	A	V	D	E	H	N	K	K	E	N	A	M	V	E	F	V	R	V	V	K	A	S	E	Q	V	V	A	G	T	L	H		
019 Cole_Ciclev10012693m	G	V	Y	D	Y	Q	N	S	A	E	I	E	G	L	A	R	A	V	Q	E	H	N	R	R	E	N	A	V	L	L	Q	F	A	R	V	L	K	A	T	E	Q	V	V	A	G	K	L	Y	
020 Csat_Cucsa_076380	G	D	V	R	D	S	Q	N	S	A	D	V	D	E	L	A	R	A	V	E	H	N	K	K	E	N	S	L	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
021 Ugih_749_g14102_mod	G	L	R	D	S	E	-	T	D	Q	V	E	S	L	A	Q	A	V	D	Q	H	N	N	K	E	N	A	L	L	Q	L	V	R	V	T	K	S	Q	E	Q	V	V	A	G	K	I	H		
022 Atha_AT3G12490_2	G	V	G	D	V	Q	N	S	G	E	V	E	S	L	A	R	A	V	D	E	H	N	K	K	E	N	A	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
023 Vrad_Vradi06g08980	G	V	R	D	S	Q	N	S	L	E	T	E	S	L	A	R	A	V	D	E	H	N	K	K	Q	N	S	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
024 Camp_Caamp_1038s0832	Q	G	T	Y	N	D	M	K	L	G	E	I	D	D	I	A	L	A	V	Q	E	H	R	R	E	N	A	V	L	E	L	A	R	V	L	K	T	E	Q	V	V	A	G	K	L	Y			
025 Pvir_Pavir_5KG180700	G	V	K	E	N	P	A	N	S	A	E	S	D	G	L	G	R	A	V	E	H	N	K	R	Q	N	A	L	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
026 Nben_101Scf21813g00014_mod	G	I	R	D	S	H	H	N	S	V	E	I	H	S	L	A	K	A	V	D	E	H	N	K	K	E	N	A	M	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
027 Cmax_Glyma_13G189500	G	L	H	D	S	Q	N	S	V	E	T	E	A	L	A	R	A	V	D	E	H	N	K	K	Q	N	S	L	L	E	F	A	R	V	V	R	A	Q	E	Q	V	V	A	G	T	L	H		
028 Desop_Desop_0227s0663	G	V	R	D	V	Q	N	S	G	E	V	E	S	L	A	R	A	V	D	E	H	N	K	K	E	N	V	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H			
029 Ptri_Potri_003G192200	G	V	H	D	S	Q	N	S	A	E	I	D	S	L	A	R	A	V	D	E	H	N	K	K	E	N	A	I	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	M	H		
030 Egra_Eucgr_J00085	G	V	H	E	S	C	E	N	S	V	E	V	E	S	L	G	R	A	V	D	E	H	N	S	K	E	N	G	L	L	E	F	A	R	V	V	K	A	H	E	Q	V	V	A	G	T	L	H	
031 Pper_Prupe_2G047100	G	V	H	E	S	H	Q	N	S	L	E	T	E	D	L	A	R	A	V	D	H	N	N	K	E	N	A	L	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
032 Sind_AF240007	G	V	H	D	S	N	-	S	N	F	D	T	H	S	L	A	R	A	V	D	Q	H	N	T	K	E	N	G	L	L	E	L	V	R	V	V	E	A	R	E	Q	V	V	A	G	T	L	H	
033 Ksub_FQIP_scaffold_2007794	G	Q	V	D	V	G	S	N	D	P	D	V	E	V	G	E	-	V	S	T	I	A	K	E	P	-	-	D	L	K	F	V	R	V	V	E	A	T	K	Q	V	V	A	G	L	M	W		
034 Salb_Sialb_0095s0048	G	V	R	D	V	Q	N	S	D	E	V	E	S	L	A	R	A	V	D	E	H	N	K	K	E	N	A	L	L	E	F	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	M	H		
035 Phal_Pahal_5G416800	G	V	K	E	N	P	A	N	S	A	E	S	D	G	L	G	R	A	V	E	H	N	K	R	Q	N	A	L	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
036 Salb_Sialb_0879s0021	G	V	R	D	V	Q	N	S	G	E	V	E	S	L	A	R	A	V	D	E	H	N	K	K	E	N	A	L	L	E	F	G	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
037 Cmax_XM_023130994	G	V	H	D	S	Q	N	S	A	D	V	D	E	L	A	R	A	V	D	E	H	N	K	K	E	N	A	L	L	E	F	V	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
038 Ptri_Potri_016G030900	G	V	H	D	C	K	Q	N	S	A	E	I	D	S	L	A	R	A	V	D	E	H	N	K	K	E	N	A	I	L	E	F	V	R	V	L	K	A	T	E	Q	V	V	A	G	K	L	Y	
039 Nben_101Scf12983g01058_mod	G	V	R	D	S	H	Q	N	D	E	I	H	N	L	A	K	A	V	D	E	H	N	K	K	E	N	A	M	I	E	L	A	R	V	V	K	A	K	E	Q	V	V	A	G	T	L	H		
040 Cvul_MWXT_scaffold_2004286	G	L	T	P	Q	N	A	N	D	L	S	I	Q	I	A	E	A	V	K	E	H	N	K	Q	R	V	E	E	V	R	L	G	I	L	S	A	A	S	Q	V	V	A	G	I	M	Y			
041 Cymi_CMI00005557	G	E	I	R	D	V	E	Q	N	S	E	I	L	G	L	G	Q	A	V	D	E	F	N	K	Q	Q	N	A	H	L	S	F	K	K	V	V	R	A	R	Q	V	V	A	G	I	M	R		
042 Bole_Bol015400	G	Y	R	D	S	T	N	G	G	E	I	D	G	L	A	L	A	V	Q	E	H	N	K	R	E	N	G	V	E	L	V	R	V	L	K	A	T	E	Q	V	V	A	G	K	L	Y			
043 Lann_Luann_0004s0723	G	F	S	D	S	T	V	S	E	I	D	D	I	A	L	A	A	V	Q	E	H	N	M	R	E	N	A	V	L	E	L	A	R	V	L	K	A	T	E	Q	V	V	A	G	K	L	Y		
044 Cqui_AUR62032567	G	V	L	R	E	S	Q	N	D	A	E	I	E	N	L	A	R	A	V	E	H	N	K	K	E	N	G	M	L	E	F	V	R	V	I	K	A	K	E	Q	V	V	A	G	T	L	H		
045 Aocc_Anaoc_0010s1343_1	G	V	H	N	C	Q	N	S	A	E	I	E	S	L	A	R	A	V	E	H	N	K	K	E	N	A	L	L	E	F	S	R	V	L	K	A	K	E	Q	V	V	A	G	M	M	Y			
046 Scuc_s0001_g000348	G	R	P	H	D	I	S	H	D	D	P	Y	L	R	D	L	A	S	A	V	D																												

091 Ahal_Araha_21141s0001
092 Bstr_Bostr_13129s0454
093 Ocho_Oropetium_20150105_09645A
094 Bnap_CDY08430
095 Mper_Myper_0043s0741
096 Euni_k31_146807_2
097 Thac_TBA0007041
098 Pdac_XM_008796910
099 Ppat_Pp3c15_24180V3
100 Lesa_Lesat_0046s0160
101 Pinf_101scf01482g02033
102 Lang_OIV89378
103 Crub_Carub_0006s0404
104 Mgut_Migut_H00005
105 Gsoj_GlysoPI483463_136151700
106 Zmar_Zosma94g0120
107 Pequ_XM_020724410
108 Aktr_CCID_scaffold_2009820
109 Cmar_Camar_0377s0027
110 Ppat_Pp3c9_24080V3
111 Omer_OMER101610740
112 Bdis_Bradi2g10146
113 Clan_22909
114 Tarv_Thlar_0016s0705
115 Ntab_0009278g0010_mod
116 Ppat_Pp3c8_17580V3
117 Mtru_Medtr8g041410
118 Grai_Gorai_0106215400
119 Esal_Thhalv10021442m
120 Paxi_162scf00516g00518
121 Dstr_Distr_0241s0826
122 Lper_LPERR01610160
123 Cara_Scaffold_637_1450
124 Sfal_Sphfalx0005s0315
125 Macu_GSMUA_Achr3T18480
126 Alin_Alyli_0149s0067
127 Lusi_Lusi0026117
128 Mdom_MD02G1275000
129 Lesa_Lesat_0165s0505
130 Alin_Alyli_0180s0053
131 Taes_CS3002G189800
132 Brap_Brapa_C00203
133 Cus_JoJo_scaffold_2007221
134 Mend_ME000129s00342
135 Brap_Brapa_A03337
136 Eves_Eruve_0771s0041
137 Itin_Isati_10464s0003
138 Efm_BFIK_scaffold_2025023
139 Pabi_MA_4477g0010
140 Mdom_MD07G1043000
141 Esyr_Eusy_0130s0928
142 Bhyb_Brahy_s0160313600
143 Gmax_Glyma_156227500
144 Cqui_AUR62031684
145 Vvin_VIT_208s0040g00770
146 Tint_Thint_0960288600
147 Stub_PGSC0003DMT40068064
148 Mnar_Mamar_0016s0858
149 Cmar_Camar_0168s0030
150 Stub_PGSC0003DMC400045978
151 Sfal_Sphfalx0093s0081
152 Lann_Luann_0020s0090
153 Pqla_PGL0010280
154 Chra_BFEA010003241
155 Psom_XM_026571053
156 Nnuc_XM_010246848
157 Itin_Isati_0492s0004
158 Spur_Sapur_0036148300
159 Chis_Crahi_1249s0012
160 Bvul_KMT17034
161 Casi_28243420
162 Cmar_Camar_0113s0027
163 Osat_LOC_Os01g16430
164 Camp_Caamp_0078s1100
165 Casi_028246879_mod
166 Sbic_Sobic_0036126800
167 Pdru_ZZEI_scaffold_2126917
168 Msin_MisinT267700
169 Psit_PSI00017843
170 Pdel_Podel_016035500
171 Msin_Misin06g106700
172 Ghir_A06g173900
173 Pmil_PQIB02000010
174 Cpep_pep_XM_023692654
175 Ahyr_AH015638
176 Zygn_MFZO_scaffold_2022157
177 Lang_OIW08132
178 Chis_Crahi_1482s0007
179 Cvio_Clevi_0011s0507
180 Kfed_Kaladp0131s0016
181 Slyc_03g097270
182 Opun_OPUNC01610650
183 Ghir_A10G190300
184 Mnar_Mamar_0004s0409
185 Vvin_VIT_213s0084g00050

G V R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T M H
G P G D S F N G G E D I D D I A L A R F A V D E H N K K R Q N A L L E F A R V V K A K E Q V V A G T L H H
G A K Q K F A N S A E A D G L A R F A V D E H N K K R Q N A L L E F A R V V K A K E Q V V A G T M H H
G V R D V F Q N S D E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G V H E S R E N S V E V E T L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G K R A V E Q S V E I D E L A R F A V D E H N T T Q Q N A L L E F A R V V K A K E Q V V A G T V Y
G V R D S Q N S L E I D E L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L Y
G K K E V D T N S L E I D E L A R F A V D E H L K S Q N S G M T F S K V V S C R K Q V V Q G T M Y
G V R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V S G T L H
G I R D S H Q N S D E I H S L A R F A V D E H N K K E N A M F E L A R V V K A K E Q V V A G T L H
G I T D S Q N T L E I Q D L A R F A V D E H N N K Q N A L L E F A R V L V K A K E Q V V S G T L H
G L S G S M N G G E I D D I A L A R F A V D E H N N R E N A L L E F A R V L V K A K E Q V V A G K L Y
G L K H T N - - T E I D S L A R F A V D E H N N K E N A L L E F A R V V N A K E Q V V A G S L H
G L H D S Q N S V E T E A L A R F A V D E H N K K Q N S L L E F A R V V V A Q E Q V V A G T L H
G V F D S Q N S A E L A D L A I R F A V D E H N N K E N T L L E F A R V V K A K E Q V V A G K L H
G V R E S K A N S V E I D E I G R F A V D E H N K K E N A T L E F A R V V K A K E Q V V S G T L H
G I R D S Q N S A E I D E I A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G A F S D Y G G G E I D D I A L A R F A V D E H N K R E N G V V E L A R V L K A K E Q V V A G K L Y
G K Q E V D S N N G L E I D E L A R F A V D E H N D R E N S K L T F S K V V S C H M Q V V A G S M Y
G A H D A P A N S V E T D A L A R F A V D E H N K R E N A L L K F F V R V V E A K E Q V V A G T L H
G K R P S Q N S L E T D G L A R F A V D E H N K R E N A L L E F V R V V K A K E Q V V A G T L H
G V H D S Q N S C V D V D E F A R F A V D E H N K K E N S L L E F V R V V K A K E Q V V A G T L H
G V R D S Q N S D E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T M H
G I R D S H Q N S D E I H N L A R F A V D E H N K K E N A M I E L A R V V K A K E Q V V A G T L H
G K K E V D T N S L E I D S L A R F A V D E H N A R Q N S K M N F S K V V S C H K Q V V S G L M Y
G I Q D S F E N S L E I E S L A R F A V D E H N A K Q N S L L E F A R V V K A K E Q V V A G T L H
G F R P S Q N S A E L A D L A I R F A V D E H N K K E K G M V E F V R V V K A K E Q V V A G T L H
G V R D V F Q N S D E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G I R D S H Q N S D E I H S L A R F A V D E H N K K E N A M F E L A R V V K A K E Q V V A G T L H
G V R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F G R V V K A K E Q V V A G T L H
G A R D V P A N S A E T D S L A R F A V D E H N K R E N A L L E F V R V V E A K E Q V V A G T L H
G L H D S Q N S S E I D A L A R F A V D E H N K K E N A L L E F V R V V E A R E Q V V A G T L H
G K P K H V E S N D L E V D L A N A I Q E H N K K E N T N T K L T F S K V V S A K K Q V V Q G T I Y
G L K E S Q N S A E I E E L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V T G T L H
G V R D V F Q N S G E V E S L A R F A V D E H N K K E N V L F E F V R V V K A K E Q V V A G T L H
G V H E S P S N S A S T E G L A R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G I H E S Q N S A E V D L A R F A V D E H N K K E N A L L E F V R V V K A K E Q V V A G T L H
G F S D S N G G G E I D D I A L A R F A V D E H N K R E N A V L E L A R V L K A K E Q V V A G K L Y
G P S D S K N G G A E I D D I A L A R F A V D E H N K R E N A V L E L S R V L K A K E Q V V A G K L Y
G K S E N P A N S L E T D G L A R F A V D E H N K R E N A L L E F V R V V E A K E Q V V A G T L H
G V R D S N G G G V I D D I A L A R F A V D E H N K R E N G V V E L A R V L K A K E Q V V A G K L Y
G K S P M K E N D L E V Q D L G R F A V S E H N S K Q N G N L E F L K V V A A Q Q Q V V A G M M Y
G N K P E K Q N D L E V Q L G R F A V S E H N S K E N A G L S F G K V V A A K E Q V V Q G K M H
G V R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T M Y
G F S D S N G G G E I D D I A L A R F A V D E H N K R E N G I L L E L A R V L K A K E Q V V A G K L Y
G V R D V F Q N S D E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L Y
G A P V D V R N D F S H E E I A R F A V E K L A E Q E N S S L Q F V R L K V T S Q V V A G L I Y
G L R D V Q N S I E T D L D L G R F A V D E H N K Q Q N G D I S F R R V V A A Q E Q V V A G T M Y
G I Q E S Q N S A E V D L A R F A V D E H N K K E N A L L E F F R V V K A K E Q V V A G T L H
G V S D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G K R E S P A N S V E T D G L A R F A V D E H N K R E N A L L E F V R V V E A K E Q V V A G T L H
G L R D S Q N S V Q T E A L A R F A V D E H N K K Q N S L L E F S R V V R T Q E Q V V A G T L H
G L R E S Q N S A E I E N L A R F A V D E H N K K E N G M L E F A R V V I K A K E Q V V A G T L H
G V R D C Q N S A E I E S I A R F A V D E H N K N Q N A L L E F A K V L K A K E Q V V A G K I Y
G K S E N P A N S V E T D G L A R F A V D E H N K R E N A L L E F V R V V E A K E Q V V A G T L H
G V H D S H Q N S D E I H S L A R F A V D E H N K K E N A M I E L A R V V K A K E Q V V A G K L H
G V R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G V H D S H Q N S D E I H S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G V H D S H Q N S D E I H S L A R F A V D E H N K K E N A M I E L A R V V K A K E Q V V A G K L H
G I K E V S N S L E M D L A R F A V D E H N S R G N S K L S F S K V V S A R T Q V V Q G T M Y
G I R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F G R V V K A K E Q V V A G T L H
G L R D V Q N S I E T D L D L G R F A V D E H N K Q Q N G D I S F R R V V A A Q E Q V V A G T M Y
G L T P Q N A N D L S M Q E I A Q V A V N E H N K Q R A E E V T L I G I L S A A S Q V V A G I M Y
G V R D S Q N S L E I E D L A R F A V D E H N K K E N A L L E F V R V V K A K E Q V V A G A M H
G V H D S Q N S V E I E S L G R F A V D E Y N K K Q N A L L E F G R V V E A K E Q V V A G T L H
G F S D S T N G G G E I D D I A L A R F A V D E H N R R E N A V L E L A R V I K A K E Q V V A G K L Y
G V H D S Q N S A E I D S L A R F A V D E H N K K G N A I L E F A R V V K A K E Q V V A G T M H
G V R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G L R E S Q N S A E I E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G I H D C M Q N S A E I E S I G R F A V D E H N K K E N A L L E F A R V V K A K E Q V V S G K M Y
G V R D V F Q N S D D V E S L A R F A V D E H N K K E N A L L E F G R V V K A K E Q V V A G T L H
G A H D A P A N S V E T D A L A R F A V D E H N K R E N A L L E F V R V V E A K E Q V V A G T L H
G V R D V F Q N S D E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G I H D S H Q N T A D I D N L A R F A V D E H N K K E N T L F E F V R V V K A K E Q V V A G T L H
G V K E N P A N S A E S D G L G R F A V D E H N K K E N A L L E F V R V V E A K E Q V V A G T L H
G L K E V H N D L E A Q D L A R F A I A Q H N T R Q N S D L E Y A G I V S A H K Q V V T G L M Y
G V K D N P A N S A E S D G L G R F A V D E H N K R E N A L L E F V R V V D T K Q V V A G T L H
G L R D V Q N S I E T D L D L G R F A V D E H N K Q Q N G D I T F R R V V A A Q E Q V V A G T M Y
G V H D S Q N S V E I E N L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G K E N P A N S A E S D G L G R F A V D E H N K R E N A L L E F V R V V E A K E Q V V A G T L H
G F R P S Q N S G E L E N L A R F A V D E H N K K E K R M V E F V R V V K A T E Q V V A G T L H
G V K E N P A N S A E S D G L G R F A V D E H N K R Q N A L L E F V R V V E A K E Q V V A G T L H
G V H D S Q N S A D V D E L A R F A V D E H N K K E N S L L E F V R V V K A K E Q V V A G T L H
G L R E S Q N D A E I E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G L S D V K E N D L E L Q E V G R F A V S E H N S K Q N A N L E F V K V V A A Q Q V V A G M M Y
G I V N S Q N S L E I Q D L G H F A V S E H N K K Q N A L L E F T R V V K A K E Q V V A G T L H
G V R D V F Q N S D E V E S L A R F A V D E H N K K E N A L L E F A R V V K A K E Q V V A G T L H
G V R D V F Q N S G E V E S L A R F A V D E H N K K E N A L L E F A R V L V K A K E Q V V A G T L H
G V R E S Q N S V E I D D L A R F A V D E H N K K E N A V L E F A R V V K A K E Q V V A G T L H
G V H D S H Q N S D E I H S L A R F A V D E H N K K E N A M I E L A R V V K A K E Q V V A G K L H
G A R D A P A N S A E T D A L A R F A V D E H N K R E N A L L E F V R V V E A K E Q V V A G T L H
G V H P S Q N S V E I E N L A R F A V D E H N K K E N A M V E F V R V V K A S E Q V V A G T L H
G V S D S T N G E V I D D I A L A R F A V E H N K K Q D A V L E L A R V L K A K E Q V V A G K L Y
N I L F L S K K G K V E D M R R L M I L Q N Q I A F N A L L E F A R V V K A K E Q V V A G T L H

186 Lusi_Lusi0008702
187 Bsta_Brast01G301000
188 Taes_CS3A02G185800
189 Sspo_03G0034930_1B
190 Abux_sb22706_1
191 Knit_kf100437_10
192 Acoe_Aqcoe1G006500
193 Cmos_XM_023089349
194 Alyr_AL3G24120
195 Spir_HAOX_scaffold_2024011
196 Spen_03g028040
197 Cole_Ciclev1002346m
198 Bstr_Bostr_3640s0021
199 Hann_HanXRCChr04g0119231
200 Chis_Crahi_0009s0246
201 Itin_Isati_4568s0011
202 Atri_AmTr_v1_0_scaffold00101_77
203 Doli_LMDX02072575
204 Ghir_D10G198100
205 Acoe_Anaoc_0016s1440
206 Ljap_Lj1g0013483_1
207 Slyd_LYD0_6ch03_mod
208 Mper_Myper_0023s0324
209 Brap_Brara_E02850
210 Olae_GGWH_scaffold_2007720
211 Cmos_XM_023101466
212 Spur_Sapur_001G026900
213 Coit_Cocit_J0835
214 Crub_Carub_0003s1182
215 Pdel_Podel_03g209600
216 Cang_03g001644
217 Pmen_PME00047088
218 Aoff_Asparagus_V1_07_1767
219 Mnot_L484_024003
220 Tarv_Tilar_0026s0267
221 Cmel_XM_008447722
222 Room_29686_m000868
223 Csin_Cs6g16640_1
224 Even_MKKT01000068
225 Psom_XM_026542585
226 Oeur_Oeu028986
227 Spim_03g097270
228 Sita_Seita_5G046700
229 Mcha_XM_022283450
230 Gmon_TnS000729105t05
231 Cpep_XM_023686782
232 Even_AMZH03000833
233 Pdru_ZZEI_scaffold_2022423
234 Egui_XM_010933246
235 Chre_YOXI_scaffold_2051387
236 Camp_Caamp_0105s1619
237 Eves_Eruve_2362s0018
238 Cmar_Camar_0113s0026
239 Atau_AET3Gv20420400
240 Gbil_GBI00017832
241 Taes_CS3B02G215400
242 Slyc_Solyc03g097270
243 Pdac_XM_008788651
244 Pmen_PME00028125
245 Pdel_Podel_16G033700

G T H D S F S N S A E I E D L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G K R E S F A N S S V E T D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G K S E N F A N S S L E T D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G V K D N F A N S S A E S D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V R E S F E N S S I E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G Q V D V G S N S D P D A E E V L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V T D S G S N S V E I E E L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V H D S G S N S A D V D E L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V R D S G S N S G E V E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G I F D V K E S D P A I Q E L A L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V H D S H Q N S D E I H S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V Q E S R E N S S N E I E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V R D S G S N S G E V E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V R N S S P N T D E I H G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G F S D S S N D G G E I D A I A L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V R D V P Q N S D E V E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G L K D V T Q N S A E I E E L A R R F A V D E H N K K S N S L L E F F A R V V K A K E Q V V A G T L Y
G G V K E S R E N S S A E S D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G V H P S Q N S V E I E N L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G L R E S C E N S S V E I E L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G A R D A T P D G A A I E N L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G V H D S H Q N S D E I H S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G F S D S S N D G G V I D D I A L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G V R D V P Q N S D E V E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G A K P V D A N D A G I A E S L A R R F A V D E H N K K E N S D L K L L K V V S A T K E Q V V A G T L Y
G G V H D S G S N S A D F D E L A R R F A V D E H N K K E N S L L E F F A R V V K A K E Q V V A G T L Y
G G V H D S G S N S V E I D S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G V H S R E N S S V E I E L A R R F A V D E H N S K E N G L L E F F A R V V K A K E Q V V A G T L Y
G G V R D V P Q N S G E V E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G V H D S Q N S A E I D S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G I H T S H E N S D E I R S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G L R E V K Q N S I E T L D L A R R F A V D E H N K K Q Q N G D I T F R R V V G A Q E Q V V A G T M Y
G G I H Q R S E N S A E I E E L A R R F A V D E H N K K Q N A L L E F F S R V V L K A K E Q V V A G I M H
G G V R Q S R E N S A D E V D L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G F S D S M N D R A A V D D I A L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G V H D S G S N S T D V D E L A R R F A V D E H N K K E N S L L E F F A R V V K A K E Q V V A G T L H
G G V H D S P A N N A E I D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G V Y D Y Q N S A E I E G L A R R F A V D E H N K K E N A L L Q F A R V V L K A K E Q V V A G T L Y
G G L K E S E Q N S A E I E E L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G N V S D S S F K S Q I S L G L A R R F A V D E H N K K I E N A V L E F F R I G K A K E Q V V A G S A V H
G G L R D S H Q N S L E I D T L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G V H D S G S N S D E I H S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G V K E N F A N S S A E S D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G V H D S G S N S A G V E E L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G K Q E I V Q N N R E I A D L A R R F A V D E H N K K Q N G A L S F N R V V K A Q E Q V V A G L A Y
G G V H D S G S N S A D V D E L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G L K E S E Q N S A E I E E L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G A R H K I T Q N D L Q L E N L A R R F A V D Q H N A K Q N S S L V Y V R L L S A E Q V V A G I M Y
G G V H D A K E N S V E I E E L A R R F A V E Q N K K A N T L L E F F V R V V K A K E Q V V A G S I Y
G I S D V K Q N S L E M T E L A Q R F A V S E H N K K K Q N S D L K F V R V V K A K E Q V V A G L M Y
G G F S D S N D G G E I D D I A L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L Y
G G V R D V P Q N S D E V E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T M H
G G V V D V N R N S S E F E S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G M L Y
G G K S E N F A N S S L E T D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G I K D V P Q H D F I Q S L A R R F A V D E H N K K Q N I L L S F S K V V K A Q E Q V V A G M I Y
G G K S E N F A N S S L E T D G L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G V H D S H Q N S D E I H S L A R R F A V D E H N K K E N A V V E F L R V V E A K E Q V V A G T L H
G G V H D A K E N S V E I E E L A R R F A V D E H N K K A N T L L E F F V R V V K A Q E Q V V A G I Y
G G L R E V K Q N S I E T L D L A R R F A V D E H N K K Q N G D I T F R R V V G A Q E Q V V A G T M Y
G V H D C K N S A E I D S L A R R F A V D E H N K K E N A I L E F F R V V L K A K E Q V V A G T L Y

001 Eves_Eruve_0015s0017
002 Cmax_XM_023111212
003 Atha_AT5G05110_1
004 Vcar_Locus_644_Transcript_10_14
005 Toac_Thecc_10G138800
006 Ppin_PPI00045831
007 Egui_XM_010910725
008 Afil_s0149_g053144
009 Pvil_Phvil_011G196600
010 Cvio_Clevi_0016s0139
011 Tdic_TRIDC3Ag025860_MOD
012 Drot_Dr09110
013 Vang_KOM28344
014 Spol_Spipo3G0066400
015 Alin_Alyli_0556s0004
016 Pvir_Pavir_5NG183500
017 Nnuc_XM_010246133
018 Grai_Gorai_011G221800
019 Cole_Ciclev10012693m
020 Csat_Cuca_076380
021 Ughl_749_g14102_mod
022 Atha_AT3G12490_2
023 Vrad_Vradi06g08980
024 Camp_Caamp_1038s0832
025 Pvir_Pavir_5KG180700
026 Nhen_101Scf21813g00014_mod
027 Gmax_Glyma_13G189500
028 Dsop_Desop_0227s0663
029 Ptri_Potri_003G192200
030 Egra_Eucgr_J00085
031 Pper_Prupe_2G047100

H L T L B I I E A G K K K L Y E A K V V V K P W L N F K E L Q E B K P S T I T P S D L G C K K D E R
H L T L B V V D A G K K K L Y E A K V V V K S W M N F K E L Q E B K H P S I T H S D L G A K K G D H
R L T L B V I E A G K K K L Y E A K V V V K P W M N F K E L Q E B K N P S I T T A D L G K P K D G N
H L T L B V I D A G K K K L Y E A K V V V K P W L N F K E L Q E B R H P S I T T A D L G A K R D G H
H L T V B A I D A G K K K L Y E A K V V V K P W M N F K E L Q E B K H P S I T A S D L G V K K D V H
H L T I B A E D G D K P K L Y T A K V V V K P W E N F K S L Q E B K P S I T A S D L G V K Q H A E
H L T L E V M E A G K K L F E A K V V V K P W L N F K E L Q E B T H P S V T A A D L G A K H G G H
H L K L K A S S G G Q T N L Y E A K V V V K P W L N F K S L E E B K R S S F T S A D L G A R I E P R
H L T L E V E A G K K K L Y E A K V V V K P W L N F K E L Q E B K P S P F T S A D L G V K K D G H
R L T L E V I E A G T K K V Y E A K V V V K P W M N F K E L Q E B K H R S F T T L D L G S K P D G N
H L T L E A L E A G K K K L Y E A K V V V K P W L D F K E L Q E B R H S T T I S D L G A K R G G H
H L T V E A I D A G T K K L Y E A K V V V K P W L N F K E L Q E B K H P S L T A S D L G V K K D G H
H L T L E A I E A G K K K L Y E A K V V V K P W L N F K E L Q E B K P S P F T S A D L G V K K D G H
H L T V E I E E G K K K L Y E A K V L V Q A W L N S K K L H E N P S P F T P A D L G A K R D P L
R L T L E V I E A G K K K L Y E A K V V V K P W M K F K Q L Q E B K N P S F S I S D L G L K P D S N
H L T L E A I E A G R K K V Y E A K V V V K P W L D F K E L Q E B V H T T F T S A D L G A K I G G H
H L T L E V I E A G K K K L Y E A K V V V K P W L N F K E L Q E B K P F T F T S A D L G A R R D G H
H L T V E A V D A G K K K L Y E A K V V V K P W M N F K E L Q E B K H A S F T A C D L G A K R D V Q
Y L T L E V I D A G K K K L Y E A K I V V K P W M N F K E L Q E B K H P S I T A S D L N L K R G C H
H L T V E V E A G K K K L Y E A K V V V K S W M N F K E L Q E B K H P S I T S D L G A K K G D H
H L T L E V I D A G K K K L Y E A K I V V K P W M N F K E L Q E B K H P S S V S T E H D A N K D E E
H L T L E I L E A G Q K K L Y E A K V V V K P W L N F K E L Q E B K P P A I T S A D L G C Q G E H
H L T L E A I E A G K K K L Y E A K V V V K P W L N S K K L Q E B K P S P F T S A D L G V K K D G H
R L T L E V I E A G K K K L Y E A K V V V K P W M N F K Q L Q E B K N P S F T T S D P G L K S D G N
H L T L E A I E A G R K K V Y E A K V V V K P W L D F K E L Q E B V H T T F T S A D L G A K K G G H
H L T L E V I D A G K K K L Y E A K V V V K P W L N F K E L Q E B T H P T L T S A D L G V K Q G E E
H L T L E A I E A G K K K L Y E A K V V V K P W L N F K E L Q E B K P S P F T S A D L G V K K D G H
H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E B K D P V F T S S D L G V K R D G H
H L T I E A V D A G K K K L Y E A K I V V K P W M N S K E L Q E B K H P A F T S S D L G A R I G G H
H L T I E A I D A G K K K L Y E A K V V V K P W L G F K E L Q E B K H P S F T P S D L G V K E G G H

032 Sind_AF240007 H L V L E V L D A G K K K L Y E A K I V V K P W M D F K Q L Q E K H P S F T S S D L G A K T D D Q

033 Ksub_FQIP_scaffold_2007794 R L T L E V K T F V G P A B L Y E A T V Y Q E G P D A T R S E L L P S D L G A H L G G A K T D D Q

034 Salb_Sialb_0095s0048 H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S V H T T F T N A D L G A K K G G H

035 Phal_Pahal_56416800 H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S V H T T F T N A D L G A K K G G H

036 Salb_Sialb_0879s0021 H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S V H T T F T N A D L G A K K G G H

037 Cmax_XM_023130994 H L T V E A D A G N N K M Y E A K V V V K S W M N F K E L L Q E S K H P S F T S S D L G A K K G G H

038 Ptri_Potri_016603090 H L T V E A D A G N N K M Y E A K V V V K S W M N F K E L L Q E S K H P S F T S S D L G A K K G G H

039 Nben_101Scf12983g01058_mod H L T L E V I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S T H P T L T S S D L G A K K G G H

040 Cvul_MWXT_scaffold_2004286 H L T L E I G T E N G H E K F L A K V V E Q A W R N S T E K L R T T A D L G A H R K G T R

041 Cymi_CM100005557 Y L T I E A V D G D Q T K I Y G A K L L T G G S Q N I Q R L L Q E S K A E H D D T F Y A G T Q E D I E

042 Bole_Bo1015400 S L T L E V E I E A G E K K I Y E A T V V V K P W M N F K Q L Q E S K N P S F T V S D L G L K S D G N

043 Lann_Luann_0004s0723 H L T L E V E I E A G E K K I Y E A K V V V K P W N F K Q L Q E S K N P S F T I S D L G S K S D G N

044 Cqui_AUR62032567 H F T I E A I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S K H A A F S T S D L G A N R E G H

045 Aocc_Anaoc_0010s1343_1 H L T L E A I D A G K K K L Y E A K V V V K P W M N F K E L L Q E S K H P S F T S S D L G A K H D G H

046 Scuc_s0001_g000348 Q L T L K A S S G G Q T R I Y E A K I V V K P W E N F K S L Q E S K M S A F T S A D L G S E I Q P E

047 Cgra_Cagra_5641s0012 H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S K P S A I T P S D L G C K K D E H

048 Svir_Sevir_56044900 H L T L E A E A G R K K V Y E A K V V V K P W L D F K E L L Q E S V H T T F T N A D L G A K K G G H

049 Oglu_OGLUM01612280 H L T L E A E A G R K K L Y E A K V V V K P W L N F K E L L Q E S R H T T F T N A D L G A K K G G H

050 Sacc_SCC59_SCVFPL4C11C01 H L T L E A E A G R K K L Y E A K V V V K P W L D F K E L L Q E S S H T T F T N A D L G A N K C G H

051 Bnap_CDX80942 S L T L E V K E A G E K K L Y E A K V V V K P W M N F K Q L Q E S K S P S F T V S D L G L K S D G N

052 Bole_Bo1010294 H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S K P S S I T P S D L G C K K G E H

053 Risl_Roisl_0074s0761 H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S K P P A I T P S D L G C K K D E H

054 Cesc_AF525880 H L T V E I E G G K K K V Y E A K V V V Q A W L N S K K L H T S P S S V T P A D L G V K R D A H

055 Csub_49027 K L L L D I A N D G E K K C V E A T V Y A P L G N Q E K Q L T S T R K Q V E Q Q T E A T S E R N G G

056 Camp_Caamp_1039s0467 H L T L E I E A G R K K L Y E A K V V V K P W L N S K E L L Q E S K P P S I T P S D L G C K K D E H

057 Clon_KF545954 Y L T L E A I D A G K K K L Y E A K V V V K P W L H F K E L L Q E S E H P S F T T A D L G A K R E E L

058 Smus_SM000037s13475 L I T L E A S D G G G A R Q Y E A K V V T K P W E N H R S L E A T Q P G G A T P A D L G A M P A S Q

059 Eves_Eruve_0061s0028 H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S K P P S I T P S D L G C K K D E H

060 Pegu_XM_020728216 H L T L E A I D A G R K K L Y E A K V V V K E W L S Y K E L L Q E S K H A S F T T A D L G A R R G G H

061 Hvul_HORVU3hr16038190 H L T L E A E A G R K K V Y E A K V V V K P W L D S K E L L Q E S R H T T F T N A D L G A K R G G H

062 Ptri_Potri_0016032900 H L T I E A I E A G K K K I Y E A E V V V K P W L N F M E L L Q E S K H P V F T S S D L G V K R D G H

063 Natt_NIATv7_g19095 H L T L E V I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S T H P T L T S S D L G V K Q E E E

064 Vung_Vigun1g018900 H L T L E A I E G G E K K L Y E A K V V V K P W M N F K E L L Q E S K P S P F T S A D L G V K R E G H

065 Cmic_RWR79807 H L T V E I D A G K K K L Y E A K V V V K P W M D F K E L L Q E S K H C P F T S S D L G A K L G D H

066 Dcat_KZ502679 H L T L E A I D A G K K K L Y E A K V V V K P W L N H K E L L Q E S K H A S F T T A D L G A K R G G H

067 Smoe_85303 H L S I E A H S Q H G T G V Y D A K V V T K P W E S F K K L E D E K P K Q G S F C C S - P E

068 Esysr_Esysr_0039s0067 Q L T L E V E I E A G E K K I Y E A K V V V K P W M N F K Q L Q E S K N P S F T I S D L G L K P D G N

069 Ccaj_XM_020346868 H L T L E A E A G R K K I Y E A K V V V K P W M N F K E L L Q E S K P S A F T S A D L G V K K D G H

070 Cpap_supercontig_228_9 H L T I E A I D A G K K K L Y E A Q V V V K P W M N F K E L L Q E S K H P S F T A S D L G V K K D G H

071 Arul_XYZO_scaffold_2011983 H L T L E A K G D S A P K L Y E A K V V V K P W E N F K K L Q D E K P S S I T P A D L G V K R E G L

072 Klax_Kalax_0745s0017 H L T L E V I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S N H P H F T S S D L G A C Q G E P

073 Abux_sb20040_1 Y L T L E V I D A G K K K I Y E A K I V V K P W I N F K Q L Q E S K H P F S A L S G L N V K R G G H

074 Dstr_Distr_0272s0263 R L T L E V E I E A G E K K I Y E A K V V V K P W M N F K Q L Q E S K N T S F T I S D L G L K P D G N

075 Gsoj_GlysoPI483463_15G194500 H L T L E A E A G R K K L Y E A K V V V K P W L N F K E L L Q E S K P S F T S A D L G V K K G G H

076 Dcar_DCAR_031840 H L T L E V E A G R K K L Y E A K I V V K P W M N F K E L L Q E S K H P S I T P S D L G V K K G G H

077 Zmay_GRMZM2G013461 H L T L E A E A G R K K L Y E A K V V V K P W L D F K E L L Q E S S H T A F T N A D L G A K Q G D H

078 Ashe_KZ451999 H L T L E A I D A G K K K L Y E A K V V V K P W L N H K E L L Q E S K Q V G F T S A D L G C K R G G H

079 Natt_OIS97424 H L T L E V I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S T H P T L T S S D L G V K Q E E E

080 Egui_19846616 Y I T I E A S D G G G K K K L Y E A K V V V K P W M K F K E L L Q E S I P - - C S S E N E T P K D G H

081 Dsoep_Desop_0094s0120 R L T L E V I E A G E K K I Y E A K V V V K P W M N F K E L L Q E S K N P S F S I S D L G L K P D G N

082 Ntab_0000021g0820 H L T L E V I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S S H P T L T S S D L G V K Q E E E

083 Ljap_LjContig00430g0007944_1 H L T L E A I D G G E K K I Y E A K V V V K S W L N F K E L L Q E S K E P L F T T S S D L G V K K D G H

084 Fves_gene21476_v1_0_hybrid H L V V E I D G G K K K L Y E A K V V V K P W L N F K Q V Q E S K H P S F T S S D L G V K Q G G H

085 Alyr_AL6614920 R L T L E V I E A G E K K I Y E A K V V V K P W M N F K Q L Q E S K N P S F T I S D L G L K S D G N

086 Bsy1_Brasy1G380800 H L T L E A I E A G R K K V Y E A K V V V K P W L D F K E L L Q E S R H A S F T T S D L G A K K G G H

087 Ohra_OB01G20920 H L T F E A L E A G R K K V Y E A K V W I K P W L D F K E L L Q E S R H T T F T N A D L G A K K G G H

088 Bnap_CDY08429 H L T L E I V E A G K K K L Y E A K V V V K P W L N F K E L L Q E S K P S S I T P S D L G C K K G E H

089 Ccap_OM095996_MOD H L T V E V M D A G K K K L Y E A K V V V K P W L N L K E L L Q E S K H P S F T A S D L G V K K D G H

090 Bnap_CDY52136 S L T L E V I E A G E K K I Y E A K V V V K P W M N F K Q L Q E S K N P S F T V S D L G L K S D G N

091 Ahal_Araha_21141s0001 H L T L E I L D A G K K K L Y E A K V V V K S W L N F K E L L Q E S K P P A I T P S D L G C K Q D E H

092 Bstr_Bostr_13129s0454 R L T L E V I E A G E K K I Y E A K V V V K P W M N F K Q L Q E S K N P S F T M S D L G L K P D G K

093 Otho_Oroptetium_20150105_09645A H L T L E A I E A G T K K L Y E A K V V V K P W L D F K E L L Q E S S H T A F T N A D L G A K K G G H

094 Bnap_CDY08430 H L T L E I V E A G K K K L Y E A K V V V K P W L N F K E L L Q E S K P S I T P S D L G S K K G E H

095 Mper_Myper_0043s0741 H L T V E I V D A G K K K L Y E A K V V V K P W L N F K E L L Q E S K P P T I T P S D L G C K K D E H

096 Euni_k31_146807_2 H L T I E A V D A G K K K L Y E A K I V V K P W M N F K E L L Q E S K H P S F T P S D L G A R R G G H

097 Thac_TBA0007041 H L T L E A K D - H S P K L F E A K V V V K S S E N F K Q L Q E S K P S P I T P A D L G V K R G L G

098 Pdac_XM_008796910 H L T L E V M E A G K K K L C E A K V V V K P W L N F K E L L Q E S R H S S V T A A D L G A K Q G G H

099 Ppat_Pp3e15_24180V3 H L V I E V E S G K L S Q Y E A K V V V K P W E N F K K E L D K P R P G V F M V L G G N S A P

100 Lesa_Lesat_0046s0160 H L T L E V I E A G E K K L Y E A K V V V K P W L N F K E L L Q E S K P P A I T P S D L G C K K D E H

101 Pinf_101Scf01482g02033 H L T L E V I D A G K K K L Y E A K V V V K P W M N F K E L L Q E S T H P A I T S S D L G V K K E E Q

102 Lang_OIV89378 H L T I E A I D A G E K K T Y H A K V V V K P W L N F K E L L Q E S K H P S F T S A D L G L K K D V H

103 Crub_Carub_0006s0404 H L T L E V I E A G E K K M Y E A K V V V K P W M N F K Q L Q E S K N P S F T I S D L G L K P D G K

104 Mgut_Migut_H00005 H L T L E V V D A G I K K L Y E A K V V V K P W M H F K E L L Q E S K H P P F T S S D L G V K K D E Q

105 Gsoj_GlysoPI483463_13G151700 H L T L E A I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S K P S F T S A D L G V K K D G H

106 Zmar_Zosma94g00120 H L T V E V E A G E K K I L Y E A T V V V K P W L N F K E L L T H N R A V F T A S D L G A V R G E H

107 Pegu_XM_020724410 H L T L E A I D A G S K K L Y E A K V V A K A W L N H K E L L Q E S K H S G F T S A D L G A K R D G H

108 Aktr_CCID_scaffold_2009820 H L T L E A I D A G K K K L Y E A K V V V K P W M N F K E L L Q E S K H P A F T S S D L G V K Q D G H

109 Cmar_Camar_0377s0027 S L T L E V K E A G R K K I Y E A R V V V K P W M N F K Q L Q E S K N P S F T V S D L G L K S D G N

110 Ppat_Pp3e9_24080V3 Y L V I E V E R G S I K K L Y E A K V V V K P W Q N F K K L E E K L R T G G P S T G R G I S A P

111 Omer_OMERI01610740 H L T L E A E A G R K K V Y E A K V V V K P W L D F K E L L Q E S R H T T F T N A D L G A K K G G H

112 Bdis_Bradi2g10146 H L T L E A I E A G R K K V Y E A K V V V K P W L D F K E L L Q E S R H T T F T N A D L G A K K G G H

113 Clan_22909 H L T L E V I D A G K K K L Y E A K V V V K S W M N F K E L L Q E S K H P S F T P S D L G A K K G D H

114 Tarv_Thlar_0016s0705 H L T L E I V E A G K K K L Y E A K V V V K P W L N F K E L L Q E S K P P A I T P S D L G C K K G E H

115 Ntab_0009278g0010_mod H L T L E V I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S T H P T L T S S D L G V K Q E D E

116 Ppat_Pp3e8_17580V3 H F V I E V E R G S Q L K K Y E A K V V V K P G G S N F K E L E E K S T E G V P L T K V S Q H S C V

117 Mtru_Medtr8g041410 H L T L E A I D A G K K K I Y E A K V V V K P W M N F K E L L Q E S K H S F T P S D L G V I K D G H

118 Grai_Grai_0106215400 H L T V E I D G G K K K L Y E A K V V V K P W M N V K E L L Q E S K H P S C T S S D L G L K K D G H

119 Esal_Thhalv10021442m H L T L E I E A G R K K L Y E A K V V V K P W L N F K E L L Q E S K P P A I T P S D L G C K K D E H

120 Paxi_162Scf00516g00518 H L T L E V V D A G K K K L Y E A K V V V K P W L N F K E L L Q E S T H P A I T S S D L G V K K E E Q

121 Dstr_Distr_0241s0826 H L T L E I E A G K K T L Y E A K V V V K P W L N F K E L L Q E S K P P T I T P S D L G C K K D E R

122 Iper_IPERR01610160 H L T L E A I E A G R K K V Y E A K V V V K P W L D F K E L L Q E S R H T T F T N A D L G A K K G G H

123 Cara_Scaffold_637_1450 H L T L E I D A G K K K L Y E A K V V V K P W L N F K E L L Q E S K H P S F T P S D L G L K K D G H

124 Sfal_Sphfalx0005s0315 H L T I E V Q E A G N P K L Y D A K V V V K P W E N F K K E L L Q E S K P A G F T A D L G V K K V S G

125 Macu_GSMIA_Achr3r18480 H L T V E A I D A G K K K L Y E A K V V V K P W L H F K E L L Q D E K H G S L T A D L G A R R G E H

126 Alin_Alyli_0149s0067 H L T L E I E I E A G E K K L Y Q A K V V V K S W L N F K E L L Q E S K P P A L T S S D L G C K K D E H

127 Lusi_Lusi0026117 H L V I E A A D A G E K K K L Y E A K V V V K P W M N F K E L Q E E K H P Q L T S S D L G A K K G C H
128 Mdom_MD02G1275000 H L T I E A T E A G K K K L Y Q A K V V V K P W M N F K E L Q E E K H P Q L T S S D L G A K K G C H
129 Lesa_Lesat_0165s0505 S L T L E V I E A G K K K V Y E A K V V V K P W M N F K E L Q E E K H P Q L T S S D L G A K K G C H
130 Alin_Alyli_0180s0053 R L T L E A L E A G K K K I Y E A K V V V K P W M N F K E L Q E E K H P Q L T S S D L G A K K G C H
131 Taes_CS3002G189800 H L T L E A L E A G K K K V Y E A K V V V K P W L D F K E L Q E E K H P Q L T S S D L G A K K G C H
132 Brap_Brapa_C00203 S L T L E V I E A G K K K L Y E A K V V V K P W M N F K E L Q E E K H P Q L T S S D L G A K K G C H
133 Ccus_JOJQ_scaffold_2007221 H L T I E A T S G G Q R G H Y E A K I W V S K P W E N F K S L Q E E K S P M V T A A D L G S - G A S
134 Mend_ME000129S00342 T L T I E A T K D G N K K L Y E A K V V V K P W E N H K S L Q E E K S P S T A A D L G A N R - A
135 Brap_Brapa_A03337 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P T T I T P S D L G C K K G E G
136 Eves_Eruve_0771s0041 S L T L E V K E G V E K K I Y E A K V V V K P W I N F K Q L Q E E K S P S F T V S D L G L K S D G K
137 Itin_Isati_10464s0003 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P P T I T P S D L G C K K D E H
138 Efin_BFK_scaffold_2025023 H V V V E V K E Q G D G A T K T Y D A R I Y Q R D Q T R T L E S Y K L P G L T P A D L G A Q T A P H
139 Pabi_Ma_4477g0010 H L T I E A E D G D H P K L Y K A K V V V K P W E N F K R R L Q E E K P P S V S R A D L G V K Q H A E
140 Mdom_MD07G1043000 H L T I E A E A G K K K L Y E A K V V V K P W M N F K E L Q E E K Q P S V T S S D L G A K K G G H
141 Esyr_Eusyri_0130s0928 H L T L E I E A G K K K L Y E A K V V V K P W M N F K E L Q E E K P P T I T P S D L G C K K D E H
142 Bhyb_Brahy_S0160313600 H L T L E A E A G K K K V Y E A K V V V K P W L D F K E L Q E E K H T S F T R S D L G A K K G E H
143 Gmax_Glyma_15G227500 H L T L E A E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P P S F T S A D L G V K K D G H
144 Cqui_AUR62031684 H F T I E A I D A G K K K L Y E A K V V V K P W L N F K E L Q E E K H A A F S S A D L G A N R E G H
145 Vvin_VIT_208s0040g00770 Y L T L E A I D A G K K K I Y E A K V V V K P W M N F K E L Q E E K H K S F A P S D L G V K Q D G H
146 Tint_Thint_09G0288600 H L T L E A E A G K K K V Y E A K V V V K P W L D F K E L Q E E K H T S F T I S D L G A K R G G H
147 Stub_PGSC003DMR400068064 H L T L E V M D A G K K K L Y E A K V V V K P W L N F K E L Q E E K H P T F T S S D L G V K Q V E Q
148 Mnar_Mamar_0016s0858 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P P A I T P S D L G C K K D E H
149 Cmar_Camar_0168s0030 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P T T I T P S D L G C K K D E R
150 Stub_PGSC003DMC400045978 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K H P T F T S S D L G V K Q V E Q
151 Sfal_Sphalx0093s0081 H L T I E V Q E G G N P R L Y D A K V V V K P W E N F K E L Q E E K P S T F T S A D L G T R T I P G
152 Lann_Luann_0020s0090 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P P A I T P S D L G C K K D E H
153 Pyla_PGL00010280 H L T I E A E D G D H P K L Y K A K V V V K P W E N F K R R L Q E E K P P S V S R A D L G V K Q H A E
154 Chra_BFEA010003241 H L T L E I E G T E N G H E K L Y K A K V V D Q A A M N S F K E L V S E E K L R T T R A D L G A H R G T R
155 Psom_XM_026571053 H L T L E V I E A G K K K F Y E A K V V V K P W M N F K E L Q E E K H S T C T P S D L G V K R E G H
156 Nnuc_XM_010246848 H L T L E V T E A G M K K L Y E A K V V V K P W L N F K E L Q E E T P P R F T S S D L G A K K D G H
157 Itin_Isati_0492s0004 R L T L E V I E A G K K K I Y E A K V V V K P W M N F K Q L Q E E K N P S F T S S D L G L K S D G N
158 Spur_Sapur_003G148300 H L T I E A E G G K K K M Y E A K V V V K P W L N F K E L Q E E K D P V F T S S D L G A K R D R H
159 Chis_Crahi_1249s0012 H L T L E I E A G K K K L Y E A K V V V K P W I N F K E L Q E E K P P T I T P S D L G C K K D E H
160 Bvul_KMT17034 H F T I E A E A G K K K L Y D A K V V V K P W M N F K E L Q E E K H P S I T P S D L G A I K E G H
161 Casi_28243420 H L T L E A V D A G K K K I Y E A K V V V K P W M N F K Q L Q E E K H P S V T S A G L G V K R D G H
162 Cmar_Camar_0113s0027 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P P S I T P S D L G C K K D E R
163 Osat_IOC_Os01G16430 H L T L E A E A G K K K V Y E A K V V V K P W L D F K E L Q E E R N T F T N A D L G A K K G G H
164 Camp_Caamp_0078s1100 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P P T I T P S D L G C K K D E H
165 Casi_028246879_mod H I T L E A V D A G K K K L Y E A K V V V K P W M N F K E L Q E E K H P A L T P S D L G V K K D G H
166 Sbic_Sobic_003G126800 H L T L E A E A G K K K L Y E A K V V V K P W L N F K E L Q E E S H T T F T N A D L G A K K G E H
167 Pdru_ZZEI_scaffold_2126917 H F K L K A T S G G R V A V Y K A N V V V K P W E N F K S L E L E E I S A F T P A D L G T H L D G S
168 Msin_Misint267700 H L T V E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E S H T T F T N A D L G A K K G G H
169 Psit_PS100017843 H L T I E A E D G D H P K L Y K A K V V V K P W E N F K R L E D E K P P S V S R A D L G V K Q H A E
170 Pdel_Podel_01G035500 H L T I E A E A G K K K I Y E A K V V V K P W L N F K E L Q E E K H P V F T S S D L G V K R D G H
171 Msin_Misin06G106700 H L T L E A E A G K K K L Y E A K V V V K P W L N F K E L Q E E S H T T F T N A D L G A K K G G H
172 Ghir_A06G173900 H L T V E A I D G G K K K V Y E A K V V V K P W M N V K E L Q E E K H P S F T S S D L G I K D G H
173 Pmil_PQIB02000010 H L T L E A I D A G K K K V Y E A K V V V K P W L D F K E L Q E E V H T F T N A D L G A K K G G H
174 Cqep_pep_XM_023692654 H L T V E V A D A G K K K L Y E A K I W V S K S W M N F K E L Q E E K H P S F S S D L G A K K G D H
175 Ahyr_AH015638 H F T I E A I D A G K K K L Y D A K V V V K P W M N F K E L Q E E K H P S F S S D L G A I B E G H
176 Zygn_MF20_scaffold_2022157 H L T I E A T L G G Q K G H Y E A K V V S K P W E N F K S L E E K S P R V T S A D L G S R A - H
177 Lang_OIW08132 Y L T I E A I D A G K K K I Y D A K V V V K P W L N F K E L T E E P S S F T S A D L G V K K D E H
178 Chis_Crahi_1482s0007 H L T L E I V E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P P S I T P S D L G C K K D E H
179 Cvio_Clevi_0011s0507 H L T L E V I E A G K K K I Y E A K V V V Q P W M N F K E L R E E K H P A I T S S D L G C K K D G H
180 Kfed_Kaladp0131s0016 H L T L E V I D A G S K K L Y E A K V V V K P W L N F K E L Q E E N H P F T S S D L G A C Q G E P
181 Slyc_03g097270 H L T L E V M D A G K K K L Y E A K V V V K P W L N F K E L Q E E K H P T F T S S D L G V K Q V E Q
182 Opun_OPUNC01G10650 H L T L E A L E A G K K K V Y E A K V V V K P W L D F K E L Q E E R H T F T N A D L G A K K G E H
183 Ghir_A10G190300 H L T L E A V D A G K K K L Y E A K V V V K P W M N F K E L Q E E K H A S F T A C D L G A K K D V H
184 Mnar_Mamar_0004s0409 H L T L E V I E A G K K K I Y E A K V V V K P W M N F K Q L K E E T N - - V I P D L G L K P D D N
185 Vvin_VIT_213s0084g00050 H L T L E V I D A G K K K L Y E A K V V V K P W M N F K E L Q E E K H P T L P S S D L G V K K D G H
186 Lusi_Lusi0008702 H L V I E A A D A G K K K L Y Q A K V V V K P W M N F K E L Q E E K H P Q L T S S D L G A K K G C H
187 Bsta_Brast01G301000 H L T L E A E A G K K K V Y E A K V V V K P W L D F K E L Q E E R H T S F T S S D L G A K K G E H
188 Taes_CS3A02G185800 H L T L E A L E A G K K K L Y E A K V V V K P W L D F K E L Q E E R H T S F T I S D L G A K R G G H
189 Sspo_03G0034930_1B H L T L E A I E A G K K K L Y E A K V V V K P W L D F K E L Q E E S H T T F T N A D L G A K K G G H
190 Abux_sb22706_1 H L T V E A I D A G K K K L Y E A K V V V K P W L N F K D L Q E E K H A S F T S S D L G V K K D G H
191 Knit_kf100437_10 H L T V E V K T P V G P A E Y K A T V Y E P G P D A T R S L E K Y E L P S D L G A H L G G A S - S A
192 Acoe_Aqcoe1G006500 H L T L E A I E A G K K K I Y E A K V V V K P W M N F K E L Q E E K H P D V T S S D L G V K Q D G H
193 Cmos_XM_023089349 H L T L E V V D A G K K K L Y E A K V V V S W M N F K E L Q D E K H P S I T H S S D L G A K K G D H
194 Alyr_A13G24120 H L T L E I L D A G K K K L Y E A K V V V K S W L N F K E L Q E E K P P A I T P S D L G C K Q D E H
195 Spir_HAOX_scaffold_2024011 Y I T L E V S E A S V R K L Y E V K I W L K P W M D Y K S L E D V K P G A V S P D S A A K V - Q
196 Spen_03g028040 H L T L E V M D A G K K K L Y E A K V V V K P W L N F K E L Q E E N H P T F T S S D L G V K Q V E Q
197 Cole_Ciclev1002346m H L T V E A I D A G T K K L Y E A K V V V K P W L N F K D L Q E E K H A S F T S S D L G V K K D G H
198 Bstr_Bostr_364s0021 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P S A V T S S D L G C K K D E H
199 Hann_HanXRChr04g0119231 H L T L E V D D V G N K K I Y L A K V V V K P W L N F K E L Q E E N H - - P N S D R D I B I E G D
200 Chis_Crahi_0009s0246 H L T L E V I E A G K K K I Y E A K V V V K P W M N F K Q L Q E E K S P S F T V S D L G P K S N G N
201 Itin_Isati_4568s0011 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P S I T P S D L G C K K D E H
202 Atri_AmTr_v1_0_scaffold00101_77 H L T L E A I D A G K K K L Y E A K V V V K P W E N F K E L Q E E N H S S L T S S D L G A K R D N H
203 Doli_LWDK02072575 H L T L E A I E A G N K K V Y E A K V V V K P W L D F K E L Q E E A H T T F T N A D L G A K K G G H
204 Ghir_D10G198100 H L T V E A V D A G K K K L Y E A K V V V K P W M N F K E L Q E E K H A S F T A C D L G A K K D V H
205 Aoco_Anaoc_0016s1440 H L T I E A I E A G K K K L Y E A K V V V K P W M N F K D L Q E E K H P V L P S L L I S V V R W D G H
206 Ljap_Ljlg0013483_1 Y L T L E A I D G G K K I Y E A K I V V K P W M N F I Q L Q E E K L S P F R S S D L G V K Q E G H
207 Slyd_LYD0_6ch03_mod H L T L E V M D A G K K K L Y E A K V V V K P W L N F K E L Q E E K H P T F T S S D L G V K Q V E Q
208 Mper_Myper_0023s0324 R L T L E V I E A G K K K I Y E A K V V V K P W M N F K Q L Q E E K N P S F T S S D L G L I S D G N
209 Brap_Brapa_E02850 H L T L E I V E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P S I T P A D L G C K K G E H
210 Olae_GWH_scaffold_2007720 K I K L W A A D S N T V T Y E A S I W E K P W E N F K E V N Q E E K P Q L V K E G K L I T E S G S E
211 Cmos_XM_023101466 H L T V E A D A G K K K L Y E A K I V V K S W M N F K E L Q E E K H P S F S S D L G A K K G D H
212 Spur_Sapur_0016026900 H L T I E A E A G K K N I Y E A K V V V K P W L N F M E L Q E E K H P V F T S S D L G V K L K D H
213 Cit_Cocit_J0835 H L T I E A V D A G K K K L Y E A K V V V K P W L N S K E L Q E E K H P S F T S S D L G A R T G G H
214 Crub_Carub_0003s1182 H L T L E I E A G K K K L Y E A K V V V K P W L N F K E L Q E E K P S A I T P S D L G A K K D E H
215 Pdel_Podel_03G209600 H L T I E A E A G K K K L Y E A K V V V K P W L N F K E L H E E K D P V F T S S D L G V K R D G H
216 Cang_03g001644 H L T L E V I D A G K K K I Y E A K V V V K P W L H F K E L Q E E K H P T F T S S D L G V K Q E E Q
217 Pmen_PME00047088 H L T I E A B D G D H P K L Y K A K V V V K P W E N F K R L Q E E K P P S V S R A D L G V K Q D G H
218 Aoff_AsparasgV1_07_1767 Y L T V E A I E A G K K K I Y E A K I V V K P W L D F K E L Q E E T H P T I T P A D L G A K R G G H
219 Mnot_L484_024003 H L T L E A E G G K K K I Y E A K V V V K P W L N F K E L Q E E K H P S F T S S D L G V K Q G G H
220 Tarv_Thlar_0026s0267 R L T L E V I E A G K K K I Y E A K V V V K P W M N F K Q L Q E E K N P S F T L S D L G L K S D G N
221 Cmel_XM_008447722 H L T V E V V E A G K K K L Y E A K V V V K S W M N F K E L Q E E K H P S I S S S D L G A K K G D H

222 Rcom_29686_m000868
 223 Csin_Cs616640_1
 224 Even_MKK701000068
 225 Psom_XM_026542585
 226 Oeur_Oeu028996
 227 Spim_03g097270
 228 Sita_Selta_56046700
 229 Mcha_XM_022283450
 230 Gmon_TnS000729105t05
 231 Cpep_XM_023686782
 232 Even_AMZ03000833
 233 Pctr_ZZEI_scaffold_2022423
 234 Egui_XM_010933246
 235 Chre_YOXI_scaffold_2051387
 236 Camp_Caamp_0105s1619
 237 Eves_Eruve_2362s0018
 238 Cmar_Camar_0113s0026
 239 Atau_AET3Gv20420400
 240 Gbil_GBI00017832
 241 Taes_CS3B02G215400
 242 Slyc_Solyc03g097270
 243 Pdac_XM_008788651
 244 Pmen_PME00028125
 245 Pdel_Podel_16G033700

H L T L E A I T E A G K K K I Y E A K V V V V K P W L N F K K E L L Q Q E S K H H P S F T S S D L G V K R R E G H
 Y L T L E V I D A G K K K L Y E A K I V V V K P W L N F K K E L L Q Q E S K H H G S L T A S D L G L K R R G N H
 H L T L E V I D A G K K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H C N S T P S S D L G F K R R E G H
 H L T L E V I E S G K K K L Y E A K V V V V K P W M N F K K E L L Q Q E S K H H P S F T S S D L G V K R R E G H
 H L T L E V M D A G K K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H P T F T S S D L G V K R R E G H
 H L T L E A I E A G K K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H A T F T S S D L G A K R R G G H
 H L T L E A V D A G K K K L Y E A K V V V V K P W M N F K K E L L Q Q E S K H H P A F T P S S D L G A K R R G D H
 H L T L E A T H A S E P K L Y E A K I V V V K A W E N F K K E L L Q Q E S K H H Q P P S L T P A S D L G V K R R L S
 H L T L E V D A G K K K L Y E A K V V V V K P W M N F K K E L L Q Q E S K H H P S I T H S S D L G A K R R G D H
 H L T L E A I D A G K K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H G S L T A A S D L G A K R R G N H
 Y L V L E A A G G G V S K L Y E A K V V V V K P W Q N F K K E L L Q Q E S K H H E T I N P D G I S T - - - S
 Y I T L E A I D G G K K L Y E A K V V V V K P W M K F K K E L L Q Q E S K H H I P - - - C S S E I E A P K D G H
 Y I T L E A S A N G Q D G L Y E A K I V S K P W E K H K K E L L Q Q E S K H H K Q N S A G Q Q G R N G - - D D E
 R L T L E V I E A G K K K I Y E A E V V V K P W M N F K K E L L Q Q E S K H H P S F T S S D L G L K S D G N
 H L T L E I E A G K K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H K P P S I T P S S D L G C K R K D E N
 Y L T L E I E A G K K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H P S I T P S S D L G C K R K D E H
 H L T L E A L E A G R K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H R H T S F T P S S D L G A K R R G G H
 Y L T L E A V D G H E S H R Y S A K I L V V E W R N I K Q L L Q Q E S K H H P S S L T P A S D L G G K Q G S R
 H L T L E A L E A G R K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H T S F T P S S D L G A K R R G R H
 H L T L E V M D A G K K K L Y E A K V V V V K P W L N F K K E L L Q Q E S K H H P T F T P S S D L G V K R R G V Q
 Y I T L E A I E G G K K L Y E A K V V V V K P W M K F K K E L L Q Q E S K H H - - - C S S E I E A P K D G H
 H L T L E A E D G D H P K L Y E A K V V V V K P W E N F K K E L L Q Q E S K H H P S V S R A S D L G V K R Q D V E
 H L T L E A T D A G N K M Y E A K V V V V K P W M N F K K E L L Q Q E S K H H - A T S S D L G V K R E D D H

001 Eves_Eruve_0015s0017
 002 Cmax_XM_023111212
 003 Atha_AT5G05110_1
 004 Vcar_Locus_644_Transcript_10_14
 005 Toac_Theoc_10G138800
 006 Ppin_PP100045831
 007 Egui_XM_010910725
 008 Afil_s0149_g053144
 009 Pval_Phval_0116196600
 010 Cvic_Clevi_0016s0139
 011 Tdic_TRIDC3A6025860_MOD
 012 Drot_Dr09110
 013 Vang_KOM28344
 014 Spol_Spipo3G0066400
 015 Alin_Alyli_0556s0004
 016 Pvir_Pavir_5NG183500
 017 Nnuv_XM_010246133
 018 Grai_Gorai_0116221800
 019 Ccle_Ciclev10012693m
 020 Csat_Cucsa_076380
 021 Ugib_749_g14102_mod
 022 Atha_AT3G12490_2
 023 Vrad_Vradi06g08980
 024 Camp_Caamp_1038s0832
 025 Pvir_Pavir_5KG180700
 026 Nben_101Scf21813g00014_mod
 027 Gmax_Glyma_13G189500
 028 Dsop_Desop_0227s0663
 029 Ptri_Potri_0036192200
 030 Egra_Euogr_J00085
 031 Pper_Prupe_2G047100
 032 Sind_AF240007
 033 Ksub_FQIP_scaffold_2007794
 034 Salb_Sialb_0095s0048
 035 Phal_Pahal_5G416800
 036 Salb_Sialb_0879s0021
 037 Cmax_XM_023130994
 038 Ptri_Potri_0166030900
 039 Nben_101Scf12983g01058_mod
 040 Cvul_MWXT_scaffold_2004286
 041 Cymi_CMI00005557
 042 Bole_Bol015400
 043 Lann_Luann_0004s0723
 044 Cqui_AUR62032567
 045 Aocc_Anaoc_0010s1343_1
 046 Scuc_s0001_g000348
 047 Cgra_Cagra_5641s0012
 048 Svir_Sevir_5G044900
 049 Oglu_OGLUM01G12280
 050 Sacc_SCC59_SCVPEFLAC11C01
 051 Bnap_CX80942
 052 Bole_Bol010294
 053 Ris1_Rois1_0074s0761
 054 Cesc_AF525880
 055 Csub_49027
 056 Camp_Caamp_1039s0467
 057 Clon_KF545954
 058 Smsu_SM000037S13475
 059 Eves_Eruve_0061s0028
 060 Pegu_XM_020728216
 061 Hvil_HORVU3hr1G038190
 062 Ptri_Potri_0016032900
 063 Natt_NATv7_g19095
 064 Vung_Vigun1g018900
 065 Cmic_RWR79807
 066 Dcat_KZ502679
 067 Smoe_85303

G S G W R E V P G D D P E V Q H V A D H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 P Q G W R S V S T P H D P H V Q D V A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G P D W R S V S T P H D P H V Q D V A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G P D W R S V S T P H D P H V Q D V A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G L G W R T V P V H D P V V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R T V P V H D P V V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G V G L Q S V S C N D P S V K E A A K E A I K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 K S G W Q S V P A H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G P E W R V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R D V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R D V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 K S G W Q S V P A H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 V P G W R D V P A H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P D W R S V S T N D P P A V K D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P D W R S V S T N D P P A V K D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G P M W R A V P A H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G P M W R A V P A H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 Q Q E W L A V S T N D L E V K N A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 P Q G W R A V P A H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 L S E W R S V P V K D D P I V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E S G W R S V P G D D P E V K H V A E Q A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 K S G W Q S V P T H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E F D W R L V S T N D L D V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R D V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G S G L K S V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 Q P G W Q S V P T H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G S G W R S V P G D D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 A P G W R A V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 C P G W Q S V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G P G W Q S V P P H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 V S G W R P V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 D A G L R T V P D D P A V E Q D A A Q F A I K Q L Q A R S N S L P P Y E L L Q E V V H A N A E V T G E
 E S G W R S V P G D D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R E V P V D D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 Q S G W R E V P G D D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 P Q G W R E V P L H D P H V Q D A A Q H A L R I I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G S G W Q P V P T N D L E V Q D A A N H A V K S I Q Q R S N S L S P Y E L L Q E V V H A N A E V T G E
 G S G L K S V P V H D P P V Q D A A Q H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 A P G R H S V S T E D P P A V K D A A I E A E R G I R A R S N S L A P P E V V H A N A E V T G E
 S S G W H T V P K Q D P P V Q D A A Q N A V K V I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G F E W R S V S T K D P P V Q D A A K H A V K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G F D W R S V S T N D P P E V V E A A K H A V K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 A P G W K D V P V H D P P V Q N A A N H A L K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 D L E W R E V P T N D L E V Q G A A S Y A V K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G L G L Q Q S V S C N D P S V K E A A K E A I K V I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G S G L R L E V P G D P P V L H V A E H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R E V P I E D P P V K D A A H A V K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R D V P V H D P P V K D A A H A V K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R S V S T K D P P E V Q A A K H A V K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E S G W R E V P G D D P P V Q H V A D H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G S G W R E V P G D D P P V Q H V A E H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E A E W L E I P T H D P P V Q D A A N H A V K S I Q Q R S N T L P P Y E L L Q E V V H A N A E V T G E
 L L G Y K E V S T D D E V L D A A N F A E Q L S Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E S G W R E V P G D D P P V Q H V A D H A V K T I Q Q R S N S L S P Y E L L Q E V V H A N A E V T G E
 G P G L R T V P V N D P L V R D A A D H A L K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 N D M K D V P S D D P A V A D A V E F A V K K L Q E Q S N S L A P Y E L L Q E V V H A N A E V T G E
 V S G W R E V P G D D P P V Q H V A D H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R P I P T H D P P V Q D A A N H A L K N I O A K S N S L A Y E L L Q E V V H A N A E V T G E
 E P G W R D V P V H D P P V Q D A A S H A V K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G P G W Q S V P V H D P P V Q D A A N H A L K S I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 G S G L K S V P V H D P P V Q D A A N H A I K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 K S G W Q S V P T H D P P V Q D A A N H A I K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E A G W R T P I P T H D P P V Q D A A S H A V K T I Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E
 E P G W R P I P T H D P P V Q D A A N H A L K A I O A K S N S L A Y E L L Q E V V H A N A E V T G E
 A A G A A I A L D D P V I N E A A E H A L K G L Q Q R S N S L P P Y E L L Q E V V H A N A E V T G E

104	Mgut_Migut_H00005	C	A	K	F	D	M	L	L	K	V	K	R	G	G	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
105	Gsoj_GlysoFI483463_13G151700	F	A	K	F	N	L	L	L	K	V	K	R	G	G	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
106	Zmar_Zosma94g00120	F	A	K	F	D	M	L	L	K	V	K	R	G	G	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
107	Pequ_XM_020724410	F	A	K	F	D	L	L	L	K	V	K	R	G	G	R	E	D	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H
108	Aktr_CCID_scaffold_2009820	F	A	K	F	D	L	L	L	K	V	K	R	G	G	S	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H
109	Cmar_Camar_0377s0027	R	V	K	F	E	L	L	L	K	L	E	K	G	N	K	P	E	K	F	K	V	E	V	M	K	Y	Q	N	G	K	F	H
110	Ppat_Pp3c9_24080V3	S	A	D	F	E	L	L	L	K	I	K	R	G	A	K	E	H	F	K	S	E	I	H	R	T	G	D	G	D	W	S	
111	Omer_OMERI01G10740	F	A	K	F	D	I	L	M	K	L	K	R	G	N	K	E	K	F	K	A	E	V	H	K	N	L	E	G	A	F	V	
112	Bdis_Bradi2g10146	F	A	K	F	D	I	L	L	K	L	K	R	G	T	K	E	K	F	K	A	E	V	H	K	N	L	E	G	A	F	V	
113	Clan_22909	A	A	K	F	D	L	L	L	K	L	K	R	G	S	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
114	Tarv_Thlar_0016s0705	T	A	K	F	D	M	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
115	Ntab_0009278g0010_mod	N	M	K	L	H	M	L	I	K	T	S	R	G	G	K	Q	E	K	F	K	V	Q	V	H	H	N	S	D	G	G	P	H
116	Ppat_Pp3c8_17580V3	S	A	V	F	D	L	L	L	K	T	K	R	G	S	K	E	N	F	K	S	E	V	H	R	A	E	D	G	S	W	S	
117	Mtru_Medtr8g041410	T	A	K	F	N	L	L	L	K	V	K	R	G	Q	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
118	Grai_Gorai_0106215400	L	T	K	L	E	M	V	L	K	V	K	R	G	D	K	E	K	L	T	V	E	V	Q	H	L	S	H	G	T	F	H	
119	Esal_Thhalv10021442m	A	A	K	F	N	M	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
120	Paxi_162Scf00516g00518	S	T	K	L	H	M	L	I	K	T	S	R	G	G	K	E	K	F	K	V	Q	V	H	H	N	S	D	G	G	P	H	
121	Dstr_Distr_0241s0826	T	A	K	Y	N	M	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
122	Lper_LPERR01G10160	F	A	K	F	D	I	L	M	K	L	K	R	G	T	K	E	K	F	K	A	E	V	H	K	N	L	E	G	A	F	V	
123	Cara_Scaffold_637_1450	S	A	R	F	D	L	H	L	K	V	R	R	G	R	E	K	F	K	V	E	V	N	R	N	S	E	G	Y	F	H		
124	Sfal_Sphfalx0005s0315	I	T	V	F	E	L	V	V	K	T	K	R	G	V	K	E	H	F	K	S	E	Q	R	T	V	D	G	V	W	S		
125	Macu_GSMIA_Achr3T18480	I	A	K	F	D	M	L	L	K	V	K	R	G	S	K	E	K	L	K	V	E	V	H	K	N	L	E	G	T	F	H	
126	Alin_Alyli_0149s0067	A	A	K	F	N	M	L	L	K	L	K	R	G	E	K	E	R	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
127	Lusi_Lus10026117	S	A	K	F	H	L	V	L	K	V	K	R	G	S	T	E	K	L	K	V	E	V	H	K	N	S	D	G	G	P	H	
128	Mdom_MD02G1275000	H	A	K	F	N	M	L	L	K	V	K	R	G	S	K	E	K	F	K	A	E	V	H	K	N	S	D	G	G	P	H	
129	Lesal_Lesat_0165s0505	H	V	K	F	E	L	L	L	K	L	E	R	G	N	K	L	E	K	F	V	E	V	M	K	Y	Q	N	G	K	F	H	
130	Alin_Alyli_0180s0053	R	V	K	F	E	L	L	L	K	L	E	R	G	N	K	L	E	K	L	M	V	E	V	M	K	Y	Q	N	G	K	F	H
131	Taes-Cs3M02G189800	F	A	K	F	D	I	L	M	K	L	K	R	G	T	K	E	K	F	K	A	E	V	H	K	N	L	E	G	A	F	V	
132	Brap_Brapa_C00203	R	V	K	F	E	L	L	L	K	L	E	K	G	N	K	P	E	K	L	M	V	E	V	M	K	Y	Q	N	G	K	F	H
133	Ccus_JoJo_scaffold_2007221	T	S	K	L	H	L	L	K	V	K	R	G	T	R	E	T	E	T	M	K	T	E	L	H	A	D	E	E	W	S		
134	Mend_ME000129s00342	G	T	L	Y	T	L	A	L	K	L	G	R	G	K	Q	E	H	I	E	A	E	V	H	H	S	A	S	G	E	W	K	
135	Brap_Brapa_A03337	A	A	K	Y	N	M	V	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
136	Eves_Eruve_0771s0041	R	V	K	F	E	L	L	L	K	L	E	K	G	N	K	A	E	K	L	M	V	E	V	M	K	Y	Q	N	G	K	F	H
137	Itin_Isati_10464s0003	A	A	K	F	N	L	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
138	Efim_BFIK_scaffold_2025023	G	V	T	F	E	L	T	E	V	A	S	G	A	H	T	E	L	T	A	V	E	H	R	T	K	E	G	K	F	T		
139	Pabi_MA_4477g0010	S	S	K	V	D	H	L	K	T	K	R	G	S	K	V	E	H	K	V	E	M	H	C	N	P	D	G	K	W	I		
140	Mdom_MD07G1043000	H	A	K	F	N	M	L	L	K	V	K	R	G	S	K	E	K	F	K	A	E	V	H	K	N	S	D	G	G	P	H	
141	Esyr_Esyr_0130s0928	A	A	K	Y	N	M	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
142	Bhyb_Brah_S01G0313600	F	A	K	F	D	I	L	L	K	L	K	R	G	T	K	E	K	I	K	A	E	V	H	K	N	L	E	G	A	F	V	
143	Gmax_Glyma_15G227500	F	A	K	F	N	L	L	L	K	V	K	R	G	Q	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
144	Cqui_AUR62031684	T	A	K	F	N	L	L	L	K	V	K	R	G	E	K	Q	S	N	V	E	V	H	K	N	S	E	G	N	Y	N		
145	Vvin_VIT_208s0040g00770	S	A	K	F	D	L	L	L	K	L	K	R	G	S	K	D	V	K	F	K	A	E	V	H	K	N	L	E	G	K	F	C
146	Tint_Thint_09G0288600	F	A	K	F	D	I	L	M	K	L	K	R	G	T	K	E	K	M	K	A	E	V	H	K	N	L	E	G	A	F	V	
147	Stub_PGSC0003DMT400068064	H	T	K	L	H	L	V	I	K	T	S	R	G	G	K	E	K	F	K	V	Q	V	H	H	N	S	D	G	G	P	H	
148	Mnar_Mamar_0016s0858	A	A	K	F	N	M	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
149	Cmar_Camar_0168s0030	A	A	K	Y	N	M	L	L	K	L	K	R	G	E	K	E	K	F	R	V	E	V	H	K	N	S	D	G	G	P	H	
150	Stub_PGSC0003DMC400045978	H	T	K	L	H	L	V	I	K	T	S	R	G	G	K	E	K	F	K	V	Q	V	H	H	N	S	D	G	G	P	H	
151	Sfal_Sphfalx0093s0081	D	V	V	F	D	L	L	L	K	T	R	G	A	T	E	H	F	K	A	E	V	Q	R	T	V	D	G	A	W	S		
152	Lann_Luann_0020s0090	A	A	K	F	N	L	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
153	Pgla_PGL00010280	S	S	K	V	D	H	L	K	T	K	R	G	S	K	V	E	H	K	V	E	M	H	C	N	P	D	G	K	W	I		
154	Chra_BFEA010003241	G	T	T	I	H	L	V	L	N	V	G	I	S	S	S	D	H	I	V	K	A	S	M	H	R	T	D	D	G	H	W	R
155	Psom_XM_026571053	S	A	K	F	H	L	L	K	L	K	R	G	S	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H		
156	Nnuc_XM_010246848	V	A	K	L	N	L	L	L	K	V	Q	R	G	S	R	E	K	F	K	V	E	L	Q	K	T	I	E	G	T	F	H	
157	Itin_Isati_0492s0004	C	V	K	F	E	L	L	L	K	L	E	R	G	N	K	L	E	K	M	V	D	M	Y	P	N	G	K	F	H			
158	Spur_Sapur_003G148300	S	A	R	F	D	M	L	L	R	V	K	R	G	S	T	E	K	F	K	V	M	V	H	K	N	S	D	G	G	P	H	
159	Chis_Crahi_1249s0012	A	A	K	F	N	M	L	L	K	L	K	R	G	E	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
160	Bvul_KMT17034	T	A	K	F	N	L	L	L	K	V	K	R	G	N	K	E	I	F	N	V	E	V	H	K	N	S	D	G	G	P	H	
161	Casi_28243420	Y	T	K	F	A	M	L	L	K	L	R	R	G	T	K	E	K	F	N	V	L	N	K	N	R	E	G	K	F	Y		
162	Cmar_Camar_0113s0027	A	A	K	F	N	M	L	L	K	L	K	R	G	N	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
163	Osat_LOC_Os01g16430	F	A	K	F	D	I	L	M	K	L	K	R	G	N	K	E	K	F	K	A	E	V	H	K	N	L	E	G	A	F	V	
164	Camp_Caamp_0078s1100	A	A	K	F	S	M	L	L	K	L	K	R	G	D	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
165	Casi_028246879_mod	S	A	K	I	D	M	V	L	K	V	K	R	G	D	K	E	K	F	K	V	E	V	H	K	N	S	D	G	G	P	H	
166	Sbic_Sobic_003G126800	F	A	K	F	D	I	L	M	K	L	K	R	G	S	K	E	K	I	K	A	E	V	H	K	N	L	E	G	A	F	V	
167	Pdru_ZZEI_scaffold_2126917	S	T	R	F	N	L	L	F	P	T	Q	R	G	S	K	E	Q	F	K	A	E	I	S	R	T	V	V	G	A	W	T	
168	Msin_Misint267700	F	A	K	F	D	I	L	M	K	L	K	R	G	S	K	E	K	I	K	A	E	V	H	K	N	L	E	G	A	F	V	
169	Psit_PSI00017843	S	S	K	V	D	H	L	K	T	K	R	G	S	K	V	E	H	K	V	E	M	H	C	N	P	D	G	K	W	I		
170																																	

009 Fvul_Fhvil_011g196600
 010 Cvio_Clevi_0016s0139
 011 Tdic_TRIDC3Ag025860_MOD
 012 Drot_Dr09110
 013 Vang_KOM28344
 014 Spo1_Spipo3G0066400
 015 Alln_Alyli_0556s0004
 016 Pvir_Pavir_5Ng183500
 017 Nnuc_XM_010246133
 018 Grai_Gorai_011G221800
 019 Cole_Ciclev10012693m
 020 Csat_Cucsa_076380
 021 Ugib_749_g14102_mod
 022 Atha_AT3G12490_2
 023 Vrad_Vradi06g08980
 024 Camp_Caamp_1038s0832
 025 Pvir_Pavir_5KG180700
 026 Nben_101Scf21813g00014_mod
 027 Gmax_Glyma_13G189500
 028 Dsop_Desop_0227s0663
 029 Ptri_Potri_003G192200
 030 Egra_Euogr_J00085
 031 Pper_Prupe_2G047100
 032 Sind_AF240007
 033 Ksub_FQLP_scaffold_2007794
 034 Salb_Sialb_0095s0048
 035 Phal_Pahal_5G416800
 036 Salb_Sialb_0879s0021
 037 Cmax_XM_023130994
 038 Ptri_Potri_016G030900
 039 Nben_101Scf12983g01058_mod
 040 Cvul_MWXT_scaffold_2004286
 041 Cymi_CMI00005557
 042 Bole_Bol1015400
 043 Lann_Luann_0004s0723
 044 Cqui_AUR62032567
 045 Aoccc_Anaoc_0010s1343_1
 046 Scuc_s0001_g000348
 047 Cgra_Cagra_5641s0012
 048 Svir_Sevir_5G044900
 049 Oglu_OGLUM01G12280
 050 Sacc_SCC59_SCVFPLA11C01
 051 Bnap_CBX80942
 052 Bole_Bol1010294
 053 Ris1_Rois1_0074s0761
 054 Cesc_AF525880
 055 Csub_49027
 056 Camp_Caamp_1039s0467
 057 Clon_KP545954
 058 Sms_SM000037S13475
 059 Eves_Eruve_0061s0028
 060 Pequ_XM_020728216
 061 Hvil_HORVU3hr1G038190
 062 Ptri_Potri_001G032900
 063 Natt_NATv7_g19095
 064 Vung_Vigun1g018900
 065 Cmic_RWR79807
 066 Dcat_K2502679
 067 Smoe_85303
 068 Esyr_Esyr_0039s0067
 069 Ccaj_XM_020346868
 070 Cpap_supercontig_228_9
 071 Arul_XTZO_scaffold1_2011983
 072 Klax_Kalax_0745s0017
 073 Abux_sb20040_1
 074 Dstr_Distr_0272s0263
 075 Gsoj_GlysoPI483463_15G194500
 076 Dcar_DCAR_031840
 077 Zmay_GRM2G013461
 078 Ashe_K2451999
 079 Natt_OIS97424
 080 Egui_19846616
 081 Dsop_Desop_0094s0120
 082 Ntab_0000021g0820
 083 Ljap_LjContig00430g0007944_1
 084 Fves_gene21476_v1_0_hybrid
 085 Alyr_AL6614920
 086 Bsyl_Brasy1G380800
 087 Obra_OB01G20920
 088 Bnap_CDY08429
 089 Coap_OM095996_MOD
 090 Bnap_CDY52136
 091 Ahal_Araha_21141s0001
 092 Bstr_Bostr_13129s0454
 093 Otho_OropeTium_20150105_09645A
 094 Bnap_CDY08430
 095 Mper_Myper_0043s0741
 096 Euni_k31_146807_2
 097 Thac_TBAD0007041
 098 Rdac_XM_008796910
 099 Ppat_Pp3cl5_24180V3
 100 Lesa_Lesat_0046s0160
 101 Pinf_101Scf01482g02033
 102 Lang_OIV89378
 103 Crub_Carub_0006s0404

FAKFNLLKVKRGEKEKFKVEVHKNNQCEFLH
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNPFL
 FAKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 SAKFDMLLKLKRGSKLEKFKVEVHKNNLEGTFH
 FAKFNLLKVKRGGKLEKFKVEVHKNNLEGALH
 SAKFNLLKIRRGTKLEKFKVEVHKNNLEGAPV
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKYH
 FTKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 VAKFDMLLKLKRGSKLEKFKVEVHKNNLEGTFH
 FAKLNMVLLKVKRGGKLEKFKVEVHKNNLEGTYH
 YAKFELHLKLRGSEEEKHFWVEIKNKSEGKPFY
 AAKFDLLKLRGSKLEKFKVEVHKNNLEGNPFL
 STKFDILMLKRGSKLEKFKVEVHKNNLEGNPFL
 AAKYMLLKLKRGSKLEKFKVEVHKNNLEGALH
 FAKFNLLKVKRGEKEKFKVEVHKNNLEGALH
 CV---LLKLRGSKLEKFKMVDVMKYPNKPFH
 FAKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 NMKLLHMLIKTSRGGKQKEKFKVQVHHNNEGAPQ
 FAKFNLLKVKRGGKLEKFKVEVHKNNQGGPFH
 AAKFNMLLKLKRGSKLEKFKVEVHKNNLEGDLH
 SAKFDMLLKLKRGSKTEKFKVIVHKNNLEGNKYH
 LAKFDMLLKLKRGSKLEKFKVEVHKNNLEGILK
 HAKFNMLLKLKRGSKLEKFKVEVHKNNLEGTFK
 SAKFDMLLKLKRGSKLEKFKVEVHKNNLEGGFN
 TTFHLLLELQHGNRTETVRFVVDRTKDGKPFH
 AAKFNMLLKLKRGSKLEKFKVEVHKNNLEGVLH
 FAKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 AAKFNMLLKLKRGSKLEKFKMAEVHKNNLEGVLH
 AAKFDLLKLRGSKLEKFKVEVHKNNLEGNPFL
 YAKFNLLKLRGSKLEKFKVEVHKNNLEGKPFH
 NMKLLHMLIKTSRGGKREKFKVQVHHNNEGAPQ
 GTTFHLLVLRVAISTSDHMMVXTSMHRTNNDGHWRR
 LEKFNLLLEIKWGTQVQVQVMSQRSLDGKVVV
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKPFH
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKPFH
 TAKFNLLKVKRGGKLEKFKVEVHKNNLEGNYN
 YARFELHLKLRGSKLEKFKQVEVHKNNLEGKPFH
 KTNFSLFLKRGSKLEKFKADIVRTVNGMWT
 AAKFNMLLKLKRGSKLEKFKVEVHKNNLEGDLH
 FAKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 FAKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 FAKFDILMMLKRGSKLEKFKMVEVHKNNLEGAPV
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKPFH
 ASKFNMLLKLKRGSKLEKFKVEVHKNNLEGATW
 TAKFNMLLKLKRGSKLEKFKVEVHKNNLEGALH
 LAKIHLMLKLRGSKREKFKVEVHKNNLEGTFH
 GRVFDLAKLSQDLSQDIMKVEVSRDLKNTTYL
 AAKFNMLLKLKRGSKLEKFKVEVHKNNLEGVLH
 TAKFDMLLKLKRGSKLEKFKMAEVHKNNLEGNPFL
 GTTFHTTLKLRGTAAEVMMNVEHRTPEGSFT
 AAKFNMLLKLKRGSKLEKFKVEVHKNNLEGVLH
 HAKLELLKLRGENEKKYKVEVHKNNLEGAFI
 FAKFDIVMMLKRGTKLEKFKMAEVHKNNLEGAPV
 SAKFDMLLKLKRGSAEKKLVVVKNNLEGNKYH
 NMKLLHMLIKTSRGGKLEKFKVQVHHNNEGAPQ
 FAKFNLLKVKRGGKLEKFKVEVHKNNLEGALH
 FAKFEMLLKLRGSKLEKFKVEVHKNNLEGNPFL
 FAKLELLKLRGSKLEKFKVEVHKNNLEGAFI
 TNALDLVLLKVRGSKREHVNAKLHRGDTGWTLL
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKPFH
 FAKFNLLKVKRGGKLEKFKVEVHKNNQGGPFH
 SPKFDMLLKLKRGSKLEKFKVEVHKNNQGNPFL
 FAKFDLLKTKRGGKLEKFKVEVHKNNLEGDWMM
 LAKFDILNMLKRGSKLEKFKVIVHKNNLEGSFH
 YAKFELHLKLRGSEEEKHFWVEIKNKSEGKPFY
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKPFH
 FAKFNLLKVKRGGKLEKFKVEVHKNNQGGPFH
 SAKFDILKVTRGDKLEKFKVEVHKNNLEGAPK
 FAKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 FAKFDLLKLRGSKLEKFKRVEVHKNNLEGTFPL
 NMKLLHMLIKTSRGGKLEKFKVQVHHNNEGAPQ
 SAKFDLLKVKRGGKLEKFKVEVDETMQGTTFH
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKYH
 NMKLLHMLIKTSRGGKQKEKFKVQVHHNNEGAPQ
 VAKFNILLKLRGSKLEKFKVEVHKNNLEGSFH
 HAKFNMLLKLKRGSKLEKFKVEVHKNNLEGAYN
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNK--
 FAKFDILKLRGSKLEKFKMAEVHKNNLEGAPV
 FAKFDILMMLKRGTKLEKFKMAEVHKNNLEGAPV
 AAKYMLLKLKRGSKLEKFKVEVHKNNLEGVLH
 FAKLEMLLKLKRGSKLEKFKVEVHKNNLEGTYH
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKPFH
 AAKYMLLKLKRGSKLEKFKVEVHKNNLEGALH
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKYQ
 FAKFDILMMLKRGSKLEKFKMAEVHKNNLEGAPV
 AAKYMLLKLKRGSKLEKFKVEVHKNNLEGVLH
 AAKFNLLKLRGSKLEKFKVEVHKNNLEGVLH
 LAKFDMLLKLKRGSKLEKFKVEVHKNNLEGILQ
 FAKFDLLKTKRGGKLEKFKVEVHKNNLEGDWLL
 SAKFDLLKLRGSKLEKFKVEVHKNNLEGTFH
 TAQFDLLKTKLGAKEQVFKAEVSRGTGDGWT
 AAKFNMLLKLKRGSKLEKFKVEVHKNNLEGALH
 STKLLHMLIKTSRGGKLEKFKVQVHHNNEGAPH
 FAKFNLLKVKRGGKLEKFKVEVHKNNLEGGPFH
 RVKPELLLKLKRGSKLEKFKMVEVHKNNQGNKYQ

199 Hann_HanRQCChr04g0119231 L V K Y D I L L K L K R N D K E E K F K A N V H R D N E G A P H
 200 Chis_Crahi_0009s0246 R V K F E L L L K L E K G M K P E K L M V E V M K Y Q N G K F H
 201 Itin_Isati_4568s0011 A A K F N L L L K L K R G E K E K F K V E V H K N H E G V L H
 202 Atri_Amrr_v1_0_scaffold00101_77 S A K F D M L L K L K R G M K E E K F K V E V H K N L E G S Y H
 203 Doli_LWDK02072575 F A K F D I L M K L K R G S K E E K L K A E V H K S L E G A F V
 204 Ghir_D10G198100 F A K L N M V L K V K R G D K E E K F K V E V H H K N E G T Y H
 205 Aocce_Anaoc_0016s1440 F A K F N M L L K V K R G D K E E K F K V E V H K N H E G T P H
 206 Ljap_Lj1g0013483_1 Y V K F N L L L K V S R G I K E E I F R V E V N K K L G G R P Y
 207 Slyd_LYD0_6ch03_mod S T K L H L V I K T S R G G K E E K F K V Q V Q H N N E G A P H
 208 Mper_Myper_0023s0324 R V K F E L L L K L E R G M K L E K F M V D V I K Y P N G K F H
 209 Brap_Brara_E02850 A A K F N M L L K L K R G D K E E K F K V E V H K N H E G V L H
 210 Olae_GGWH_scaffold_2007720 S A L F P L A L E L T R P S N A V E Q E V E V Q V S D G K W S
 211 Cmos_XM_023101466 A A K F D L L L K L K R G S K E E K F K V E V H K N N E G N P L
 212 Spur_Sapur_0016026900 S A K F D L L L K V K R G S A E E K L R V V V H K N S E G T Y H
 213 Ccit_Cocit_J0835 S A K F D M L L K V K R G D K E E K F K V E V H K N H E G I L K
 214 Crub_Carub_0003s1182 A A K F N M L L K L K R G D K E E K F K V E V H K N H E G D L H
 215 Pdel_Podel_03G209600 S A K F D M L L K V K R G S T E E K F K V L V H K N N E G N Y H
 216 Cang_03g001644 S T K I H L V I K T S R G G K E E K F K V Q V H H N N E G A P H
 217 Fmen_FME00047088 L A K V D L H L K T K R G S K V E E H K V E M H R N L D G K W I
 218 Aoff_AsparagusV1_07_1767 L V K F D L L L K L R R G S K E E K Y K V E V H K N L E G N P H
 219 Mnot_L484_024003 S A K F D M L L K V K R G T K E E K Y K A E V H K N S E G T P H
 220 Tarv_Thlar_0026s0267 R V K F E L L L K L E R G N K P E K F M V E V M K Y E N G K F H
 221 Cmel_XM_008447722 A A K F D L L L K L K R G S K E E K F K V E V H K N N E G N P L
 222 Rcom_29686_m000868 F A K F D M L L K V K R G T S E E K F K V E V H K N N E G T P L
 223 Csin_Cs6g16640_1 Y A K F E H L K L R R G S E E E K H W V E I I K N S E G K F Y
 224 Even_MKKT01000068 E Q G I D P T G R Y T G T T D L Q L E R V N V Y N E A L V G G
 225 Psom_XM_026542585 A V K F H I L L T V K W G S Q E E K Y K V E V H K D V E G R L L
 226 Oeur_Oeu028986 S A K F D M L L K V N R G G K E E K F K V E V H K N N E G G F H
 227 Spim_03g097270 S T K L H L V I K T S R G G K E E K F K V Q V Q H N N E G A P H
 228 Sita_Seita_5G046700 F A K F D I L M K L K R G S K E E K L K A E V H K S L E G V F V
 229 Mcha_XM_022283450 S A K F D L H L K V M R G S K E E K F K V K V H K N N E G N P L
 230 Gmon_TnS000729105t05 A A K L D L H L K V K R G T K E E E H K V E V H R N L D G K W L
 231 Cpep_XM_023686782 A A K F D L L L K L K R G S K E E K F K V E V H K N N E G N P L
 232 Even_AMZH03000833 E Q G I D P T G R Y T G T T D L Q L E R V N V Y N E G S Y E C
 233 Pdru_ZZEI_scaffold_2022423 S T I Y D L V L I T Q R G S K E E Q I D A E V V R M G V G A W T
 234 Egui_XM_010933246 S A R F E L L L K V A R G Q R E E K M V E V D K T M Q G I F H
 235 Chre_YOXI_scaffold_2051387 G S K F H L V M K L K R G A M V E T R K A L V H F D G G D K W T
 236 Camp_Caamp_0105s1619 R V K F E L L L K L E R G N K L E K F M V E V M K Y Q N G K F H
 237 Eves_Eruve_2362s0018 A A K F N M L L K L K R G D K E E K F K V E V H K N H E G V L H
 238 Cmar_Camar_0113s0026 A A K F N M L L K L K R G D K E E K F K V E V H K N S E G V L H
 239 Atau_AET3gv20420400 F A K F D I L M K L K R G T K E E K M K A E V H K N L E G A F V
 240 Gbil_GBI00017832 S E T F D L D L K I K W G T K E E Q C K V E I L Q S P H G R W T
 241 Taes_CS3B02G215400 F A K F D I L M K L K R G T K E E K M K A E V H K N L E G A F V
 242 Slyc_Solyc03g097270 S T K L H L V I K T S R G G K E E K F K V Q V Q H N N E G A P H
 243 Pdac_XM_008788651 S A K F E L L L K V R R G Q E E E K F M V E V D E T M Q G T P H
 244 Fmen_FME00028125 L A K V D L H L K T K R G S K V E E H K V E M H R N L D G K W I
 245 Pdel_Podel_16G033700 Y A K F N L L L K L R R G I K E E N F K V E V I K N N E G K F H

1 2 3 4 5 6 7 8 9
 Variable Average Conserved

X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

CURRICULUM VITÆ

BALBINOTT, N.

1. DADOS PESSOAIS

NOME: Natalia Balbinott

DATA DE NASCIMENTO: 25/03/1996

LOCAL DE NASCIMENTO: Passo Fundo – Rio Grande do Sul, Brasil

TELEFONE PROFISSIONAL: (54) 999 594 594

E-MAIL: nati.balbinott@gmail.com

2. FORMAÇÃO

- 2019 – Atual Mestrado em andamento
Programa de Pós-graduação em Biologia Celular e Molecular (PPGBCM) da Universidade Federal do Rio Grande do Sul (UFRGS)
Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
- 2013 – 2018 Graduação em Ciências Biológicas (Bacharelado)
Universidade de Passo Fundo (UPF), Passo Fundo, Brasil
com período sanduíche em University of Melbourne, Austrália
Bolsista do Programa Ciência Sem Fronteiras - Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

3. ESTÁGIOS

- 2018 Estágio curricular obrigatório
Laboratório de Genomas e Populações de Plantas
Departamento de Biofísica, Universidade Federal do Rio Grande do Sul
- 2017 – 2018 Bolsista de Iniciação Científica (PIBIC/CNPq)
Laboratório de Biotecnologia
Projeto: Potencial da construção gênica RNAi-*chs3b* para conferir resistência à giberela do trigo via silenciamento gênico induzido pelo hospedeiro em *Arabidopsis thaliana*
Empresa Brasileira de Pesquisa Agropecuária, Embrapa Trigo
- 2016 – 2017 Bolsista de Iniciação Científica (PROBIC/FAPERGS)
Laboratório de Microbiologia e Imunologia Avançada
Projetos: Desenvolvimento e caracterização de IL-1 e IL-6 recombinantes de suínos
Faculdade de Medicina Veterinária, Universidade de Passo Fundo

2013 – 2014 Bolsista de Iniciação Científica (PIBIC/UPF)
Laboratório de Herpetologia
Projeto: Taxocenose de anfíbios em ambiente urbano
Instituto de Ciências Biológicas, Universidade de Passo Fundo

4. PRÊMIOS E DISTINÇÕES

- 2018 2º lugar na XIII Mostra de Iniciação Científica da Embrapa Trigo
- 2011 Menção Honrosa do Concurso "Química Verde e seus benefícios na sociedade atual", UNESCO.

5. EXPERIÊNCIA PROFISSIONAL OU DIDÁTICA ANTERIOR

- 2018 Monitora da disciplina de Ecologia Urbana
Curso de Ciências Biológicas (Bacharelado), Universidade de Passo Fundo
- 2014 Monitora da disciplina de Zoologia dos Invertebrados II
Curso de Ciências Biológicas (Licenciatura), Universidade de Passo Fundo
- 2014 Projeto de Extensão: Biodiversidade na Escola
Universidade de Passo Fundo

6. ARTIGOS COMPLETOS PUBLICADOS

RODRIGUES, N. F.; BALBINOTT, N.; PAIM, I.; GUZMAN, F.; MARGIS, R. Comparative analysis of the complete chloroplast genomes from six Neotropical species of Myrteae (Myrtaceae). *Genetics and Molecular Biology*, v. 43, p. e20190302, 2020.

PIRES ESPÍNDOLA, J.; BALBINOTT, N.; TREVISAN GRESSLER, L.; MACHADO, G.; SILENE KLEIN, C.; REBELATTO, R.; GUTIÉRREZ MARTÍN, C. B.; KREUTZ, L. C.; SCHRYVERS, A. B.; FRANDOLOSO, R. Molecular serotyping of clinical strains of *Haemophilus (Glaesserella) parasuis* brings new insights regarding Glässer's disease outbreaks in Brazil. *PeerJ*, v. 7, p. e6817, 2019.

GUIZZO, J. A.; CHAUDHURI, S.; PRIGOL, S. R.; YU, R.; DAZZI, C. C.; BALBINOTT, N.; FRANDOLOSO, G. P.; KREUTZ, L. C.; FRANDOLOSO, R.; SCHRYVERS, A. B. The amino acid selected for generating mutant TbpB antigens defective in binding transferrin can compromise the in vivo protective capacity. *Scientific Reports*, v. 8, p. 7372, 2018.

7. RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS

BALBINOTT, N.; MARGIS, R. Group II Phytocystatins: the bifunctional plant cystatins. *VII Simpósio Brasileiro de Genética Molecular de Plantas*, 2019, Campos do Jordão.

BALBINOTT, N.; PASA, J. B.; NUNES, G. C. M.; BOCK, L.; GRIGOLO, R.; ZANELLA, N. Distribuição espacial e temporal de anuros em poças permanentes da UPF, RS. In: *7º Congresso Brasileiro de Herpetologia*, 2015, Gramado.

ESPINDOLA, J. P.; BALBINOTT, N.; GRESSLER, L.; MACHADO, G.; KLEIN, C. S.; REBELATTO, R.; GUTIÉRREZ-MARTÍN, C. B.; KREUTZ, L. C.; FRANDOLOSO, R. . Evidence that most of the isolated *Haemophilus parasuis* serovars are absent in commercial vaccine composition in Brazil. In: *29º Congresso Brasileiro de Microbiologia*, 2017, Foz do Iguaçu. *29º Congresso Brasileiro de Microbiologia*, 2017.

FONTANA, R. B.; BALBINOTT, N.; SEIDLER, S. S.; ZANELLA, N. . Biologia reprodutiva de *Dendropsophus minutus* (Peters, 1872) no norte do estado do Rio Grande do Sul, Brasil. In: *7º Congresso Brasileiro de Herpetologia*, 2015, Gramado. *7º Congresso Brasileiro de Herpetologia*, 2015.

PASA, J. B.; NUNES, G. C. M.; SOARES, J. P.; BOCK, L.; BALBINOTT, N.; GRIGOLO, R.; ZANELLA, N. . Anurofauna do Parque Estadual Papagaio Charão, RS, BR. In: *7º Congresso Brasileiro de Herpetologia*, 2015, Gramado. *7º Congresso Brasileiro de Herpetologia*, 2015.