



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

**JOÃO ABNER SOUZA LIMA**

**GENES CODIFICANDO QUITINASES DE *Arabidopsis thaliana* E ESPÉCIES  
RELACIONADAS: UM ESTUDO *IN SILICO* DA VARIABILIDADE DAS  
ESTRUTURAS PRIMÁRIAS DAS ENZIMAS, SUAS RELAÇÕES FILOGENÉTICAS  
E DIVERSIDADE DE ORGANIZAÇÃO DE DOMÍNIOS**

**FORTALEZA**

**2026**

JOÃO ABNER SOUZA LIMA

GENES CODIFICANDO QUITINASES DE *Arabidopsis thaliana* E ESPÉCIES  
RELACIONADAS: UM ESTUDO *IN SILICO* DA VARIABILIDADE DAS ESTRUTURAS  
PRIMÁRIAS DAS ENZIMAS, SUAS RELAÇÕES FILOGENÉTICAS E DIVERSIDADE  
DE ORGANIZAÇÃO DE DOMÍNIOS

Trabalho de conclusão de curso apresentado ao  
Curso de Graduação em Biotecnologia do  
Centro de Ciências da Universidade Federal do  
Ceará como requisito parcial à obtenção do  
grau de Bacharel em Biotecnologia.

Orientador: Prof. Dr. Thalles Barbosa  
Grangeiro.

FORTALEZA

2026

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas

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

---

- L698g Lima, João Abner.  
Genes Codificando Quitinases de *Arabidopsis thaliana* e Espécies Relacionadas: um estudo *in silico* da variabilidade das estruturas primárias das enzimas, suas relações filogenéticas e diversidade de organização de domínios / João Abner Lima. – 2026.  
75 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Biotecnologia, Fortaleza, 2026.  
Orientação: Prof. Dr. Thalles Barbosa Grangeiro.
1. Filogenia. 2. Domínios Conservados . 3. PR proteins. I. Título.

CDD 661

---

JOÃO ABNER SOUZA LIMA

GENES CODIFICANDO QUITINASES DE *Arabidopsis thaliana* E ESPÉCIES  
RELACIONADAS: UM ESTUDO *IN SILICO* DA VARIABILIDADE DAS ESTRUTURAS  
PRIMÁRIAS DAS ENZIMAS, SUAS RELAÇÕES FILOGENÉTICAS E DIVERSIDADE  
DE ORGANIZAÇÃO DE DOMÍNIOS

Trabalho de conclusão de curso apresentado ao  
Curso de Graduação em Biotecnologia do  
Centro de Ciências da Universidade Federal do  
Ceará como requisito parcial à obtenção do  
grau de Bacharel em Biotecnologia.

Aprovado em 05/01/2026

BANCA EXAMINADORA

---

Prof. Dr. Thalles Barbosa Grangeiro (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dra. Cristina Paiva da Silveira Carvalho  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. José Hélio Costa  
Universidade Federal do Ceará (UFC)

## AGRADECIMENTOS

Agradeço a Deus e sua esposa, Igreja Católica Apostólica Romana, e todos os instrumentos pastorais e suas lideranças, em destaque a pessoa do administrador paroquial Pe. Ademar Cason, SDS, que acompanharam com fervor todas as etapas do meu desenvolvimento pessoal, espiritual e acadêmico, além da verdadeira amizade infinita e revigorante.

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pelo fomento à presente pesquisa.

À Universidade Federal do Ceará e aos professores do curso de Biotecnologia, que tiveram papel direto em minha formação.

Ao Prof. Dr. Thalles Barbosa Grangeiro pelo referencial científico, crítico e racional, sendo uma influência concreta e amiga no campo da pesquisa, da administração e da ciência.

Aos participantes da banca examinadora Prof. Dra. Cristina Paiva e Prof. Dr. Hélio Costa pela disponibilidade, interesse e sugestões para aperfeiçoamento dessa pesquisa.

Aos que me orientaram e iniciaram na pesquisa científica, Prof. Dra. Cristina, Dra. Diva Correia e Prof. Dr. Thalles pelo carinho, dedicação, paciência e empolgação de ensinar uma fração dos seus conhecimentos.

Aos demais integrantes do Laboratório de Genética Molecular (LabGeM), Filipe, Edvar, Beatriz, Ádria, Arthur, Viviane, Carol, Cecília, Ricardo e Taynara pelo apoio, orientação e por todos os cafês tomados às 15 h, pontualmente.

À minha família, em especial meu pai, Benedito, e minha mãe, Cleyse, por acreditarem no meu potencial e terem investido tanto em meu futuro.

Aos Kobras, André, Markin e Bruno pela amizade verdadeira e duradoura que serviu de base para superação de diversas dificuldades durante a graduação.

Aos meus queridos amigos da paróquia Imaculada Conceição e São João XXIII, principalmente Gabriel e Cecília, que foram fonte de amor, dedicação e diversão nas atividades de coordenação do grupo de Coroinhas, do qual tive a sorte de coordenar em conjunto com eles durante essa caminhada formativa.

## RESUMO

Quitinases (EC 3.2.1.14) são hidrolases glicosídicas que atuam sobre a quitina, um polissacarídeo estrutural presente em fungos e artrópodes. Essas enzimas são divididas em duas principais famílias em vegetais: GH18 (que inclui as classes III e V) e GH19 (que inclui as classes I, II, IV, VI e VII). Cada classe possui diferenças específicas em sua estrutura primária e arquitetura de domínios. Estudar a distribuição de quitinases em *Arabidopsis thaliana* e em genomas relacionados é uma forma representativa de avaliar a diversidade de organização de domínios de quitinases vegetais e inferir raciocínios filogenéticos que auxiliam na compreensão do grupo de genes que codifica essas enzimas, além de ser de interesse científico compreender a função e a dinâmica de suas proteínas codificadas no metabolismo vegetal. As sequências foram obtidas nos bancos de dados do GenBank, Phytozome, Ensembl plants e Plants Gramene. Em seguida, foram preditas, nas sequências proteicas, a presença de peptídeos sinal, a localização subcelular e os domínios conservados (CDDs). Os alinhamentos foram feitos via MAFFT, e árvores filogenéticas via IQ-TREE, com imagens geradas no MEGA. Os alinhamentos para as árvores foram submetidos a inferência de taxa de substituição sinônima e não sinônima pelo teste aBSREL e FEL, do DataMonkey. Com o pacote Genes\_on\_Chr foram inferidas as localizações cromossômicas dos genes estudados para *Arabidopsis thaliana*, *Arabidopsis arenosa* e para o alotetraploide originado dos cruzamento das duas espécies citadas, *Arabidopsis suecica*. *Arabidopsis thaliana* possui 24 genes que codificam quitinases, agrupadas nas classes I (uma quitinase), II (duas), III (uma), IV (oito), V (nove), VI (duas) e VII (uma), além das *Chitinase-like-protein* (CLP) CLP1, CPL2 e CPL3, com um exemplar de cada. Com exceção de *Arabidopsis suecica*, todas as outras espécies estudadas apresentaram número menor ou igual de quitinases codificadas no genoma, com a halófito relativa *Eutrema salsugineum* não apresentando quitinases de classe VII. As análises de taxa de substituição revelaram pressão purificadora em quitinases de classe I, II, III, VI e VII e pressão diversificadora para quitinases de classe IV e V. Revelando que para membros da classe IV e V existe uma maior influência de mutações acumuladas em suas sequências polipeptídicas, alinhando-se com o fato de proteínas dessas classes estarem localizadas em tandem nos cromossomos, sendo origem de eventos de duplicação, o que se relaciona com mecanismos de senescência, germinação e desenvolvimento vegetal, e não diretamente na defesa vegetal, diferentemente das classes I, II, III e VI. A ausência de quitinases de classe VII no genoma de *Eutrema salsugineum* pode indicar perda de função de enzimas dessa classe.

**Palavras-chave:** Filogenia; Domínios Conservados; PR proteins

## ABSTRACT

Chitinases (EC 3.2.1.14) are glycosyl hydrolases that act on chitin, a structural polysaccharide present in fungi and arthropods. These enzymes are divided into two main families in plants: GH18 (which includes classes III and V) and GH19 (which includes classes I, II, IV, VI, and VII). Each class possesses specific differences in its primary structure and domain architecture. Studying the distribution of chitinases in *Arabidopsis thaliana* and related genomes is a representative way to evaluate the diversity of domain organization in plant chitinases and to infer phylogenetic reasoning that helps in the understanding of the gene group encoding these enzymes, in addition to the scientific interest in understanding the function and dynamics of their encoded proteins in plant metabolism. The sequences were obtained from the GenBank, Phytozome, Ensembl Plants, and Plants Gramene databases. Subsequently, the presence of signal peptides, subcellular localization, and conserved domains (CDDs) were predicted in the protein sequences. Alignments were performed via MAFFT, and phylogenetic trees via IQ-TREE, with images generated in MEGA. The alignments for the trees were submitted to synonymous and non-synonymous substitution rate inference using the aBSREL and FEL tests from DataMonkey. With the Genes\_on\_Chrom package, the chromosomal localizations of the studied genes were inferred for *Arabidopsis thaliana*, *Arabidopsis arenosa*, and for the allotetraploid originated from the crossing of the two mentioned species, *Arabidopsis suecica*. *Arabidopsis thaliana* possesses 24 genes that encode chitinases, grouped into classes I (one chitinase), II (two), III (one), IV (eight), V (nine), VI (two), and VII (one), in addition to the Chitinase-like-proteins (CLP) CLP1, CLP2, and CLP3, with one unity each. With the exception of *Arabidopsis suecica*, all other studied species presented a lower or equal number of chitinases encoded in the genome, with the relative halophyte *Eutrema salsugineum* not presenting class VII chitinases. Substitution rate analyses revealed purifying pressure in class I, II, III, VI, and VII chitinases and diversifying pressure for class IV and V chitinases. This reveals that for members of classes IV and V, there is a greater influence of mutations accumulated in their polypeptide sequences, aligning with the fact that proteins of these classes are located in tandem on the chromosomes, being the origin of duplication events, which relates to mechanisms of senescence, germination, and plant development, and not directly to plant defense, unlike classes I, II, III, and VI. The absence of class VII chitinases in the genome of *Eutrema salsugineum* may indicate loss of function of enzymes of this class.

**Keywords:** Phylogeny; Conserved Domains; PR proteins

## LISTA DE ABREVIATURAS E SIGLAS

Ala/A	Alanina
Asn/N	Asparagina
Arg/R	Arginina
Asp/D	Aspartato
Cys/C	Cisteína
Gly/G	Glicina
Glu/E	Glutamato
His/H	Histidina
Lys/K	Lisina
Pro/P	Prolina
Phe/F	Fenilalanina
Ser/S	Serina
Thr/T	Treonina
Tyr/Y	Tirosina
Val/V	Valina
dS/ $\alpha$	Taxa de substituição sinônima
dN/ $\beta$	Taxa de substituição não sinônima
CatD	Domínio Catalítico
ChtBD	Domínio de ligação à quitina
GH (18 ou 19)	Família de Hidrolase de Glicosídeo (18 ou 19)
N-terminal; C-terminal	Amino-terminal (porção inicial da cadeia de AA) e Carboxi-terminal
(Porção	final da cadeia de AA)
CLP	<i>Chitinase-Like-Protein</i>
CDS	<i>Coding DNA Sequence</i>
AA	Aminoácido
ML	Máxima Verossimilhança
NCBI	<i>National Center for Biotechnology Information</i>
CDD	<i>Conserved Domains Database</i>
MAFFT	<i>Multiple Alignment using Fast Fourier Transform</i>
aBSREL	<i>Adaptative Branch-Site Random Effects Likelihood method</i>
FEL	<i>Fixed Effects Likelihood</i>

## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	14
2	<b>OBJETIVOS</b> .....	16
2.1	Objetivo Geral .....	16
2.2	Objetivos específicos .....	16
3	<b>MATERIAL E MÉTODOS</b> .....	17
3.1	Mineração das sequências em bancos de dados públicos .....	17
3.2	Predição de características gerais .....	17
3.3	Análises filogenéticas e alinhamentos .....	18
3.4	Predição de taxa de substituição sinônima (dS) e não sinônimas (dN) de mutações.....	18
3.5	Mapeamento cromossômico de genes que codificam quitinases em <i>Arabidopsis thaliana</i> , <i>Arabidopsis arenosa</i> e <i>Arabidopsis suecica</i> .....	19
3.6	<i>Eutrema salsugineum</i> .....	19
4	<b>RESULTADOS</b> .....	20
4.1	<i>Arabidopsis thaliana</i> .....	20
4.2	<i>Arabidopsis halleri</i> , <i>Arabidopsis suecica</i> , <i>Arabidopsis lyrata</i> , <i>Arabidopsis arenosa</i> e <i>Eutrema salsugineum</i> .....	25
4.2.1	<i>Estrutura primária, taxas dS e dN</i> .....	25
4.2.1.1	<i>Classe I</i> .....	25
4.2.1.2	<i>Classe II</i> .....	26
4.2.1.3	<i>Classe III</i> .....	28
4.2.1.4	<i>Classe IV</i> .....	29
4.2.1.5	<i>Classe V</i> .....	30
4.2.1.6	<i>Classe VI</i> .....	31
4.2.1.7	<i>Classe VII</i> .....	33
4.2.2	<i>Mapeamento cromossômico</i> .....	34
5	<b>DISCUSSÃO</b> .....	36
5.1	<i>Classe I</i> .....	37
5.2	<i>Classe II</i> .....	37
5.3	<i>Classe III</i> .....	38
5.4	<i>Classe IV e classe V</i> .....	38
5.5	<i>Classe VI</i> .....	39

5.6	Classe VII .....	40
5.7	<i>Eutrema salsugineum</i> .....	40
6	CONCLUSÃO .....	42
	REFERÊNCIAS .....	43
	APÊNDICE A – ALINHAMENTO DO ChtBD DE HEVEÍNA de <i>Hevea brasiliensis</i> COM QUITINASES DE CLASSE I DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> , <i>A. arenosa</i> E <i>E. salsugineum</i> .....	46
	APÊNDICE B – ALINHAMENTO DE QUITINASES DE CLASSE II DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> , <i>A. arenosa</i> E <i>E. salsugineum</i> .....	47
	APÊNDICE C – ALINHAMENTO DE QUITINASES DE CLASSE III DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> , <i>A. arenosa</i> E <i>E. salsugineum</i> .....	48
	APÊNDICE D – ALINHAMENTO DO ChtBD DE HEVEÍNA DE <i>Hevea brasiliensis</i> COM QUITINASES DE CLASSE IV DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> , <i>A. arenosa</i> E <i>E. salsugineum</i> .....	50
	APÊNDICE E – ALINHAMENTO DE QUITINASES DE CLASSE IV COM DOMÍNIOS ADICIONAIS DE <i>A. halleri</i> , <i>A. suecica</i> E <i>A. arenosa</i> .....	53
	APÊNDICE F – ALINHAMENTO DE QUITINASES DE CLASSE V DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> , <i>A. arenosa</i> E <i>E. salsugineum</i> .....	55
	APÊNDICE G – ALINHAMENTO DO ChtBD DE HEVEÍNA DE <i>Hevea brasiliensis</i> COM QUITINASES DE CLASSE VI DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> , <i>A. arenosa</i> E <i>E. salsugineum</i> .....	59
	APÊNDICE H – ALINHAMENTO DO ChtBD DE HEVEÍNA DE <i>Hevea brasiliensis</i> COM QUITINASES DE CLASSE VI COM DOMÍNIOS ADICIONAIS DE <i>A. thaliana</i> E <i>A. arenosa</i> .....	60
	APÊNDICE I – ALINHAMENTO DE QUITINASES DE CLASSE VII DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> E <i>A. arenosa</i> .....	62
	APÊNDICE J – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE I DE <i>A. thaliana</i> , <i>A. halleri</i> , <i>A. lyrata</i> , <i>A. suecica</i> , <i>A. arenosa</i> E <i>E. salsugineum</i> .....	63

<b>APÊNDICE K – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE II DE <i>A. thaliana</i>, <i>A. halleri</i>, <i>A. lyrata</i>, <i>A. arenosa</i> E <i>E. salsugineum</i> .....</b>	<b>64</b>
<b>APÊNDICE L – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE III DE <i>A. thaliana</i>, <i>A. halleri</i>, <i>A. lyrata</i>, <i>A. arenosa</i> E <i>E. salsugineum</i> .....</b>	<b>65</b>
<b>APÊNDICE M – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE IV DE <i>A. thaliana</i>, <i>A. halleri</i>, <i>A. lyrata</i>, <i>A. arenosa</i> E <i>E. salsugineum</i> .....</b>	<b>66</b>
<b>APÊNDICE N – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE V DE <i>A. thaliana</i>, <i>A. halleri</i>, <i>A. lyrata</i>, <i>A. suecica</i>, <i>A. arenosa</i> E <i>E. salsugineum</i>.....</b>	<b>67</b>
<b>APÊNDICE O – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE VI DE <i>A. thaliana</i>, <i>A. halleri</i>, <i>A. lyrata</i>, <i>A. suecica</i>, <i>A. arenosa</i> E <i>E. salsugineum</i>.....</b>	<b>68</b>
<b>APÊNDICE P – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE VII DE <i>A. thaliana</i>, <i>A. halleri</i>, <i>A. lyrata</i>, <i>A. suecica</i> E <i>A. arenosa</i> .....</b>	<b>69</b>
<b>APÊNDICE Q – MAPA CROMOSSÔMICO COM GENES DE QUITINASES DE <i>Arabidopsis arenosa</i> .....</b>	<b>70</b>
<b>APÊNDICE R – MAPA CROMOSSÔMICO COM GENES DE QUITINASES DE <i>Arabidopsis suecica</i> .....</b>	<b>71</b>
<b>APÊNDICE S – QUANTIDADE DE GENES QUE CODIFICAM CADA UMA DAS CLASSES DE QUITINASES EM <i>Arabidopsis thaliana</i>, <i>Arabidopsis suecica</i>, <i>Arabidopsis arenosa</i>, <i>Arabidopsis lyrata</i>, <i>Arabidopsis halleri</i> e <i>Eutrema salsugineum</i> .....</b>	<b>72</b>
<b>ANEXO A – DOMÍNIOS CONSERVADOS DA PROTEÍNA CAE6117724_ <i>A.arenosa</i>_III .....</b>	<b>73</b>
<b>ANEXO B – DOMÍNIOS CONSERVADOS DAS PROTEÍNAS KAG7574778.2_ <i>A.suecica</i>_class_IV, CAE6047466.3_ <i>A.arenosa</i>_class_IV, CAL9217362.1_ <i>A.halleri</i>_class_IV e CAE6076249.1_ <i>A.arenosa</i>_class_IV, respectivamente .....</b>	<b>74</b>
<b>ANEXO C – DOMÍNIOS CONSERVADOS DA PROTEÍNA CAE5956658_ <i>A.arenosa</i>_VI.....</b>	<b>75</b>

## 1 INTRODUÇÃO

Quitinases (EC 3.2.1.14) são classificadas como hidrolases de glicosídeos (*glycoside hydrolases*, GHs) que clivam aleatoriamente resíduos de N-acetil- $\beta$ -D-glucosamina (NAG) da quitina [nomenclatura IUPAC: (1 $\rightarrow$ 4)-2-acetamido-2-desoxi- $\beta$ -D-glucano], polissacarídeo presente em diversos seres vivos, principalmente em invertebrados como moluscos, nematoides, artrópodes, esponjas e fungos (Crini, 2019; Moussian, 2019).

Essas enzimas são encontradas em grupos bem distribuídos na natureza, estando presentes desde plantas, animais, vírus e até em organismos pertencentes aos domínios Archaea, Bacteria e Eukarya. Em vegetais, adicionalmente, constituem uma das principais frentes moleculares de resposta das plantas contra patógenos, especialmente fungos (Hermans et al., 2010; Passarinho; De Vries, 2002; Su et al., 2015). Além de ação fungicida e fungistática, essas proteínas podem também apresentar influência no desenvolvimento, na senescência, na germinação, na remodelação da parede celular e na adaptação a estresses abióticos (Sivaramakrishnan; Veeraganti Naveen Prakash; Chandrasekar, 2024).

Por possuir um potencial biotecnológico elevado, essas enzimas são amplamente estudadas em nível de estrutura tridimensional (Landim et al., 2017) e estrutura primária (Passarinho; De Vries, 2002). Bem como são exploradas diversas estratégias para melhoramento de atividade enzimática, purificação dessas enzimas e sua produção heteróloga (Motlagh et al., 2025; Yang et al., 2025).

As quitinases presentes em vegetais distribuem-se em duas famílias principais de hidrolases de glicosídeos (GHs) principais, de acordo com o banco de dados CAZy (*Carbohydrate-Active enZymes database*; [www.cazy.org](http://www.cazy.org)) (Drula et al., 2022): GH18 e GH19, diferenciadas estruturalmente e evolutivamente. Membros da família GH18 podem ser encontrados em diversos outros organismos animais, além de plantas. Possuem um sítio catalítico conservado que envolve dois resíduos ácidos, frequentemente Asp (D) e Glu (E), dispostos em uma topologia barril beta. Já as quitinases da família GH19 são típicas de plantas e bacteriófagos, possuindo um enovelamento distinto (tipo  $\alpha+\beta$ ), além de apresentarem dois resíduos de Glu (E) catalíticos usualmente não homólogos aos da GH18. Essas diferenças estruturais refletem diretamente na forma como cada grupo interage com o substrato (Grover, 2012; Cletus et al., 2013).

Em vegetais, quitinases são geralmente proteínas modulares, apresentando um domínio catalítico (*catalytic domain*; CatD), domínio de ligação ao substrato, para algumas proteínas da família GH19 (*Chitin Biding Domain*; ChtBD) e um ou mais domínios auxiliares.

No mais, são proteínas codificadas por grandes famílias gênicas, com membros presentes em diversos cromossomos e que codificam enzimas tanto GH18 como GH19 (Lv et al., 2022; Mir et al., 2020).

A arquitetura de domínios em quitinases permite uma separação de suas enzimas em sete classes (I-VII), das quais são distribuídas dentre as duas principais famílias de hidrolases de glicosídeos em vegetais, GH18 e GH19. Quitinases GH18 possuem classe III e classe V, com CatD GH18 de classe III e CatD GH18 de classe V. Membros da família GH19 dispõem-se entre as classes I, II, IV, VI e VII, possuindo, portanto, um CatD da família GH19 (Landim et al., 2017). Proteínas de classe I apresentam tipicamente um ChtBD na porção N-terminal do tipo *hevein-like*, seguido de uma região *linker* rica em resíduos de treonina, prolina e glicina, um domínio catalítico GH19 e, por fim, uma extensão C-terminal. Membros da classe II, por sua vez, apresentam uma deleção do domínio de ligação, embora conservem a estrutura geral do CatD de classe I (Shinshi et al., 1990). Quitinases de classe IV são semelhantes a membros da classe I, mas possuem um tamanho reduzido devido uma deleção no ChtBD *hevein-like* N-terminal e 3 deleções no CatD, além da ausência total da extensão C-terminal (Collinge et al., 1993; Margis-Pinheiro et al., 1991; Taira et al., 2011). Segundo Tyler *et al.* (Tyler et al., 2010) membros da classe VI são caracterizados por um domínio catalítico GH19 e um ChtBD, o primeiro apresentando uma substituição do resíduo catalítico canônico das enzimas dessa família, o glutamato (Glu/E), por uma lisina (Lys/K), o segundo sendo distantemente relacionado ao domínio de tipo *hevein-like* presentes na classe I, uma vez que não possuem quatro dos oito resíduos conservados de cisteína responsáveis pela formação de pontes dissulfeto, importantes para a estabilidade da estrutura. Quitinases de classe VII assemelham-se com membros da classe IV, possuindo as mesmas deleções no CatD e a deleção do *loop* C-terminal. Contudo, não possuem domínio de ligação com o substrato (Kolossova; Breuil; Bohlmann, 2014; Peery et al., 2021).

Ademais, com o advento de novas tecnologias de sequenciamento o número de quitinases relatadas em genomas vegetais, bem como a qualidade de sua anotação cresce significativamente. Com isso, o número de dados relativos às classes de I a VII demanda uma organização e classificação sistemática quanto a número de enzimas, organização de domínios proteicos e características gerais nos diversos genomas disponibilizados nos bancos de dados globais. Além disso, é pertinente investigar a existência de possíveis novas classes, bem como questionar a classificação atual ao inferir raciocínios filogenéticos com base em sua evolução e entender melhor sua função na natureza.

## 2 OBJETIVOS

### 2.1 Objetivo Geral

Analisar *in silico* a distribuição e classificação de domínios de quitinases e *Chitinase-like* (CLPs) em *Arabidopsis thaliana*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Arabidopsis suecica*, *Arabidopsis arenosa* e *Eutrema salsugineum*.

### 2.2 Objetivos específicos

- Quantificar o número de quitinases nas espécies *Arabidopsis thaliana*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Arabidopsis suecica*, *Arabidopsis arenosa* e *Eutrema salsugineum*.
- Obter dados de endereçamento subcelular, predição de peptídeo sinal, e classificação de domínios conservados de quitinases.
- Compreender o tipo de seleção que atua sobre os genes que codificam quitinases.
- Alinhar e inferir distribuição de classes com base nas relações filogenéticas comparativas entre as espécies.
- Identificar e comparar a presença de resíduos catalíticos e de ligação presentes nas sequências.
- Mapear a localização cromossômica dos genes envolvidos na codificação de quitinases nessas espécies.

### 3 MATERIAL E MÉTODOS

#### 3.1 Mineração das sequências em bancos de dados públicos

As sequências de *Arabidopsis thaliana*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Arabidopsis suecica*, *Arabidopsis arenosa* e *Eutrema salsugineum* foram obtidas por meio dos bancos de dados públicos do GenBank do NCBI (Sayers et al., 2020) (<https://www.ncbi.nlm.nih.gov/genbank/>) do Phytozome, (Goodstein et al., 2012) (<https://phytozome-next.jgi.doe.gov/>) do Ensembl Plants (Bolser et al., 2016) (<https://plants.ensembl.org/index.html>) e do Plants Gramene (Ware et al., 2002) (<https://www.gramene.org/>).

Para encontrar ortólogos de quitinases nos genomas, proteomas e transcriptomas das plantas estudadas no trabalho, utilizando BLAST (Altschul et al., 1990) (*Basic Local Alignment Search Tool*), ferramenta de busca comparativa entre uma sequência de polipeptídeos ou nucleotídeos contra um banco de dados de proteínas ou nucleotídeos, foram realizadas as variações BLASTp, isca de proteína contra banco de dados de proteína, e tBLASTn, isca de proteína contra banco de dados de nucleotídeos. As iscas utilizadas foram sequências representantes de cada classe, sendo OsChia1b (*Oryza sativa*) quitinase modelo de classe I, RSC-c (*Secale cereale*), de classe II, AAG02504.1 (*Oryza sativa*), de classe III, PvChia4 (*Phaseolus vulgaris*), de classe IV, AtCTL1 (quitinase AT1G05850.1 de *Arabidopsis thaliana*) de classe VI e PcChia7-1 (*Picea engelmannii*) de classe VII. As sequências codificadoras das proteínas (*Coding Sequences*; CDS) e as sequências polipeptídicas foram obtidas após remoção de variantes idênticas.

O banco de dados Ensembl Plants disponibiliza versões alternativas da sequência polipeptídica de proteínas em *Arabidopsis thaliana*, sendo a versão canônica indicada no banco de dados. Os alinhamentos posteriores foram realizados somente com as versões canônicas dos transcritos para *A. thaliana*.

#### 3.2 Predição de características gerais

Para todas as sequências primárias de quitinases, foram preditos peptídeo sinal, utilizando o software SignalP 6.0 (Nielsen et al., 2024) (<https://services.healthtech.dtu.dk/services/SignalP-6.0/>), endereçamento subcelular, via DeepLoc 2.1 (Almagro Armenteros et al., 2017)

(<https://services.healthtech.dtu.dk/services/DeepLoc-2.1/>), programas fornecidos pelo DTU Healf Tech (*Danmarks Tekniske Universitet*) baseados em *deep learning*. Domínios conservados e mapeamento de resíduos envolvidos na catálise e na ligação com substrato foram preditos nos bancos CDD (*Conserved Domains Database*) (Marchler-Bauer et al., 2015) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) e PANTHER (Thomas et al., 2022) (<https://pantherdb.org/>).

### 3.3 Análises filogenéticas e alinhamentos

Os alinhamentos foram dirigidos utilizando o software MAFFT (*Multiple Alignment using Fast Fourier Transform*) (Kato; Rozewicki; Yamada, 2019) (<https://mafft.cbrc.jp/alignment/server/index.html>) e com o método de refinamento iterativo L-INS-i, ideal para grupos proteicos com domínios conservados. As sequências de *Arabidopsis thaliana* foram inicialmente alinhadas respeitando a separação entre família GH19 e GH18, mantendo somente a união de classes. Em seguida, as demais quitinases de outras espécies foram separadas por classes e alinhadas em conjunto com seus respectivos representantes de cada classe do modelo *A. thaliana* (Apêndices A, B, C, D, E, F, G, H e I).

As árvores filogenéticas foram feitas via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025) (Apêndices J, K, L, M, N, O e P).

### 3.4 Predição de taxa de substituição sinônima (dS) e não sinônimas (dN) de mutações

Foram retiradas as colunas que apresentavam *gaps* nos alinhamentos separados por classes das sequências polipeptídicas e CDS (Apêndices A, B, C, D, E, F, G, H e I). As sequências alinhadas foram submetidas ao teste aBSREL (*adaptive Branch-Site Random Effects Likelihood method*) e ao teste FEL (*Fixed Effects Likelihood*), baseados primariamente no teste de Máxima Verossimilhança (ML) incluídos no pacote HyPhy (Pond; Frost; Muse, 2005), disponibilizado pelo DataMonkey (Weaver et al., 2018) (<https://www.datamonkey.org/>).

Alinhamentos de sequências que apresentaram domínios adicionais além dos domínios catalíticos foram apresentados separadamente, para classe III (Anexo A), IV (Anexo B) e

classe V (Anexo C). No entanto, essas sequências foram incluídas nesses testes após retirada *in silico* manual desses domínios adicionais.

### **3.5 Mapeamento cromossômico de genes que codificam quitinases em *Arabidopsis thaliana*, *Arabidopsis arenosa* e *Arabidopsis suecica***

As informações de tamanho de cromossomos e localização dos genes que codificam quitinases de *Arabidopsis thaliana*, *Arabidopsis arenosa* e *Arabidopsis suecica* foram obtidas via GenBank, do NCBI (Sayers et al., 2020). *Arabidopsis suecica* é um alotetraploide (4n), resultante da união de gametas das espécies *Arabidopsis thaliana* e *Arabidopsis arenosa*. A fim de investigar a distribuição de genes que codificam quitinases nos cromossomos e analisar a deleção ou duplicação de algum gene com a origem desse alotetraploide, *A. thaliana*, *A. arenosa* e *A. suecica* foram utilizadas nessa etapa. As figuras 3, 4 e 5 foram geradas via pacote personalizado `Genes_on_Chr` do R, disponível no endereço [https://github.com/xukaili/Genes\\_on\\_Chr/tree/master](https://github.com/xukaili/Genes_on_Chr/tree/master).

Proteínas descritas como *Chitinase-Like-Protein* (CLPs) de *A. suecica* e *A. arenosa* foram utilizados apenas para essa etapa, mediante obtenção da sua localização cromossômica. Transcritos CLP de *A. thaliana* além de serem utilizados nessa etapa também foram alinhados e tiveram sua estrutura primária discutida (Figura 2). A divergência de sequências desses transcritos para as outras espécies estudadas impossibilitou a aplicação das predições de taxa de substituição para eles.

### **3.6 *Eutrema salsugineum***

Essa espécie foi incluída nas análises dada sua proximidade filogenética com o gênero *Arabidopsis*, mesmo fazendo parte do gênero *Eutrema*, compartilhando aproximadamente 80% dos genes (Yang et al., 2013). Adicionalmente é considerada halófita relativa, tolerando níveis de estresse salino maiores (Khanal; Moffatt; Gray, 2015). Além disso, *Eutrema salsugineum* já foi classificada, anteriormente, como uma participante do gênero *Arabidopsis*, dada sua semelhança anatômica e fisiológica com representantes clássicos desse gênero e de outras brassicáceas (Kuznetsov et al., 1901).

## 4 RESULTADOS

### 4.1 *Arabidopsis thaliana*

As análises revelaram que no genoma de *A. thaliana*, existem 27 genes codificando quitinases e proteínas *Chitinase-like*, sendo 14 para quitinases GH19 e 13 para quitinases GH18 (Tabelas 1 e 2). Por comparação com as quitinases canônicas por meio de alinhamentos de suas estruturas primárias, foi possível classificar todas as quitinases GH18 e GH19 codificadas no genoma da planta modelo (Figuras 1 e 2). Desta maneira, as quitinases de *A. thaliana* foram identificadas como pertencentes às classes I (uma enzima), II (duas enzimas), III (uma enzima), IV (oito enzimas), V (nove enzimas), VI (duas enzimas) e VII (uma enzima) (Apêndice S), além de três quitinases organizadas da seguinte forma GH18\_SI\_CLP1 (uma proteína), GH18\_SI\_CLP2 (uma proteína) e GH18\_SI\_CLP3 (uma proteína). Distribuição feita de acordo com a presença/ausência de um domínio de ligação a quitina do tipo I (ChtBD1) e da família de GH à qual pertence o domínio catalítico.

**Tabela 1 – Características gerais das quitinases GH19 de *Arabidopsis thaliana*.**

Gene	AA	Pep. sinal	ChtBD	CatD	Endereçamento celular	Classe
AT2G43600.2	294	1-43	ChtBD1 (50-78)	chitinase_GH19 (105-294)	Extracelular	IV
AT2G43620.1	283	1-28	ChtBD1 (30-62)	chitinase_GH19 (94-283)	Extracelular	IV
AT2G43610.1	281	1-28	ChtBD1 (30-62)	chitinase_GH19 (92-281)	Extracelular	IV
AT2G43580.1	265	1-24	ChtBD1 (29-55)	chitinase_GH19 (73-265)	Extracelular	IV
AT2G43570.1	278	1-31	ChtBD1 (36-62)	chitinase_GH19 (81-277)	Extracelular	IV
AT2G43590.1	264	1-24	ChtBD1 (29-55)	chitinase_GH19 (72-264)	Extracelular	IV
AT4G01700.1	280	1-27	-	chitinase_GH19 (41-273)	Extracelular	II
AT1G56680.2	303	1-51	ChtBD1 (59-86)	chitinase_GH19 (111-303)	Extracelular	IV
AT1G02360.3	324	-	-	chitinase_GH19 (85-317)	Lisossomo/ Vacúolo	II
AT1G05850.2	321	1-26	-	chitinase_GH19 (68-304)	Extracelular	VI
AT3G16920.2	357	1-51	-	chitinase_GH19	Extracelular	VI

AT3G54420.1	273	1-28	ChtBD1 (33-57)	(100-337) chitinase_GH19 (76-273)	Extracelular	IV
AT3G47540.2	284	1-20	-	chitinase_GH19 (34-211)	Extracelular	VII
AT3G12500.1	335	1-33	ChtBD1 (34-73)	chitinase_GH19 (89-320)	Extracelular	I

Fonte: Autoral (2025).

Quitinases da família GH19 possuem peptídeo sinal que endereçam a proteína para o meio extracelular, com exceção da proteína AT1G02360.3, de classe II, a qual não teve peptídeo sinal detectado pelo Signal p6, embora a análise com o programa DeepLoc 2.1 tenha revelado um endereçamento para as organelas lisossomo ou vacúolo. O endereçamento de quitinases é relatado, usualmente, para o meio extracelular, haja vista sua relação com a defesa vegetal, sendo caracterizadas como *Pathogenesis Related protein* (PR) (Barbosa et al., 2020; Collinge et al., 1993; Declercq et al., 2023).

**Tabela 2 – Características gerais de quitinases GH18 de *Arabidopsis thaliana*.**

Gene	AA	Pep. sinal	ChtBD	CatD	Endereçamento celular	Classe
AT4G19720.1	363	-	-	chitinase_GH18_class_V (3-343)	Citoplasma/extracelular	V
AT4G19810.2	379	1-24	-	chitinase_GH18_class_V (25-359)	Extracelular	V
AT4G19800.1	398	-	-	chitinase_GH18_class_V (4-337)	Citoplasma/extracelular	V
AT4G19820.1	368	1-22	-	chitinase_GH18_class_V (23-357)	Extracelular	V
AT4G19750.1	362	-	-	chitinase_GH18_class_V (11-348)	Citoplasma/extracelular	V
AT4G19760.1	369	-	-	chitinase_GH18_class_V (11-358)	Citoplasma/extracelular	V
AT4G19730.1	332	-	-	chitinase_GH18_class_V (11-332)	Citoplasma/extracelular	V
AT4G19770.1	261	-	-	chitinase_GH18_class_V (1-250)	Citoplasma/extracelular	V
AT4G19740.1	211	-	-	chitinase_GH18_class_V (23-211)	Citoplasma/extracelular	V
AT5G24090.	302	1-27	-	chitinase_GH18_class	Extracelular	III

1	_III (30-298)					
AT1G08530. 3	257	-	-	Chitinase-like protein	Plastídeo	-
AT5G09995. 3	257	-	-	Chitinase-like protein	Plastídeo	-
AT4G01040. 2	430	-	-	Chitinase-like protein	Reticulo endoplasmático	-

Fonte: Autorial (2025).

Quitinases GH19, ainda, possuem um domínio catalítico do tipo GH19, além de representantes das classes I, IV e VI com domínios de ligação a quitina (ChtBD) do tipo *hevein-like* (Tabela 1). Esse domínio de ligação é rico em oito resíduos de cisteína (Cys; C) importantes para o enovelamento correto do domínio via formação de pontes dissulfeto. No entanto, quitinases da classe VI de *A. thaliana* não possuem quatro desses resíduos, além de terem seu tamanho reduzido (Figura 1).

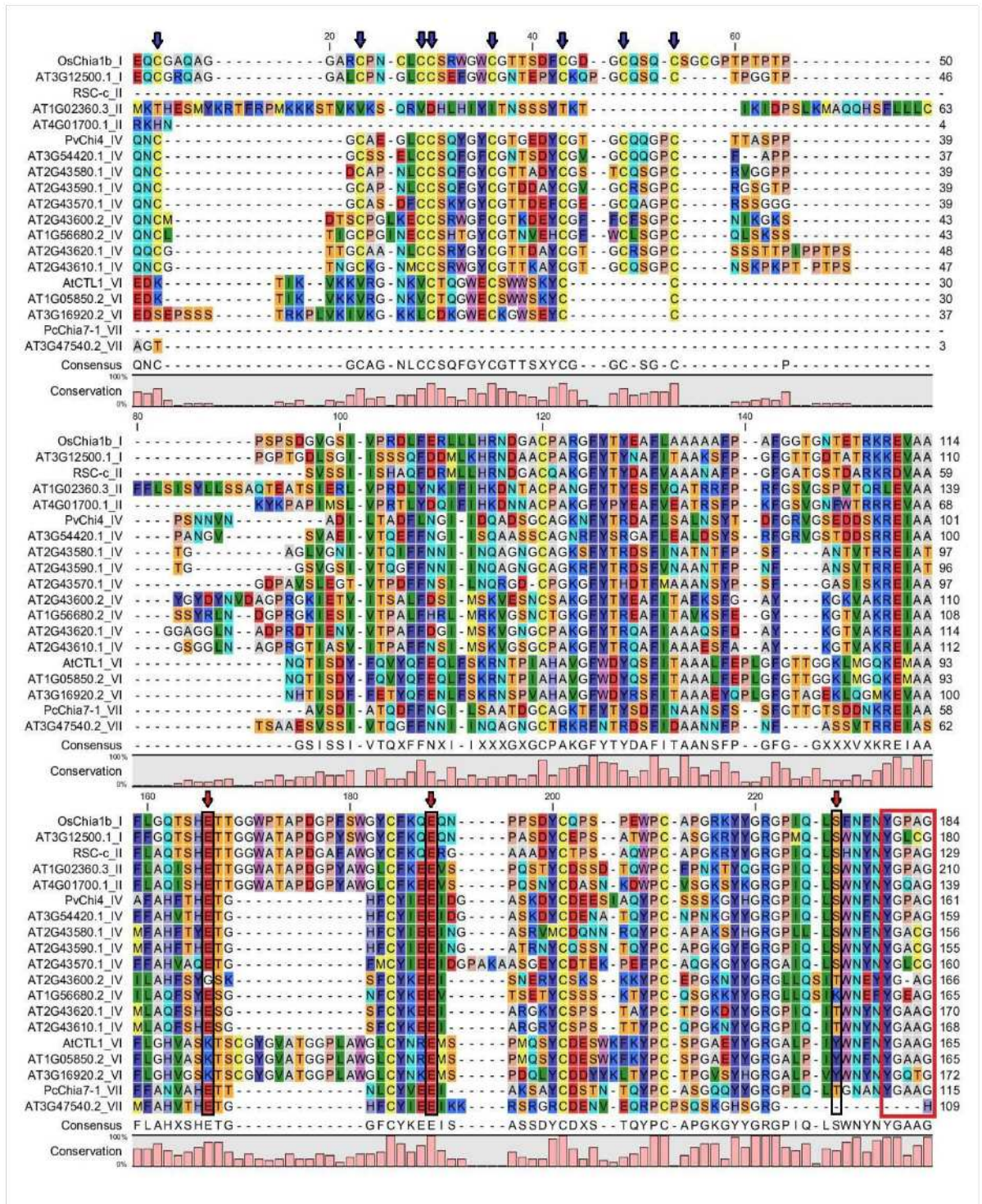
O CatD de representantes da família GH19 é pouco variado, com algumas classes possuindo características pontuais e conservadas, como o *gap* de proteínas de classe IV de quatro resíduos entre as posições 140 e 160 e a deleção de 13 resíduos entre as posições 160 e 185 do alinhamento da Figura 1 (Passarinho; De Vries, 2002). Representantes da classe VII também possuem a mesma deleção de 13 resíduos de aminoácidos na mesma posição (Figura 1). Quitinases dessa família apresentam resíduos importantes para ligação com o substrato, próximos a região C-terminal, região geralmente rica em glicina e alanina (Gly/G e Ala/A). A quitinase de classe VII AT3G47540.2 apresenta deleção desses resíduos, embora o modelo da mesma classe utilizado, Pc-Chia7-1 possua essa região (Figura 1).

A presença dos dois glutamatos (Glu/E) catalíticos é conservada nas classes I, II, IV e VII, com exceção da sequência AT2G43600.2 de classe IV (Figura 1). O primeiro glutamato catalítico relatado nas outras classes é substituído por um resíduo de lisina (Lys/K) em classe VI em *A. thaliana* e de todas as outras espécies estudadas no presente trabalho (Apêndice G). Ainda, um terceiro resíduo catalítico, serina (Ser/S), é relatado por Hahn et al. (Hahn et al., 2000) e é presente em quitinases de classe I, II e em quatro quitinases de classe IV e ausente nas demais classes para *Arabidopsis thaliana*.

Figura 1 – Alinhamento de OsChia1b (quitinase de classe I de *Oryza sativa*), RSC-c (quitinase de classe II de *Secale cereale*), PvChia4 (quitinase de classe IV de *Phaseolus*

*vulgaris*), AtCTL1 (quitinase de classe VI de *Arabidopsis thaliana*; At1g05850.1), PcChia7-1 (quitinase de classe VII de *Picea engelmannii*) com outras quitinases de *A. thaliana*.

Fonte: Autoral. Imagem gerada via CLC Sequence Viewer 8. O domínio de ligação a quitina é evidenciado no início da sequência. Os resíduos destacados com uma seta vermelha são



essenciais na catálise enzimática. Os resíduos evidenciados com uma seta azul no ChtBD são importantes para o enovelamento desse domínio (Passarinho; De Vries, 2002). Resíduos em caixa são importantes para a ligação com o substrato.



ambas as classes, e o segundo resíduo catalítico varia entre aspartato e glutamato, para a classe V. Em representantes da classe III, o segundo resíduo não foi encontrado em proteínas de *A. thaliana* (Figura 2).

Proteínas de tipo quitinase (*Chitinase-Like Protein*; CLP) AT1G08530.3 (GH18\_CLP1), AT5G09995.3 (GH18\_CLP2), AT4G01040.2 (GH18\_CLP3), não tiveram peptídeo sinal detectado, além de possuírem um padrão de endereçamento atípico para quitinases. A estrutura primária dessas enzimas não é homóloga a nenhuma outra classe de quitinase relatada (Figura 2). No entanto, os domínios conservados dessas proteínas se assemelham aos de *Chitinase-Like* de proteínas da família GH18. Seus resíduos catalíticos não são homólogos a nenhum outro relatado para nenhuma das outras classes, além de não haver predição de domínio de ligação com o substrato. Suas sequências apresentaram um padrão de tamanho menor do que outras quitinases, tendo a região C-terminal típica de quitinases da família GH18 ausente. Além disso, diversos *gaps* no alinhamento são encontrados na região do domínio catalítico (Figura 2).

A região N-terminal da proteína deduzida da proteína AT4G01040.2\_CLP3 apresentou um tamanho significativamente maior ao das outras quitinases e CLP's e possui sua sequência primária mais divergente em relação às outras duas proteínas caracterizadas como CLP's apresentados, AT1G08530.3\_CLP1 e AT5G09995.3\_CLP2 (Figura 2).

## ***4.2 Arabidopsis halleri, Arabidopsis suecica, Arabidopsis lyrata, Arabidopsis arenosa e Eutrema salsugineum***

### ***4.2.1 Estrutura primária, taxas dS e dN***

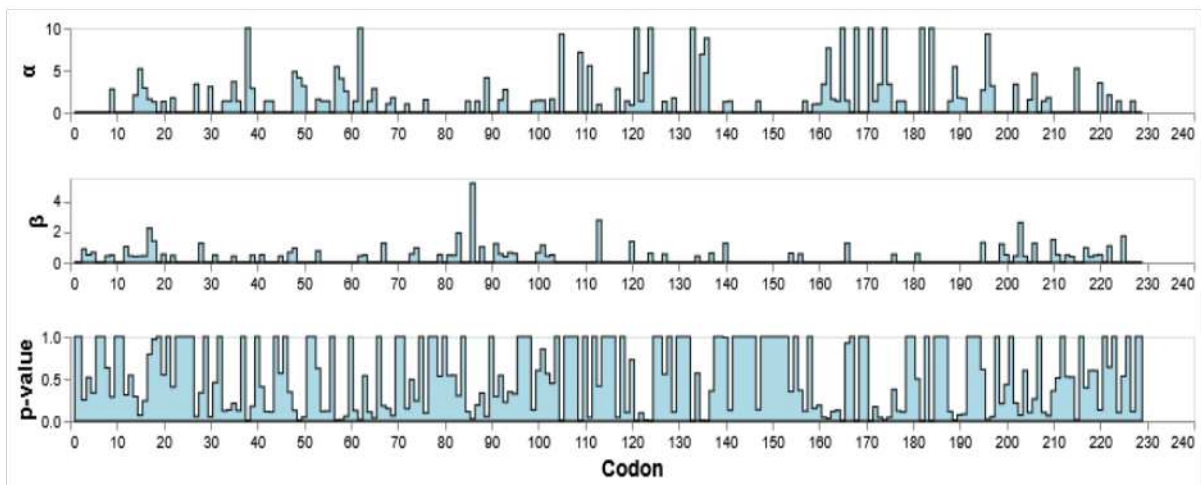
#### ***4.2.1.1 Classe I***

A quantidade de quitinases dessa classe para as espécies estudadas foram *A. suecica*, três; *A. arenosa*, duas; e uma para *A. lyrata*, *A. halleri* e *E. salsugineum* (Apêndices A e S).

A estrutura de domínios é conservada majoritariamente para essas quitinases, sendo CAE5967005.1 de *A. arenosa* e KAG7581678.1 de *A. suecica* as mais variadas, apresentando deleções na região *linker* entre o ChtBD-hevein-like e o CatD de classe I. Na região próxima ao primeiro glutamato catalítico H-E-T-T, KAG7581678.1 apresenta duas mutações nos dois primeiros resíduos H e E para duas glutaminas (Gln/Q), respectivamente, enquanto CAE5967005.1 conserva a primeira mutação mas não sofre modificação no Glu catalítico.

Ainda, para o segundo resíduo catalítico, as mesmas proteínas apresentam uma troca na sequência Q-E por E-Q. Para o terceiro resíduo participante da catálise, Ser (S), ambas as sequências citadas apresentam uma mutação para leucina (Leu/L), além das sequências CAL9220541.1, de *A. halleri* e XP\_006407328.1 de *E. salmugineum* que tiveram esse resíduo substituído por uma treonina (Thr/T) (Apêndice A).

Figura 3 – Estimativas de Máxima Verossimilhança para as taxas sinônimas ( $\alpha$ ) e não-sinônimas ( $\beta$ ) em cada sítio de CatD de quitinases de classe I (Apêndice A). P-valores acima de 10 são fixados neste valor.



Fonte: Fixed Effects Likelihood (FEL).

O teste aBSREL não encontrou nenhum ramo ou sequência do alinhamento presente no Apêndice A sobre pressão episódica diversificadora.

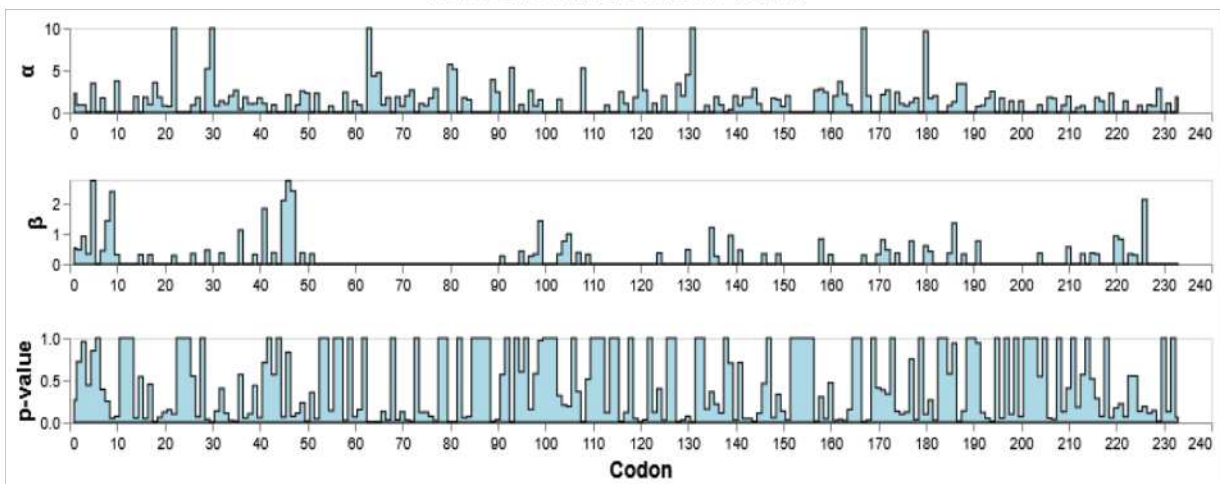
O teste FEL aplicado evidenciou em nível de sequência primária, que para essa classe maioria dos sítios estão sobre seleção purificadora, com altas taxas de substituições sinônimas ( $\alpha/dS$ ). Portanto, os eventos estocásticos de mutações ocorrentes em maior parte dos sítios da sequência nucleotídica do gene não modificaram o aminoácido codificado. Como exemplo, os sítios catalíticos, entre as posições 60 e 65, nas regiões de resíduos de ligação com o substrato 120 até 125, na região C-terminal (Figuras 1 e 3) tiveram altas taxas de substituição sinônima ( $\alpha/dS$ ), o que indica preservação desses sítios na sequência polipeptídica final. Os sítios que sofreram pressão diversificadora, com isso, denotando taxas de mutações não sinônimas ( $\beta/dN$ ) nessas regiões, não estão envolvidos na catálise, mas estão próximos de regiões catalíticas, sendo eles o codon 86, codificando A, G, S ou D; e o codon 203, codificando G, S ou Q.

#### 4.2.1.2 Classe II

Para classe II os números foram, para *A. suecica*, três; *A. arenosa*, *A. lyrata*, *A. halleri* e *E. salsugineum*, duas cada (Apêndices B e S).

A estrutura do CatD, semelhante à classe I, também se manteve, sendo AT1G02360, de *A. thaliana*, CAE5956335.1, de *A. arenosa*, CAL9213418.1, de *A. halleri*, XP\_002892104.2 de *A. lyrata*, KAG7595559.1, de *A. suecica* e XP\_006418351, de *E. salsugineum* mais semelhantes entre si. Os três resíduos catalíticos relatados foram conservados (Apêndice B), além da estrutura de resíduos importantes para a ligação com o substrato.

Figura 4 – Estimativas de Máxima Verossimilhança para as taxas sinônimas ( $\alpha$ ) e não-sinônimas ( $\beta$ ) em cada sítio de CatD de quitinases de classe II (Apêndice B). P-valores acima de 10 são fixados neste valor.



Fonte: Fixed Effects Likelihood (FEL).

O teste aBSREL detectou dois nós sobre pressão de seleção episódica diversificadora na árvore gerada para enzimas dessa classe (Apêndice K), sendo o nó 10, referente ao ramo da sequência CAL9213418\_1\_A\_halleri\_II e o nó 13, de KAG7619285\_1\_A\_suecica\_II, com p-valores respectivos de 0,00010 e 0,00228, evidenciando que as taxas de substituição não sinônimas foram majoritárias para os ramos que originam essas sequências.

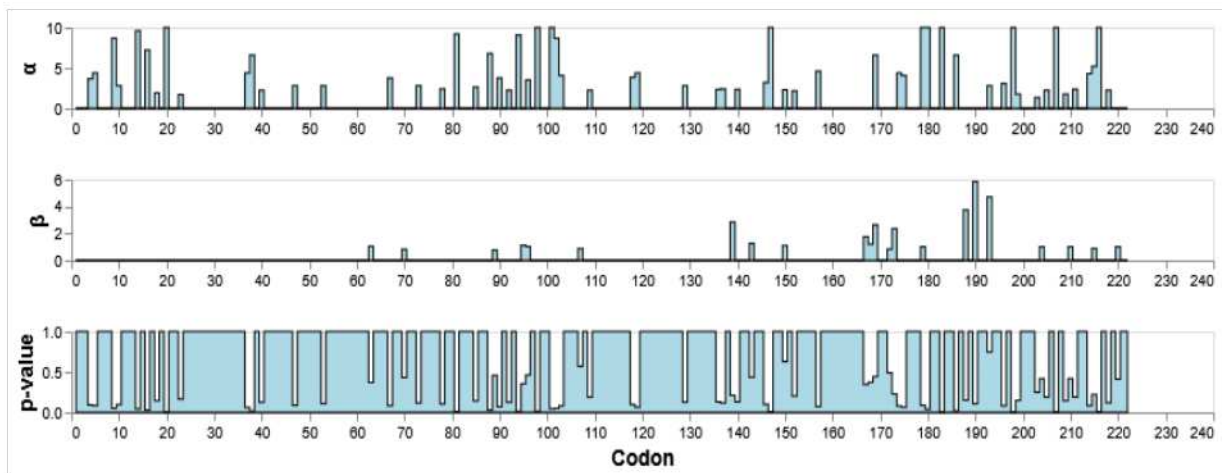
Já o teste FEL evidenciou mutações sinônimas em maioria dos sítios da sequência, mantendo o padrão observado em classe I, com destaque nos sítios catalíticos e de ligação com o substrato. Os sítios onde atuam mutações não sinônimas foram focalizados na porção N-terminal do CatD, observando, com isso, as divergências citadas inicialmente das sequências AT1G02360, CAE5956335.1, CAL9213418.1, XP\_002892104.2, KAG7595559.1

e XP\_006418351, referentes às posições entre os resíduos 5 e 10 e 45 e 50 (Figura 4) (Apêndice B).

#### 4.2.1.3 Classe III

Foram detectadas, em classe III, quitinases no padrão de duas para *E. salsugineum* e *A. suecica* e uma para *A. arenosa*, *A. halleri* e *A. lyrata* (Apêndices C e S).

Figura 5 – Estimativas de Máxima Verossimilhança para as taxas sinônimas ( $\alpha$ ) e não-sinônimas ( $\beta$ ) em cada sítio do CatD de quitinases de classe III (Apêndice C). P-valores acima de 10 são fixados neste valor.



Fonte: Fixed Effects Likelihood (FEL).

A sequência CAE6117724\_A.arenosa\_III apresenta após o CatD de *Chitinase-Like* GH18 um domínio de B\_lectina, um de S\_locus glicoproteína, um de tipo PAN\_APPLE e um de tipo proteína cinase, possuindo, portanto, um tamanho significativamente maior em comparação com os outros representantes da classe III (Apêndice C) (Anexo A). Além disso, a região do domínio catalítico de tipo quitinase GH18 é fortemente conservada, com exceção da sequência XP\_024007186\_E.salsugineum\_III, que apresenta uma longa deleção na região C-terminal do CatD. O aspartato catalítico localizado na posição 155 do alinhamento do Apêndice C se mostrou conservado em todas as sequências.

Nenhum nó foi identificado sobre seleção episódica diversificadora mediante o teste aBSREL para quitinases de classe III.

O p-valor para o teste FEL, em média, para classe III, demonstra uma diferença não significativa para maioria dos sítios do CatD dessas proteínas. Para os sítios que substituição sinônima foram identificados próximos à porção N-terminal, na posição 10 e na posição 20;

próximo à posição 80 e, em destaque, o resíduo catalítico de Asp, próximo à posição 100, no gráfico da Figura 5. Além disso, a porção C-terminal do CatD apresentou altas taxas de substituição sinônimas, especificamente nos sítios 180, 198 e 215 (Figura 5).

#### 4.2.1.4 Classe IV

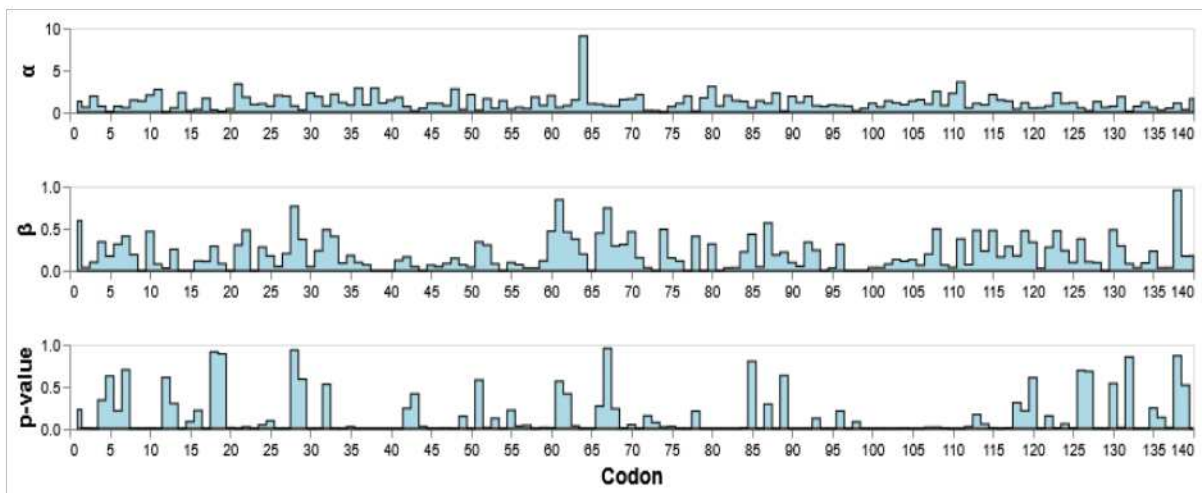
Em classe IV: *A. suecica*, 14; *A. halleri*, 11; *A. arenosa* e *A. lyrata*, 8 e *E. salsugineum* com 10 representantes (Apêndices D e S)

As sequências em classe IV se agrupam em 3 grupos com maior grau de similaridade, sendo as sequências de AT2G43620\_A.thaliana\_class\_IV até XP\_024010844.1\_E.salsugineum\_class\_IV as maiores, contendo uma adição cerca de 14 resíduos de aminoácidos na região do ChtBD (Apêndice D). A variação no CatD dessas sequências é significativa, com as quatro primeiras proteínas do alinhamento com domínios maiores, evidenciando um *gap* de quatro resíduos no restante das sequências. Adicionalmente, uma proteína possui um tamanho reduzido, XP\_024010844.1\_E.salsugineum\_class\_IV, o qual foi retirado do alinhamento final para os testes de substituição sinônima e não sinônima. No mais, a proteína KAG7574777.1\_A.suecica\_class\_IV possui uma deleção de 26 resíduos próxima à região C-terminal. As proteínas AT2G43600.2\_A.thaliana\_class\_IV, KAG7644176.1\_A.suecica\_class\_IV, CAE6047496.1\_A.arenosa\_class\_IV, CAL9236781.1\_A.halleri\_class\_IV, XP\_020884058.1\_A.lyrata\_class\_IV, XP\_006397534.1\_E.salsugineum\_class\_IV, e XP\_006397533.1\_E.salsugineum\_class\_IV, não possuem o primeiro glutamato catalítico conservado das outras classes de GH19 (Apêndice D). Fora a sequência reduzida de *E. salsugineum* que não foi incluída nos próximos testes (Figura 6), outra sequência da mesma espécie, XP\_006402100.1, dessa classe, não possui o segundo glutamato catalítico conservado. Por fim, para o resíduo de serina catalítico, além da proteína XP\_006410096.1\_E.salsugineum\_class\_IV, da proteína AT2G43620\_A.thaliana\_class\_IV até as outras proteínas conseguintes há uma mutação desse resíduo em treonina (Thr/T) (Apêndice D).

Uma quitinase de classe IV com domínios adicionais que não sejam o catalítico foi encontrada, tendo um domínio CAP\_PR-1, sendo ela a proteína CAL9217362.1\_A.halleri\_class\_IV. Outras duas proteínas foram encontradas, possuindo domínios de quitinase GH19 adicionais, sendo KAG7574778.2\_A.suecica\_class\_IV encontrada com um domínio catalítico e um domínio de ligação adicional semelhantes aos primeiros dois domínios N-terminais, e CAE6047466.3\_A.arenosa\_class\_IV com um CatD

adicional logo após o primeiro CatD N-terminal, seguido de um ChtBD semelhante ao primeiro na porção N-terminal e por mais um CatD de quitinase GH19. No mais, uma última sequência CAE6076249.1\_A.arenosa\_class\_IV possui um domínio de função desconhecida na porção N-terminal, anterior ao domínio de ligação típico dessa classe (Apêndice E) (Anexo B).

Figura 6 – Estimativas de Máxima Verossimilhança para as taxas sinônimas ( $\alpha$ ) e não-sinônimas ( $\beta$ ) em cada sítio do CatD de quitinases de classe IV (Apêndice D). P-valores acima de 10 são fixados neste valor.



Fonte: Fixed Effects Likelihood (FEL).

O teste aBSREL não encontrou nenhum ramo sobre seleção positiva episódica em quitinases de classe IV.

O teste FEL, no entanto, acusa, em maioria, mutações não sinônimas próximas às regiões de ligação ao substrato, na porção C-terminal. Além disso, um pico na posição 64 (Figura 6) é referente a uma cisteína altamente conservada em todas as sequências dessa classe, próxima ao último glutamato catalítico. Em média, a classe IV foi a classe que mais apresentou sítios sobre pressão diversificadora.

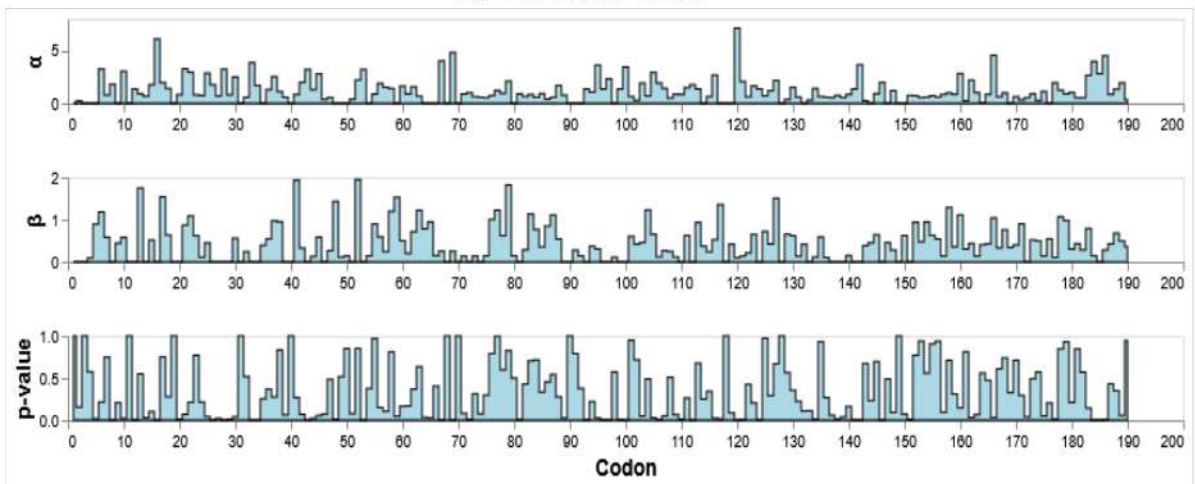
#### 4.2.1.5 Classe V

Em quitinases de classe V foram 11 para *A. suecica*, 6 para *A. halleri*, 3 para *A. lyrata* e *A. arenosa* e uma para *E. salsugineum* (Apêndices F e S).

AT4G19730, de *A. thaliana* e KAG7621107.1 de *A. suecica* são as maiores sequências dentre as outras quitinases de classe V, possuindo uma porção C-terminal maior. Ao passo

que AT4G19740, de *A. thaliana* apresenta uma deleção nessa mesma porção. AT4G19770 e KAG7546872.1, de *A. thaliana* e *A. suecica*, respectivamente, possuem uma deleção significativa na região N-terminal, mas não atingindo o CatD. O aspartato catalítico inicial é conservado, com exceção da sequência KAG7546872.1 de *A. suecica*, pela deleção no N-terminal e CAL9226112.1 e *A. halleri*, a qual sofreu uma mutação do resíduo catalítico para um resíduo de valina (Val/V) (Apêndice F). AT4G19740\_V e KAG7621108.1\_A.suecica\_V possuem uma substituição do segundo aspartato catalítico por um glutamato (Apêndice F).

Figura 7 – Estimativas de Máxima Verossimilhança para as taxas sinônimas ( $\alpha$ ) e não-sinônimas ( $\beta$ ) em cada sítio de quitinasases de classe V (Anexo F). P-valores acima de 10 são fixados neste valor.



Fonte: Fixed Effects Likelihood (FEL).

O teste aBSREL acusou um nó sobre pressão positiva episódica, 26, sendo o p-valor 0,00008. O nó 26 é referente ao ramo de origem da proteína AT4G19730, de *A. thaliana*.

O teste FEL apresentou resultados variados em nível de sequência primária para classe V, com picos variados de substituição sinônima e não sinônima. Em destaque, o sítio próximo à posição 15 do alinhamento de CatD possui uma alta taxa de substituições sinônimas, sendo uma região N-terminal importante para a ligação com o substrato. A região 120 demonstrou também estar sobre alta pressão purificadora, um resíduo de serina conservado em maioria das sequências (Figura 7).

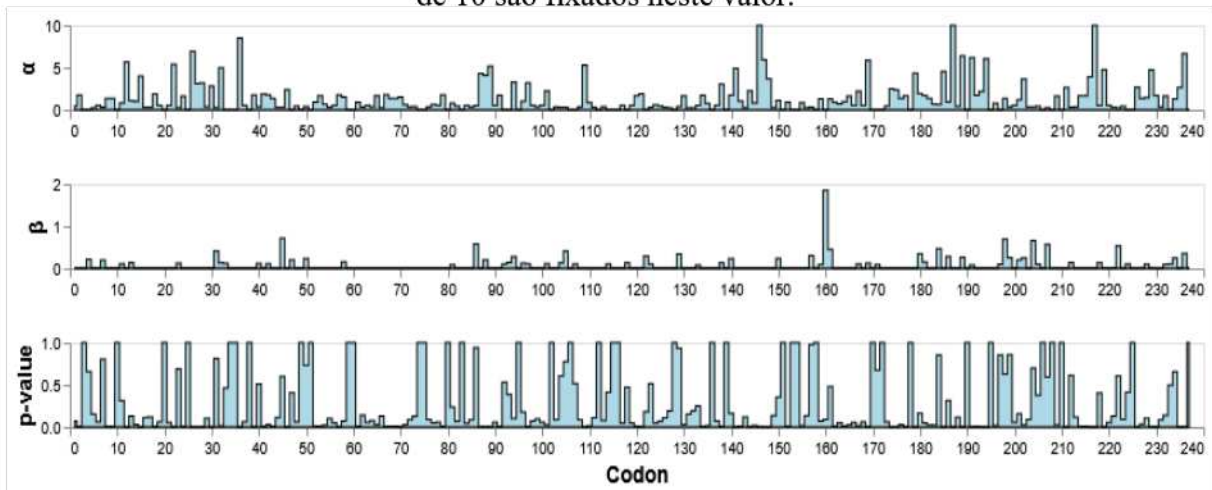
#### 4.2.1.6 Classe VI

A distribuição dos genes foi quatro para *A. suecica*, três para *A. lyrata*, duas para *A. halleri* e *E. salsugineum* e um para *A. arenosa* (Apêndices G e S).

As estruturas primárias de quitinases de classe VI se agrupam em dois padrões de sequência diferentes. As proteínas AT1G05850, de *A. thaliana*, KAG7595984, de *A. suecica*, CAL9213868, de *A. halleri*, XP\_006417979, de *E. salisugineum* apresentam uma sequência menor, com a região do peptídeo sinal reduzida e com maiores variações na região do ChtBD, o qual já possui, para essa classe, uma similaridade menor com o ChtBD de heveína, fortemente relacionado com o ChtBD-*hevein-like* de classe I. No entanto, todas as quitinases de classe VI estudadas mantiveram o padrão de deleção de quatro resíduos de cisteína no ChtBD. Para classe VI a substituição do glutamato catalítico canônico por um resíduo de lisina (Lys/K) se conservou em todas as sequências, da mesma forma para o segundo glutamato catalítico canônico. Proteínas dessa classe tipicamente também não apresentam a serina como terceiro resíduo catalítico, sendo substituído por um resíduo de tirosina (Tyr/Y), o que foi conservado (Tyler et al., 2010).

CAE5956658\_A.arenosa\_VI não foi incluída no alinhamento do Apêndice G pois possui um domínio adicional de tipo pepsina vegetal A, além da deleção do ChtBD.

Figura 8 – Estimativas de Máxima Verossimilhança para as taxas sinônimas ( $\alpha$ ) e não-sinônimas ( $\beta$ ) em cada sítio do CatD de quitinases de classe VI (Apêndice G). P-valores acima de 10 são fixados neste valor.



Fonte: Fixed Effects Likelihood (FEL).

O teste aBSREL não acusou nenhum ramo ou sequência sobre pressão de seleção diversificadora (Apêndice O).

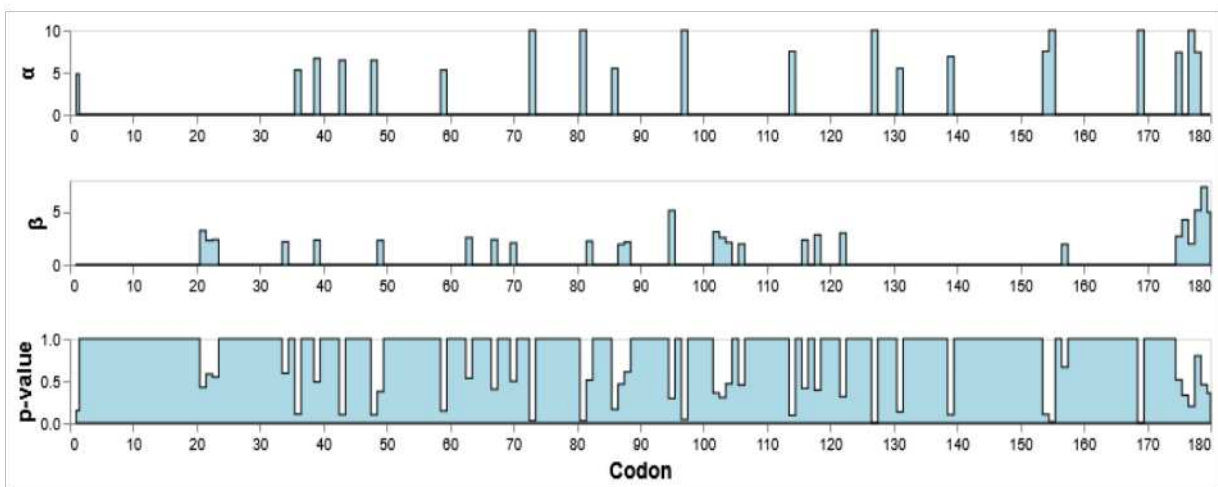
A predição com o teste FEL demonstrou forte influência de substituição sinônima na posição 146, um resíduo conservado de fenilalanina em todas as sequências do alinhamento de classe VI (Apêndice G). O códon que codifica a posição 160 apresenta uma forte taxa de substituição não sinônima, referente aos aminoácidos asparagina (Asn/N), histidina (His/H),

alanina (Ala/A), tirosina (Tyr/Y) e lisina (Lys/L), não relatados como participantes cruciais da catálise ou da ligação com o substrato. Uma fraca taxa de substituição sinônima ocorre entre as posições 160 e 170, referente a um motivo conservado P-S-A-H-D em todos os representantes da classe VI.

#### 4.2.1.7 Classe VII

Para classe VII, dois exemplares foram encontrados em *A. suecica* e uma proteína para as outras espécies estudadas, salvo *E. salisugineum*, a qual não apresentou nenhuma quitinase de classe VII (Apêndices I e S).

Figura 9 – Estimativas de Máxima Verossimilhança para as taxas sinônimas ( $\alpha$ ) e não-sinônimas ( $\beta$ ) em cada sítio do CatD de quitinases de classe VII (Apêndice I). P-valores acima de 10 são fixados neste valor.



Fonte: Fixed Effects Likelihood (FEL).

Maior parte das sequências dessa classe possui alta similaridade entre si. XP\_020882332\_A.lyrata\_VII e KAG7564793\_A.suecica\_VII apresentam uma região do peptídeo sinal maior em comparação com as outras sequências, embora o transcrito canônico de *A. thaliana*, AT3G47540.2 apresente uma região C-terminal maior em comparação com as outras duas proteínas representantes. KAG7633600\_A.suecica\_VII apresenta um maior grau de similaridade com a versão canônica de classe VII de *A. thaliana*, do que em comparação com as outras proteínas das outras espécies (Apêndice I).

Os dois glutamatos catalíticos foram conservados em todas as sequências. No entanto, o resíduo de serina foi substituído por um resíduo de arginina (Arg/R) nas duas sequências

mais semelhantes entre si KAG7633600\_A.suecica\_VII e AT3G47540.2, sendo mantido nas outras proteínas (Apêndice I).

O gene que codifica a proteína AT3G47540.2 de *A. thaliana* foi a única sequência dentre todas as classes cuja taxa de substituição não sinônima foi significativamente maior em comparação a outras quitinases de classe VII, evidenciando que o gene está sobre pressão episódica positiva mediante o teste aBSREL (Apêndice P).

O teste FEL acusou apenas taxas de seleção purificadoras em nível de sequência para esses genes de classe VII. Dentre os sítios com taxa de substituição sinônima os codons que codificam aminoácidos T, A, G, Q e S foram os evidenciados pela sequência polipeptídica. Nenhum desses resíduos nas posições testadas participa canonicamente da catálise de quitina em classe VII (Apêndice I). O teste FEL não acusou nenhum resíduo sobre pressão seletiva diversificadora.

#### 4.2.2 Mapeamento cromossômico

Quitinases e proteínas de tipo quitinase em *Arabidopsis thaliana* apresentam localização cromossômica no padrão de genes em tandem para classe IV e classe V, com algumas exceções, como a sequência AT3G54420\_IV. Representantes de outras classes permanecem distribuídos pelos outros cromossomos separadamente, como o exemplar de classe I AT3G12500, localizado no cromossomo 3, e a quitinase de classe III AT5G24090, encontrada no cromossomo 5 (Figura 10).

Sequências CLP's, tiveram o seguinte padrão de localização, para CLP1, cromossomo 1; CLP3, cromossomo 4 e CLP2, cromossomo 5 (Figura 10).

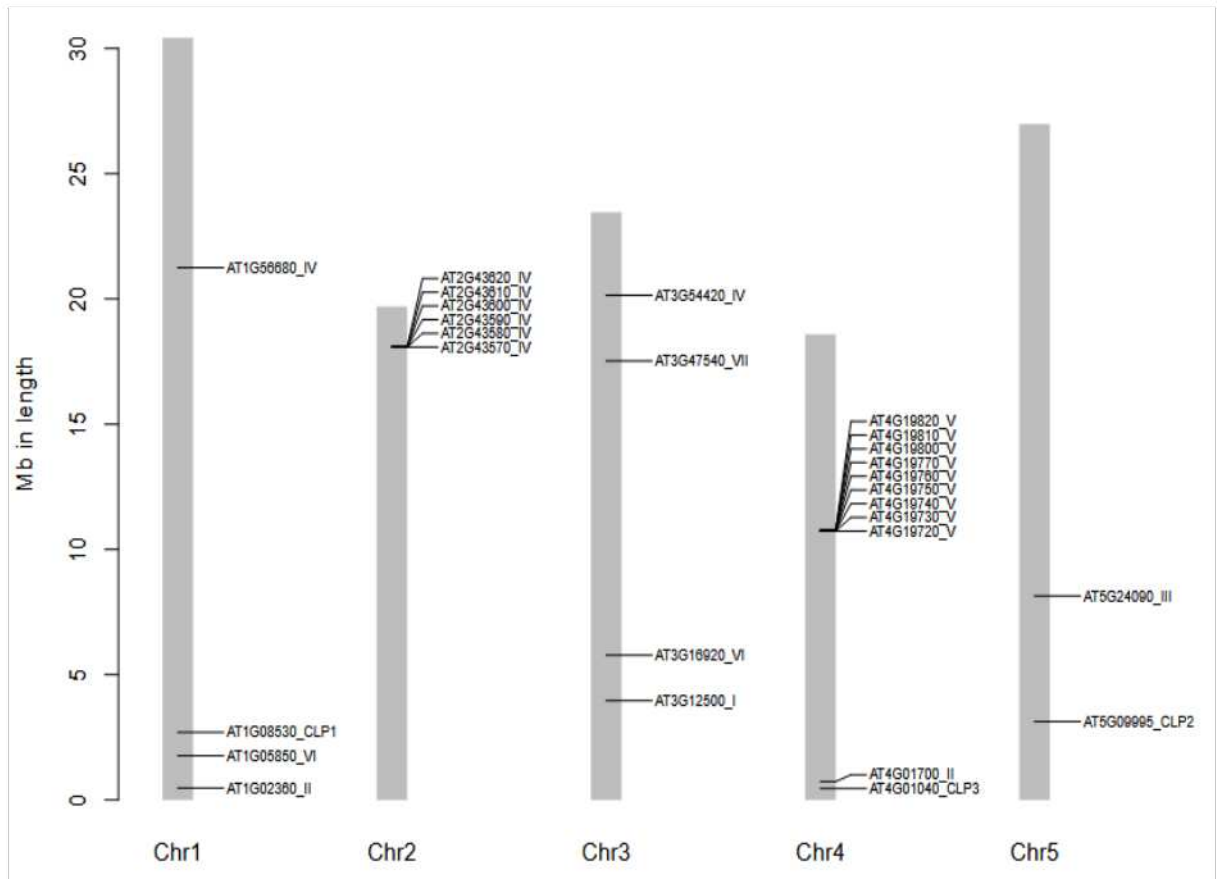
Quitinases de *Arabidopsis arenosa* para a classe IV e V foram também localizadas em tandem no cromossomo (Apêndice Q). Além disso, a quitinase de classe II AT1G02360 em *A. thaliana* se encontra em uma posição alinhada com a quitinase de classe II CAE5956335.1, no cromossomo 1. No entanto, a outra quitinase de classe II, em conjunto com a CLP3 de cada uma das duas espécies variaram de posição entre a porção inferior do cromossomo quatro de *A. thaliana* até a porção superior do cromossomo 6 de *A. arenosa*.

No mais, CLP1 conservou sua posição nas duas espécies, além das duas CLP2 de *A. arenosa* em uma posição do cromossomo 6 similar à posição da CLP2 de *A. thaliana* (Figura 10) (Apêndice Q).

O mapeamento de cromossomos do alotetraploide *A. suecica* evidencia a soma de ambos os padrões de *A. thaliana* e *A. arenosa* em uma espécie resultante, conservando as

localizações das sequências mais semelhantes entre cada uma das espécies parentais e relação à filial. O número de genes que codificam as proteínas, no entanto, não é a soma conservada dos genes de *A. arenosa* e *A. thaliana* (Apêndice R).

Figura 10 – Mapa cromossômico com localização de genes de quitinasases em *Arabidopsis thaliana*.



Fonte: Autoral. Imagem gerada com o pacote Genes\_on\_chr ([https://github.com/xukai/Genes\\_on\\_Chrtree/master](https://github.com/xukai/Genes_on_Chrtree/master)).

## 5 DISCUSSÃO

Quitinases vegetais, usualmente, podem ter sua ação específica no metabolismo vegetal inferida pela quantidade de genes codificados no genoma da espécie em questão, uma vez que sua função pode estar relacionada à defesa vegetal, diretamente, ou em mecanismos gerais celulares, tais como regulação de crescimento, germinação e em modulação em resposta a estresses abióticos (Grover, 2012; Zhong et al., 2002). Por exemplo, quitinases de classe I e classe III possuem poucas cópias nos genomas vegetais, e, além disso, são as mais relatadas contra ataques bióticos (Grover, 2012; Hamid et al., 2013; Passarinho; De Vries, 2002). Em contrapartida, genes que codificam quitinases presentes em maior número de cópias, como em classe IV, classe V e classe VI, podem participar de outros processos celulares de resposta. Em classe IV, exemplificando, um estudo recente relata que uma quitinase de cana-de-açúcar (*Saccharum* spp) atua como sinalizadora em rotas metabólicas contra ataques fúngicos, elevando a atividade, ainda, de outras quitinases da planta (Chen et al., 2025).

O número de genes que codificam quitinases de classe VI em *A. thaliana*, *A. halleri* e *A. lyrata* variou pouco. Essa classe não possui os resíduos essenciais para ligação com substrato e para a catálise, respectivamente histidina e glutamato, mas possui domínio de ligação à quitina (Figura 1). A conservação em seu número pode revelar uma função secundária no metabolismo. Em caju (*Anacardium occidentale*), no entanto, o resíduo de glutamato foi substituído por um resíduo de lisina, alinhando-se com os resultados encontrados em *Arabidopsis* e *Eutrema salsugineum*, e, apesar de a enzima apresentar atividade antifúngica, ainda foi significativamente inferior às demais classes com o resíduo catalítico convencional (Oliveira et al., 2020).

Quitinases da família GH19 podem ou não possuir domínio de ligação, e maioria é endereçada para o meio extracelular (Tabela 1), padrão de proteínas de defesa em plantas (Vaghela et al., 2022). A quantidade de genes que codificam quitinases em *A. thaliana* é concordante com análises de Passarinho e De Vries (Passarinho; De Vries, 2002) e de Tyler *et al.* (Tyler et al., 2010), revelando conservação nesse número de enzimas para essa espécie. Contudo, Passarinho e De Vries não detectaram as CLP's no genoma de *Arabidopsis thaliana*, ou não as abordaram como relacionadas à quitinases.

A identificação de genes homólogos aos de *A. thaliana* nos genomas das espécies *Arabidopsis lyrata* e *Arabidopsis halleri* foi realizada por meio de consultas no banco de dados genômicos de vegetais Gramene. Em *A. lyrata*, o número total de genes para quitinases

é menor (19), enquanto em *A. halleri* o número é conservado. No entanto, para *Arabidopsis halleri* houve redução em três unidades para quitinases de classe V, ao passo que três unidades adicionais foram encontradas na classe IV. Essa alteração pode explicar a plasticidade da quantidade de quitinases necessárias no genoma vegetal para as classes IV e V, assumindo sua função geral no metabolismo, o que demonstra uma possível sobreposição de função para genes que codificam suas proteínas (Apêndice S).

Em *A. arenosa*, no alotetraploide *A. suecica* e na halófita relativa *E. salsugineum*, o total para cada um, respectivamente foi 18 genes (14 para GH19 e 4 para GH18), apresentando redução na classe IV e V; 39 genes (26 para GH19 e 13 para GH18), com adição de genes em todas as classes; e 18 genes (15 para GH19 e 3 para GH18), com apenas um representante da classe V, redução do número de quitinases de classe IV e ausência de quitinases de classe VII (Apêndice S).

## 5.1 Classe I

O gene que codifica a proteína de classe I que apresenta mutações em resíduos catalíticos em *A. suecica*, KAG7581678.1, provavelmente foi originado do gene CAE5967005.1 de *A. arenosa*, o qual se assemelha mais com a sequência de *A. suecica*, o que comprova a passagem de genes mediante o cruzamento entre *A. arenosa* e *A. thaliana*. É suposto que o mecanismo de defesa a ataques bióticos de *A. suecica* não seja avariado, uma vez que também foi herdada a sequência codificadora da quitinase de classe I de *A. thaliana* AT3G12500. O gene equivalente a essa sequência é KAG7630970.1\_A.suecica\_I, o qual não possui deleções nos sítios de catálise e ligação com o substrato.

O fato de que as sequências CAL9220541.1 de *A. halleri* e XP\_006407328.1 de *E. salsugineum* apresentarem deleções em regiões catalíticas importantes, ao passo que os testes de substituição sinônima apresentaram altas taxas nessas sequências pode significar que essas mutações detectadas nos genes dessas duas espécies podem não ter ação deletéria no mecanismo de catálise das enzimas codificadas.

Além disso, a região de ligação com o substrato do CatD (Figura 1) está sob pressão purificadora, o que infere um mecanismo de correção de mutações para essa região para essa classe de enzimas, indicando sua importância para a atividade biológica, em consonância com Ohuma et al. (Ohnuma et al., 2014).

## 5.2 Classe II

Sequências dessa classe se distribuem internamente em outros dois grupos próximos entre si, sendo um deles mais semelhante com a sequência de *A. thaliana* AT4G01700\_II e outro com AT1G02360.3\_II, sugerindo uma duplicação ancestral no gene que codifica essas proteínas em todas as espécies estudadas. Hipótese sustentada na sugestão de Blanc et al. (Blanc et al., 2000) de que *Arabidopsis thaliana* é um paleopoliploide fracionado.

Dada a variação em dois grupos internos, é esperado que quitinases de classe II apresentem maior variação nas taxas de substituição. Como ocorrido com as duas sequências de *A. suecica* e *A. halleri*, além da pressão seletiva diversificadora nos sítios da porção N-terminal do CatD (Figura 4). Essa classe de quitinase, ainda, quando analisada no genoma de *A. suecica*, sofre uma deleção de uma sequência, quando comparada com a soma teórica de genes dos dois genomas parentais, *Arabidopsis thaliana* e *Arabidopsis arenosa*. Os genes mantidos foram mais semelhantes a AT4G01700\_II, sugerindo maior conservação na função de genes próximos em sequência primária a ele.

### 5.3 Classe III

Membros da classe III apresentaram alto grau de conservação das sequências. Mediante a lógica descrita correlacionando número de quitinases com sua atividade (Grover, 2012; Passarinho; De Vries, 2002), quitinases de classe III provavelmente possuem ação fundamental na defesa vegetal. As taxas de substituição também mantiveram o padrão de seleção purificadora para genes dessa classe, o que indica atuação no mecanismo molecular a fim de manter a estrutura primária das enzimas, e, com isso, sua ação fisiológica.

Ademais, a sequência CAE6117724\_A.arenosa\_III apresentou domínios adicionais não relatados em quitinases, usualmente (Anexo A). Os domínios de B\_lectina, S\_locus\_glycop, PAN\_APPLE e Pkc\_like podem ser originado de um evento de *splicing* alternativo, que não foi mantido em outras proteínas dessa classe para outras espécies.

### 5.4 Classe IV e classe V

Classes com maior número de representantes e com genes mais correlacionados com metabolismo geral vegetal e relatados em menor frequência contra ataques bióticos. Proteínas dessa classe estão localizadas em tandem nos cromossomos das três espécies *Arabidopsis thaliana*, *Arabidopsis arenosa* e no alotetraploide *Arabidopsis suecica* (Figuras 10)

(Apêndices Q e R), sugerindo um ou mais eventos de duplicação desses genes. Adicionalmente, a presença de três proteínas com duplicação de domínios ou domínios atípicos para classe IV (Anexo B) é concordante com o fato de que sequências dessa classe estão sob constante pressão de seleção diversificadora, como evidenciado pelos testes aBSREL e FEL (Figura 6).

O número de quitinases de classe IV e de classe V é também o mais variado dentre as espécies estudadas (Apêndice S). Como exemplo, em *Eutrema salsugineum* há apenas um representante da classe V, o que pode indicar uma perda da atividade no metabolismo celular geral dessas proteínas nessa espécie, mediante o estresse salino, e sua ação tenha sido suprimida ou substituída por quitinases de outras classes, principalmente de classe IV, que teve o número de genes acrescido em *Eutrema salsugineum*, quando comparado com as espécies *Arabidopsis thaliana*, *Arabidopsis arenosa* e *Arabidopsis lyrata* (Apêndice S). Além disso, o número de genes de classe IV em *Arabidopsis thaliana*, quando comparado com o número de genes de classe V, é similar, padrão não encontrado para as outras espécies, salvo o alotetraploide *Arabidopsis suecica*. Isso pode indicar que *A. arenosa*, *A. lyrata* podem ter sofrido o mesmo efeito de substituição de função de quitinases de classe V por membros da classe IV, como especulado para *Eutrema salsugineum*. Esse fenômeno demonstra que essas espécies, por possuírem maiores mecanismos de adaptação a estresses bióticos e abióticos quando comparadas com *Arabidopsis thaliana* (Bouزيد et al., 2019; Kaplenig et al., 2022) podem possuir a mesma lógica de distribuição do balanço quitinases de classe IV e de classe V encontrado em *Eutrema salsugineum* (Marín-de la Rosa et al., 2019).

Nota-se, também, que a presença ou ausência do ChtBD não induz uma resposta mais específica a ataques bióticos, uma vez que quitinases de classe III não possuem ChtBD e são relatadas nessas ocasiões, enquanto quitinases de classe IV e de classe VI possuem esse domínio de ligação mas não são fortemente expressas contra agentes bióticos quanto classe I e classe III (Chen et al., 2025; Hamid et al., 2013).

## 5.5 Classe VI

Quitinases dessa classe são relatadas como com fraca ação sobre o substrato, uma vez que a deleção do Glu (E) canônico para a catálise é relatada em todos os representantes dessa classe (Tyler et al., 2010). Além disso, seu ChtBD com as deleções citadas podem produzir um domínio com *folding* incorreto.

Ao passo que sua função catalítica pode não estar relacionada diretamente com estresse biótico, as taxas de substituição sinônima foram consideravelmente altas em comparação com não sinônimas (Figura 8), indicando, provavelmente que sua função no metabolismo geral é conservada. Argumento reforçado pelo número intermediário de cópias desses genes nos genomas, sendo maior que para a classe II, por exemplo, nas plantas estudadas.

## 5.6 Classe VII

Kolosova et al. (Kolossova; Breuil; Bohlmann, 2014) consideraram proteínas dessa classe como da família GH18. No entanto, em nível de sequência primária, os resíduos catalíticos e de ligação dos exemplares da classe VII para os genomas estudados nesse trabalho estão de acordo com o citado (Kolossova; Breuil; Bohlmann, 2014).

Novamente, a serina (S/Ser) catalítica sofreu uma mutação em todas as proteínas da classe VII para alanina (A/Ala) (Apêndice P), o que pode indicar que o padrão do resíduo canônico unicamente se manteve em classe I e classe II, com poucas ou nenhuma proteína apresentando a serina canônica. Com isso, pode ser considerado um resíduo não crucial para a catálise em outras classes (Peery et al., 2021). Mais informações sobre os resíduos participantes da catálise ou ligação com o substrato não foram encontrados no banco de dados CDD para essa classe.

O fato de AT3G47540.2 estar sob pressão seletiva episódica diversificadora mas não possuir codons sobre pressão diversificadora mediante o teste FEL pode ser esperado. O teste aBSREL pode detectar pulsos episódicos de mutações em uma linhagem evolutiva. O teste FEL, por sua vez, infere taxas de substituição de forma fixa, unicamente com base nas correlações de homologia e identidade dos aminoácidos codificados por cada trinca do conjunto de CDS's disponibilizadas e alinhadas com as sequências polipeptídicas correspondentes (Weaver et al., 2018). Dessa forma, AT3G47540.2 provavelmente sofreu mutações não sinônimas, de tal sorte que gerou um pulso evolutivo episódico, mas de forma fixa as mutações não permaneceram, como sugerem os resultados do FEL (Figura 9).

## 5.7 *Eutrema salsugineum*

A quantidade de genes que codificam quitinases nessa espécie pode ser um objeto comparativo de distância evolutiva média, em comparação com as outras espécies estudadas,

dada sua não participação do gênero *Arabidopsis*. Além disso, entender a dinâmica de quantidade de quitinases e correlacioná-la ao fato de *Eutrema salsugineum* ser uma halófito relativa modelo pode ajudar a compreender qual classe dessas enzimas pode estar mais envolvida com o mecanismo de estresse abiótico.

De posse disso, para quitinases de classe I e VI, *Eutrema salsugineum* manteve o padrão observado na planta modelo *Arabidopsis thaliana*. Para quitinases de classe III houve a adição de mais um gene, o que pode sugerir uma maior resposta a ataques bióticos mediante o estresse salino (Khanal; Moffatt; Gray, 2015). Em classe IV a adição de dois genes pode sugerir também uma maior amplitude nas respostas a influências diversas no metabolismo, o que justificaria a duplicação na região cromossômica equivalente que codifica proteínas de classe IV. As reduções nas classes V e VII sugerem uma perda de função dos genes em *Eutrema salsugineum*, ao passo que as funções dessas classes podem ter sido substituídas, para classe V, pela classe IV e para classe VII, pelas classes I e II, seguindo um raciocínio de proximidade de função por proximidade de número de cópias desses genes.

## 6 CONCLUSÃO

*Arabidopsis thaliana* é um modelo coerente para avaliar quantidade de quitinases e sua distribuição e classificação de domínios em plantas. *Arabidopsis halleri*, *Arabidopsis lyrata* e *Eutrema salsugineum* apresentam números menores de quitinases em relação ao modelo, provavelmente por perda de função de quitinases de classe V e de classe IV. *Arabidopsis suecica* demonstra as consequências do cruzamento entre *Arabidopsis arenosa* e *Arabidopsis thaliana* ao evidenciar a soma entre maior parte das sequências que codificam quitinases da geração parental em seu genoma alotetraploide.

Quitinases dos genomas vegetais estudados, com exceção para classe IV e V, possuem fortes taxas de seleção purificadora, condizente com a função central bem relatada contra estresses bióticos enfrentados pelo vegetal. Ao passo que quitinases de classe IV e V apresentam taxas de substituição não sinônimas em maior número, indicando que quitinases dessa classe estão em maior pressão diversificadora, o que se alinha com o raciocínio de participação majoritária em funções diversas do metabolismo vegetal, como na germinação, senescência, adaptação a estresses abióticos, e desenvolvimento da planta.

## REFERÊNCIAS

- ALMAGRO ARMENTEROS, Jose Juan et al. DeepLoc: prediction of protein subcellular localization using deep learning. **Bioinformatics**, v. 33, n. 24, p. 4049, 15 dez. 2017.
- BOLSER, Dan et al. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. **Methods in Molecular Biology**, v. 1374, p. 115–140, 2016.
- BOUZID, W. et al. *Arabidopsis* species deploy distinct strategies to cope with drought stress. **Journal of Experimental Botany**, v. 70, n. 4, p. 1375–1389, fev. 2019.
- COLLINGE, D. B. et al. Plant chitinases. **The Plant Journal**, v. 3, n. 1, p. 31–40, jan. 1993.
- CRINI, Grégorio. Historical review on chitin and chitosan biopolymers. **Environmental Chemistry Letters**, v. 17, n. 4, p. 1623–1643, 1 dez. 2019.
- GOODSTEIN, David M. et al. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178–1186, jan. 2012.
- HERMANS, Christian et al. Chitinase-like protein CTL1 plays a role in altering root system architecture in response to multiple environmental conditions. **Plant Physiology**, v. 152, n. 2, p. 904–917, fev. 2010.
- KAPLENIG, D. et al. Repeated colonization of alpine habitats by *Arabidopsis arenosa* viewed through freezing resistance and ice management strategies. **Plant Biology**, v. 24, n. 6, p. 939–949, set. 2022.
- KATOH, Kazutaka; ROZEWICKI, John; YAMADA, Kazunori D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1160–1166, 19 jul. 2019.
- KHANAL, N.; MOFFATT, B. A.; GRAY, G. R. Acquisition of freezing tolerance in *Arabidopsis* and two contrasting ecotypes of the extremophile *Eutrema salsugineum* (*Thellungiella salsuginea*). **Journal of Plant Physiology**, v. 180, p. 35–44, 15 maio 2015.
- KOLOSOVA, N.; BREUIL, C.; BOHLMANN, J. Cloning and characterization of chitinases from interior spruce and lodgepole pine. **Phytochemistry**, v. 101, p. 32–39, maio 2014.
- KUZNETSOV, N. I. et al. **Flora Caucasica critica**: Materialy dlia flory Kavkaza: kriticheskoe sistematicheskoe-geograficheskoe izslidovanie. IUrev: Tip. K. Mattisena, 1901. Disponível em: <https://www.biodiversitylibrary.org/bibliography/9632>.
- LANDIM, Patrícia G. Castro et al. Production in *Pichia pastoris*, antifungal activity and crystal structure of a class I chitinase from cowpea (*Vigna unguiculata*): Insights into sugar binding mode and hydrolytic action. **Biochimie**, v. 135, p. 89–103, abr. 2017.
- LV, Peiyun et al. Genome-Wide Identification and Expression Analyses of the Chitinase Gene Family in Response to White Mold and Drought Stress in Soybean (*Glycine max*). **Life**, v. 12, n. 9, p. 1340, 29 ago. 2022.
- MARCHLER-BAUER, Aron et al. CDD: NCBI's conserved domain database. **Nucleic Acids Research**, v. 43, n. D1, p. D222–226, jan. 2015.

- MARGIS-PINHEIRO, M. et al. Isolation of a complementary DNA encoding the bean PR4 chitinase: an acidic enzyme with an amino-terminus cysteine-rich domain. **Plant Molecular Biology**, v. 17, n. 2, p. 243–253, ago. 1991.
- MARÍN-DE LA ROSA, N. et al. Genome-wide identification of drought-responsive genes in *Arabidopsis lyrata* and *Eutrema salsugineum*. **Frontiers in Plant Science**, v. 10, p. 1116, 18 set. 2019.
- MINH, Bui Quang et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. **Molecular Biology and Evolution**, v. 37, n. 5, p. 1530–1534, 1 maio 2020.
- MIR, Zahoor Ahmad et al. Genome-wide identification and characterization of Chitinase gene family in *Brassica juncea* and *Camelina sativa* in response to *Alternaria brassicae*. **Genomics**, v. 112, n. 1, p. 749–763, jan. 2020.
- MOTLAGH, N. M. et al. Enhancing enzyme activity of heterologously expressed Chit42 in *Nicotiana benthamiana* for chitin degradation. **Biotechnology Reports**, v. 48, p. e00934, dez. 2025.
- MOUSSIAN, Bernard. Chitin: Structure, Chemistry and Biology. **Advances in Experimental Medicine and Biology**, v. 1142, p. 5–18, 2019.
- NIELSEN, Henrik et al. SignalP: The Evolution of a Web Server. **Methods in Molecular Biology**, v. 2836, p. 331–367, 2024.
- OLIVEIRA, Simone T. et al. Structural and functional features of a class VI chitinase from cashew (*Anacardium occidentale* L.) with antifungal properties. **Phytochemistry**, v. 189, p. 112814, set. 2021.
- PASSARINHO, Paul A.; DE VRIES, Sacco C. Arabidopsis Chitinases: a Genomic Survey. **The Arabidopsis Book**, v. 1, p. e0023, 2002.
- PEERY, Rhiannon M. et al. Comparative genomics of the chitinase gene family in lodgepole and jack pines: contrasting responses to biotic threats and landscape level investigation of genetic differentiation. **Botany**, v. 99, n. 6, p. 355–378, jun. 2021.
- SAYERS, Eric W. et al. GenBank. **Nucleic Acids Research**, v. 48, n. D1, p. D84–D86, 8 jan. 2020.
- SHINSHI, H. et al. Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. **Plant Molecular Biology**, v. 14, n. 3, p. 357–368, mar. 1990.
- STECHER, Glen et al. MEGA 12.1: Cross-Platform Release for macOS and Linux Operating Systems. **Journal of Molecular Evolution**, 17 nov. 2025.
- SU, Yachun et al. Identification, Phylogeny, and Transcript of Chitinase Family Genes in Sugarcane. **Scientific Reports**, v. 5, p. 10708, 2 jun. 2015.
- TAIRA, Toki et al. Cloning and characterization of a small family 19 chitinase from moss (*Bryum coronatum*). **Glycobiology**, v. 21, n. 5, p. 644–654, maio 2011.

THOMAS, Paul D. et al. PANTHER: Making genome-scale phylogenetics accessible to all. **Protein Science**, v. 31, n. 1, p. 8–22, jan. 2022.

TYLER, Ludmila et al. Annotation and comparative analysis of the glycoside hydrolase genes in *Brachypodium distachyon*. **BMC Genomics**, v. 11, p. 600, 25 out. 2010.

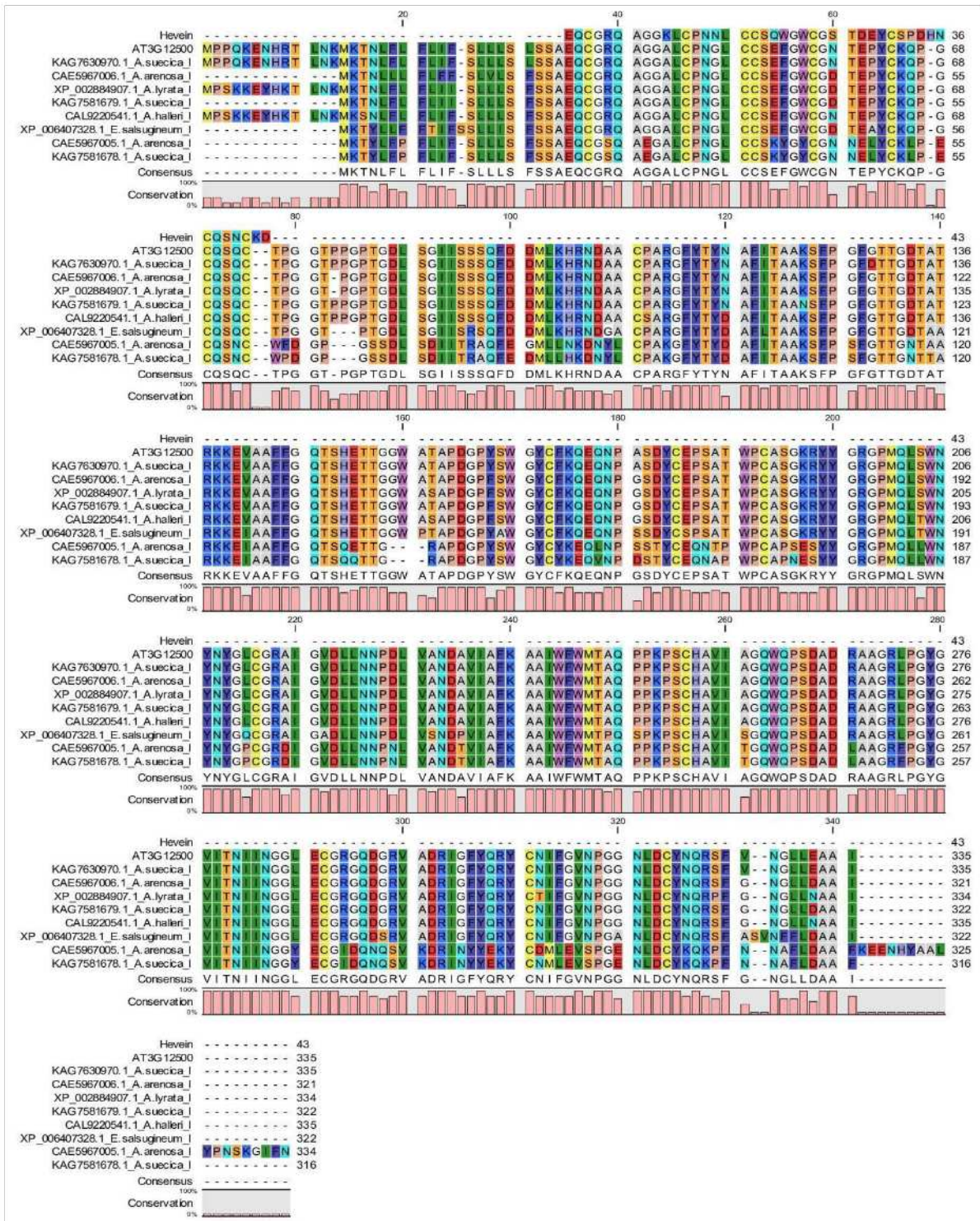
WARE, Doreen H. et al. Gramene, a tool for grass genomics. **Plant Physiology**, v. 130, n. 4, p. 1606–1613, dez. 2002.

YANG, L. et al. Characterization and Biocontrol Potential of a Novel Endochitinase CaChi19B from *Chitinilyticum aquatile*. **Journal of Agricultural and Food Chemistry**, v. 73, n. 32, p. 20019–20030, 13 ago. 2025.

YANG, R. et al. The Reference Genome of the Halophytic Plant *Eutrema salsugineum*. **Frontiers in Plant Science**, v. 4, p. 46, 2013.

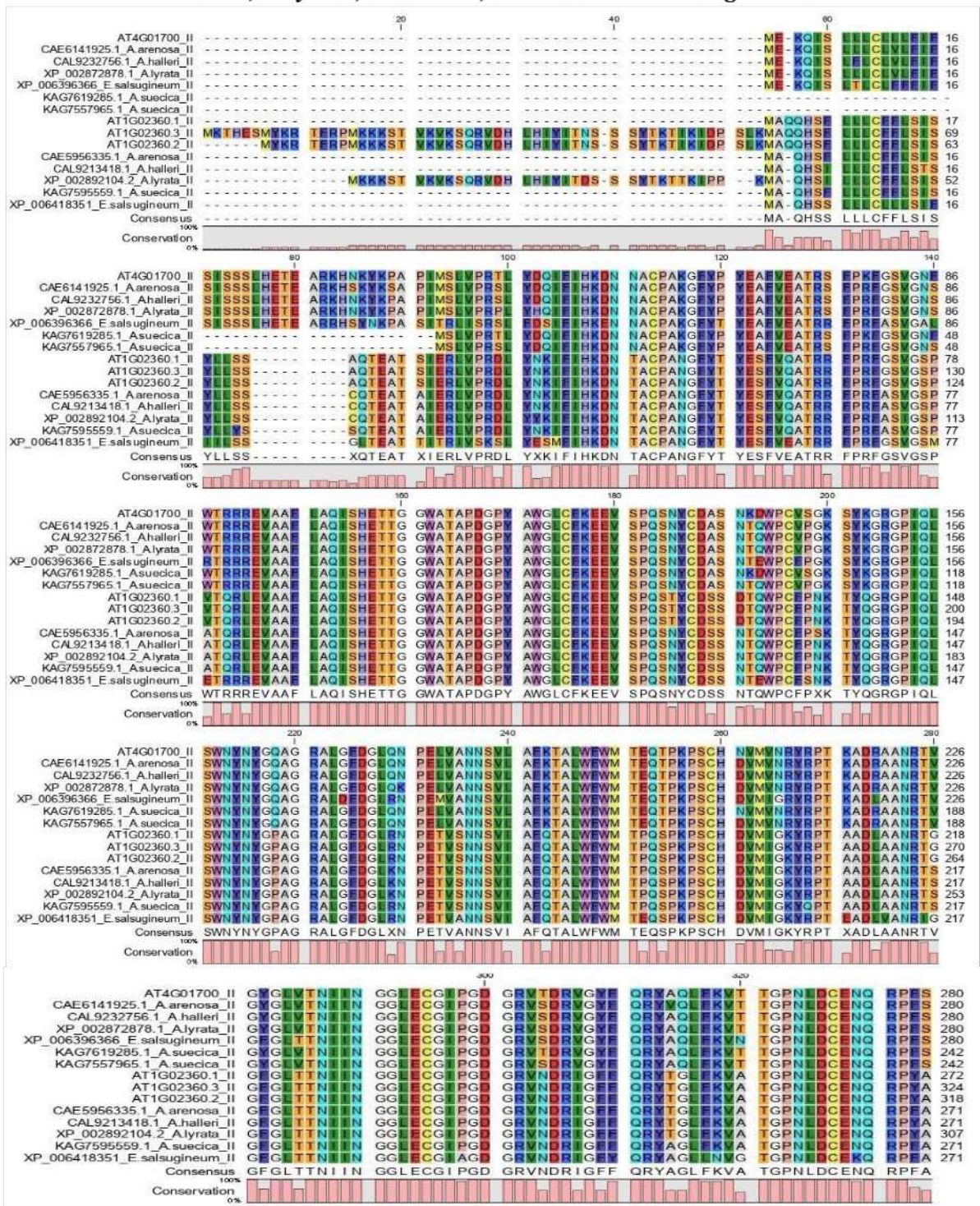
ZHONG, Ruiqin et al. Mutation of a chitinase-like gene causes ectopic deposition of lignin, aberrant cell shapes, and overproduction of ethylene. **The Plant Cell**, v. 14, n. 1, p. 165–179, jan. 2002.

## APÊNDICE A – ALINHAMENTO DO ChtBD DE HEVEÍNA DE *Hevea brasiliensis* COM QUITINASES DE CLASSE I DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. suecica*, *A. arenosa* E *E. salsugineum*.



Fonte: Autoral. Imagem gerada via CLC Sequence Viewer 8. Sequências de *A. thaliana* possuem apenas o identificador do transcrito do projeto TAIR10 (Swarbreck et al., 2008).

APÊNDICE B – ALINHAMENTO DE QUITINASES DE CLASSE II DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. suecica*, *A. arenosa* E *E. salsugineum*.

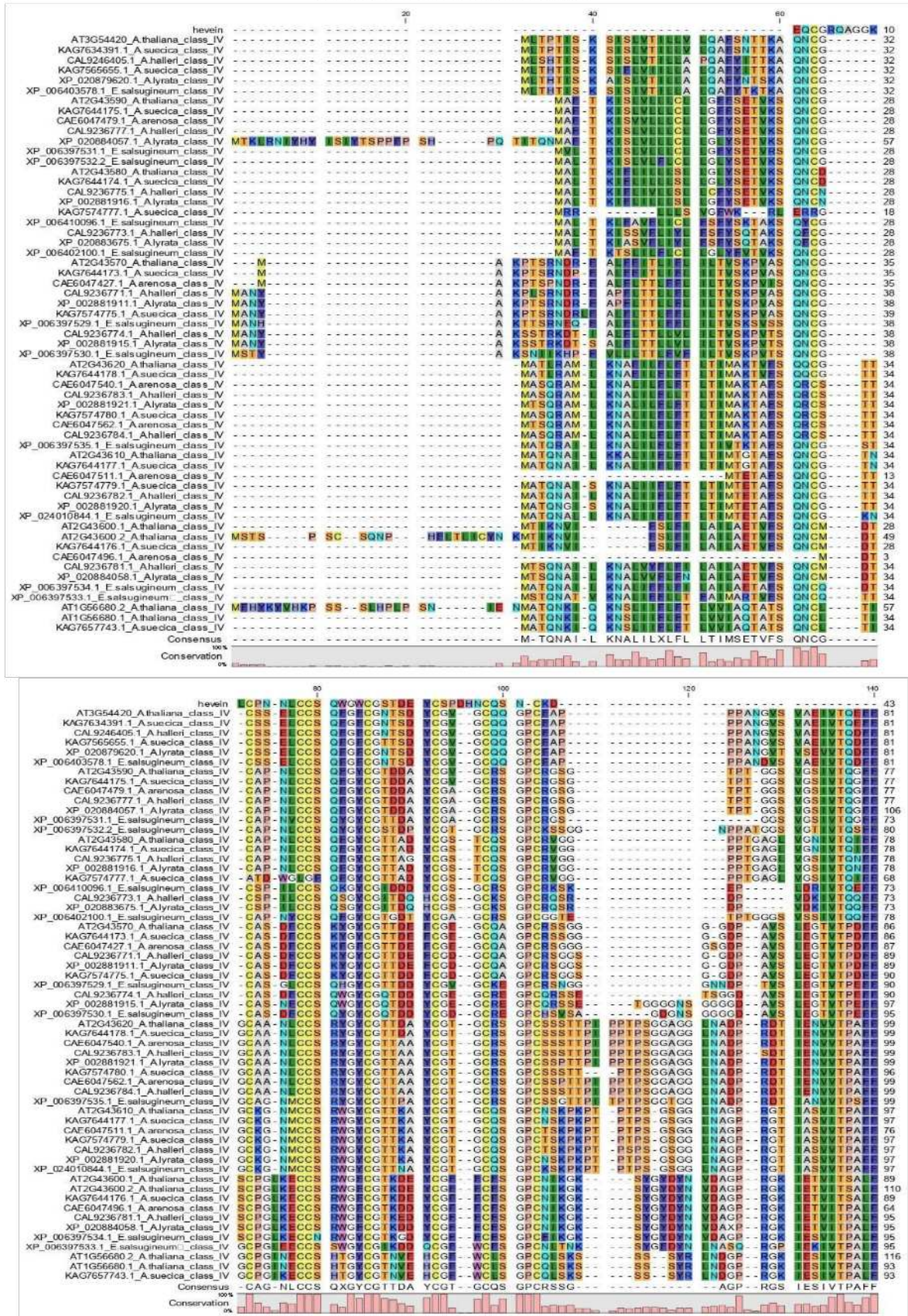


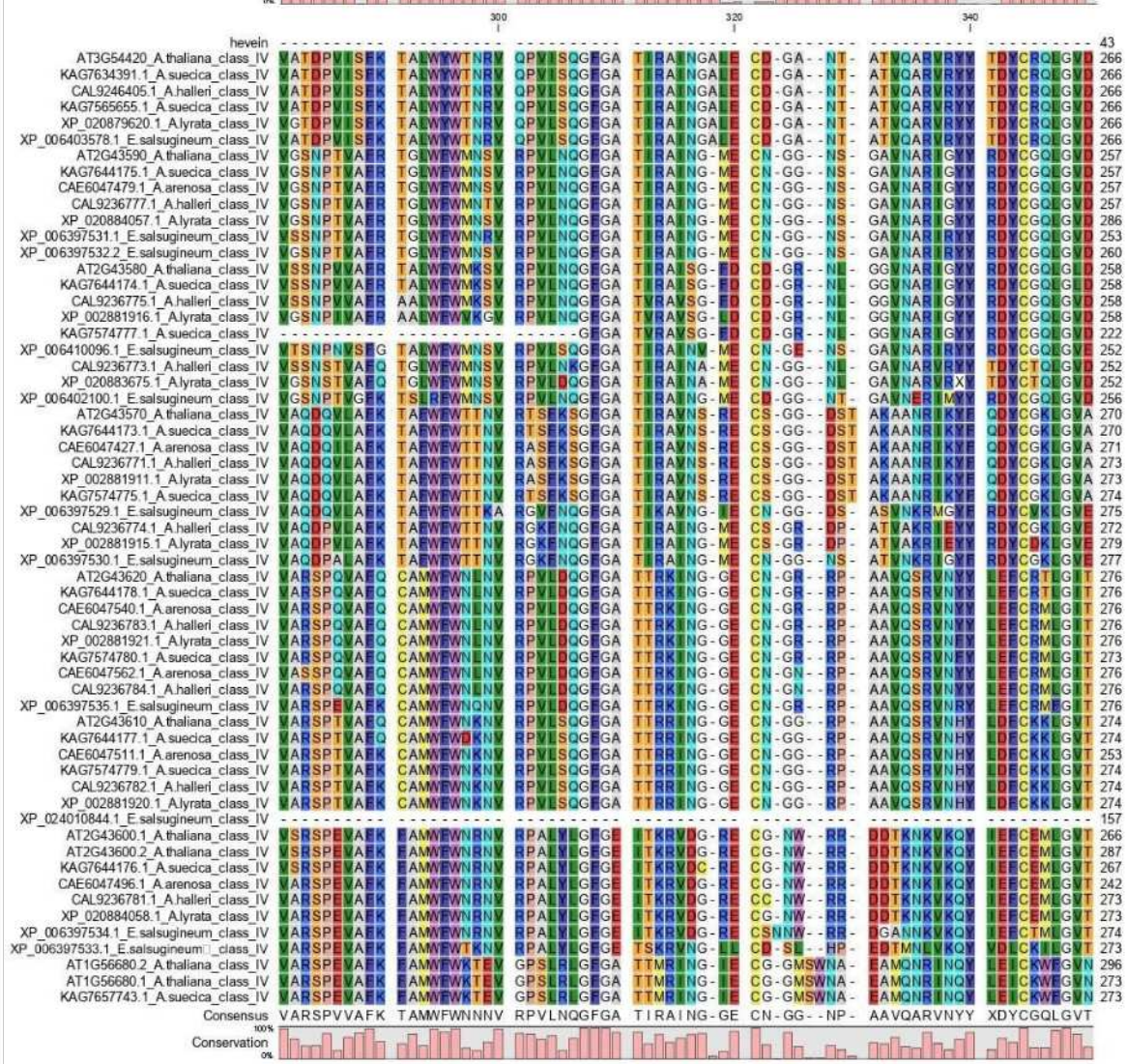
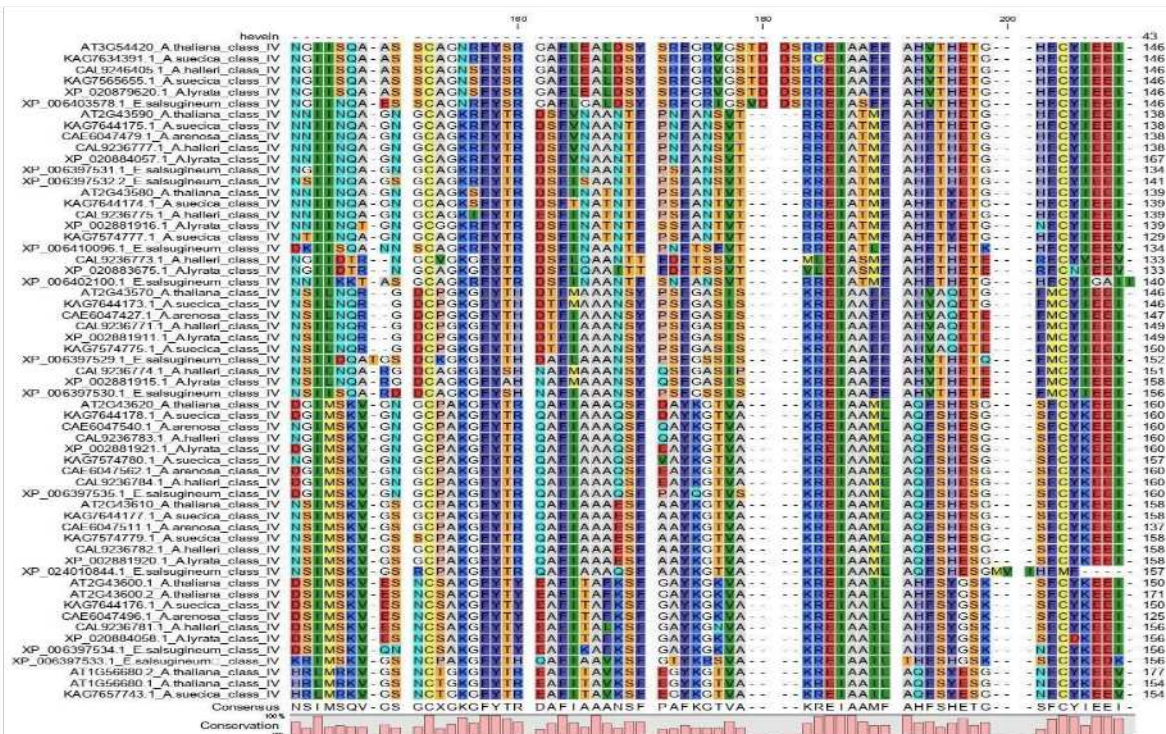
Fonte: Autoral. Imagem gerada via CLC Sequence Viewer 8. Sequências de *A. thaliana* possuem apenas o identificador do transcrito do projeto TAIR10 (Swarbreck et al., 2008).

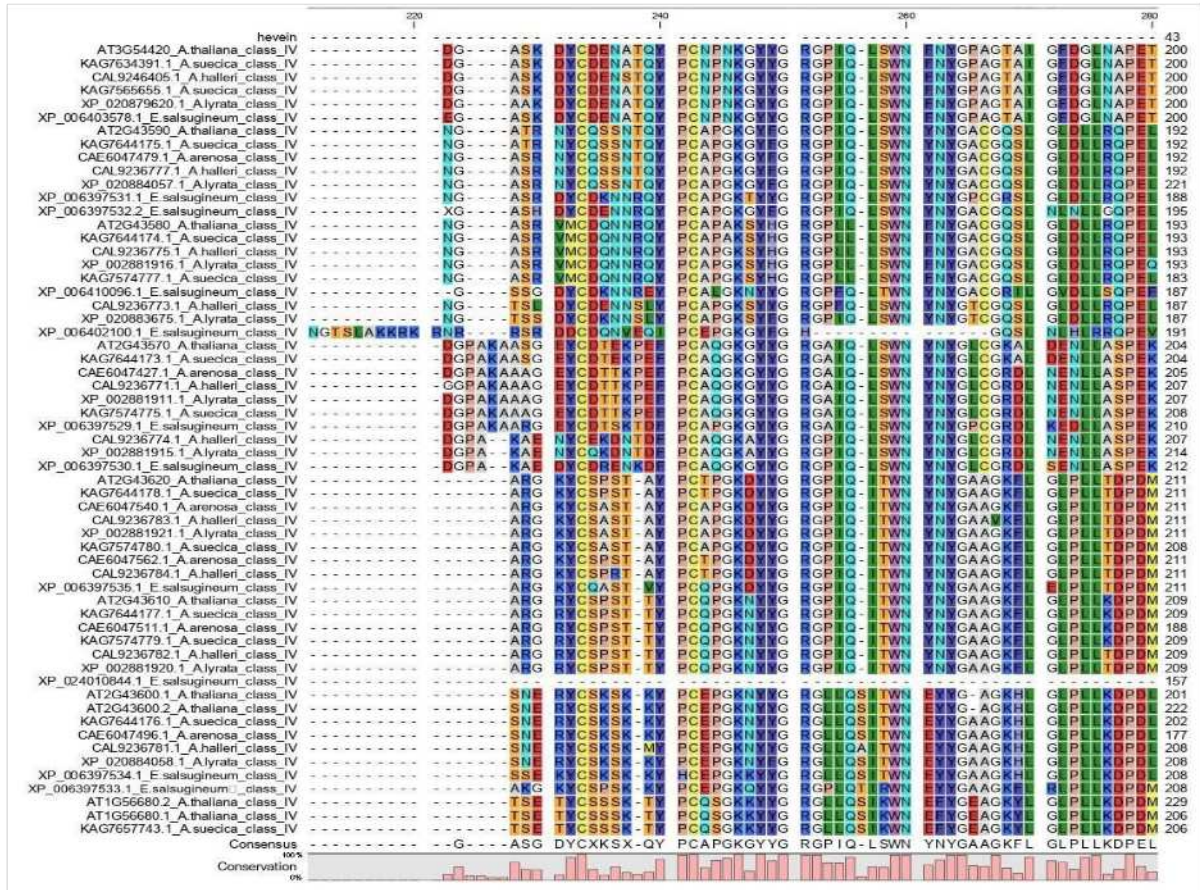




# APÊNDICE D – ALINHAMENTO DO ChtBD DE HEVEÍNA DE *Hevea brasiliensis* COM QUITINASES DE CLASSE IV DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. suecica*, *A. arenosa* E *E. salsugineum*.

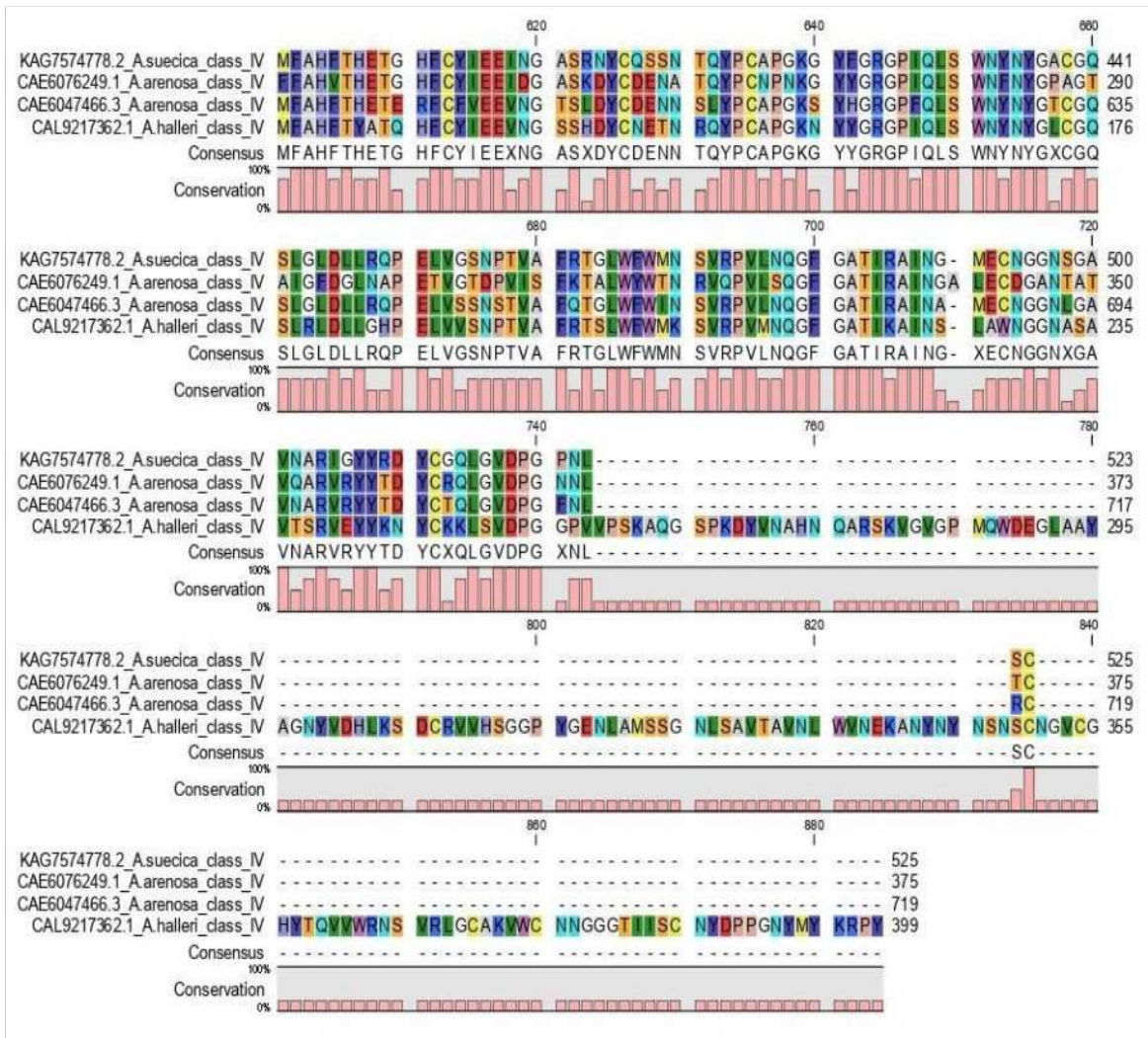






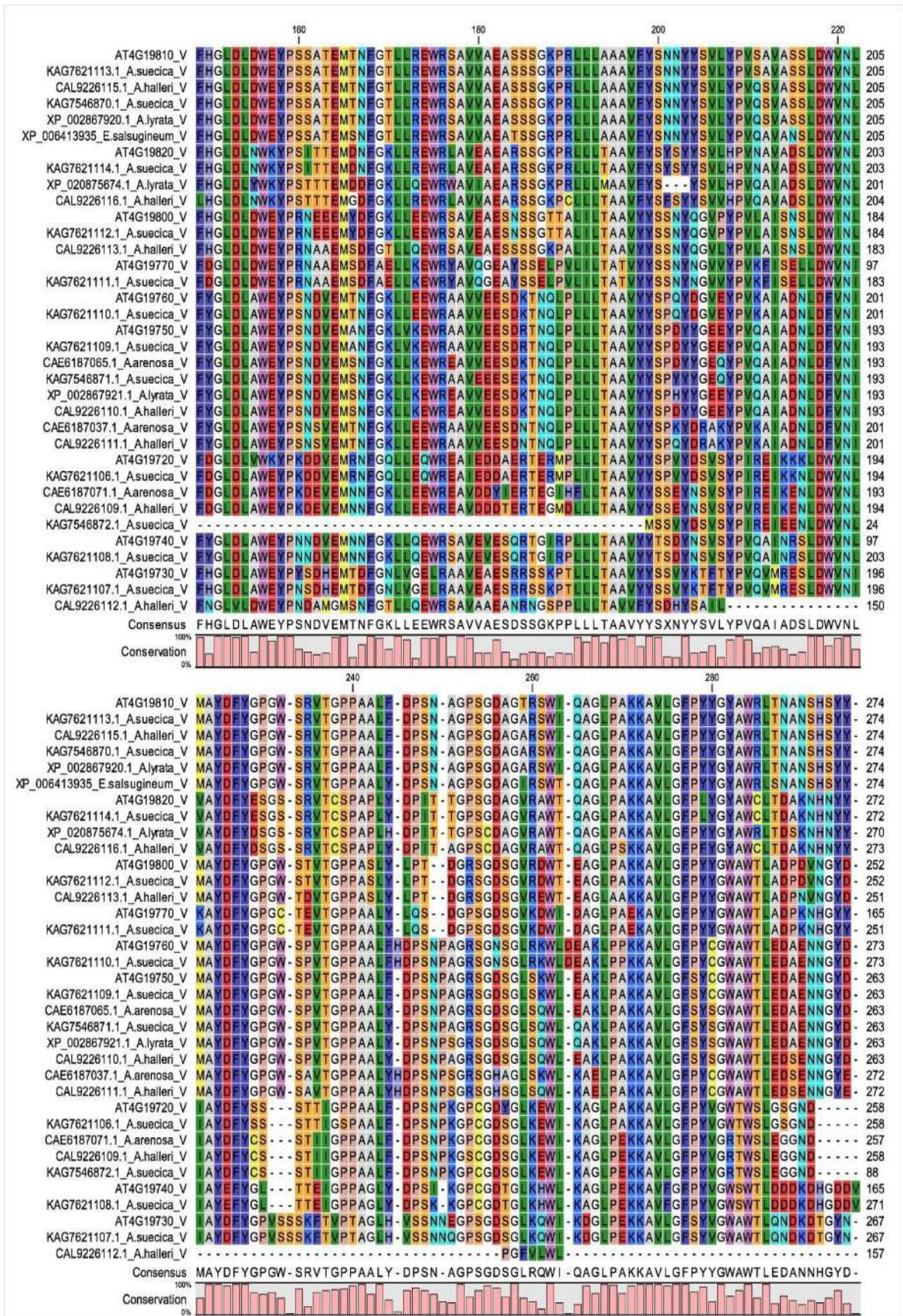
Fonte: Autoral. Imagem gerada via CLC Sequence Viewer 8.

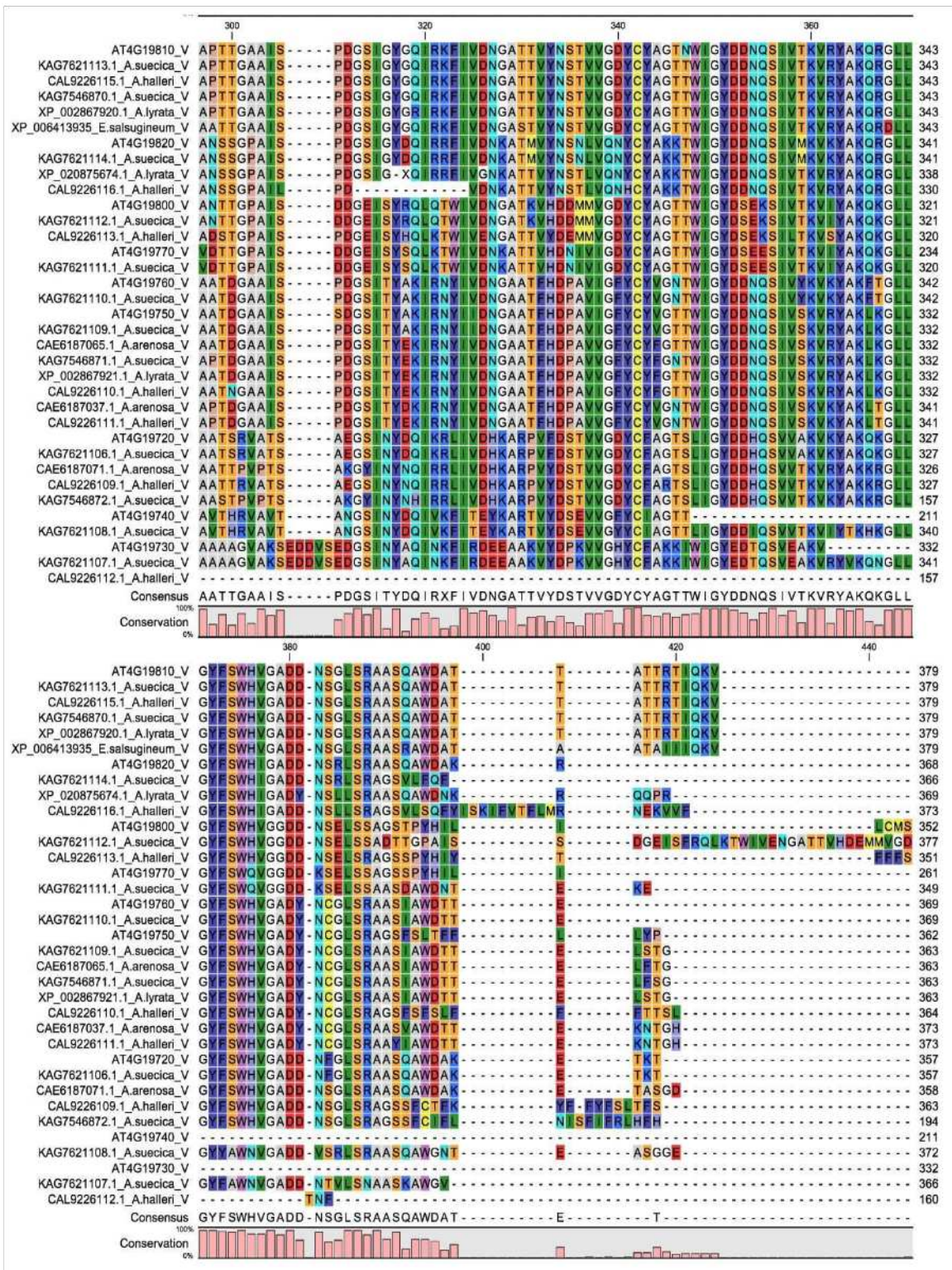




Fonte: Autor. Imagem gerada via CLC Sequence Viewer 8.

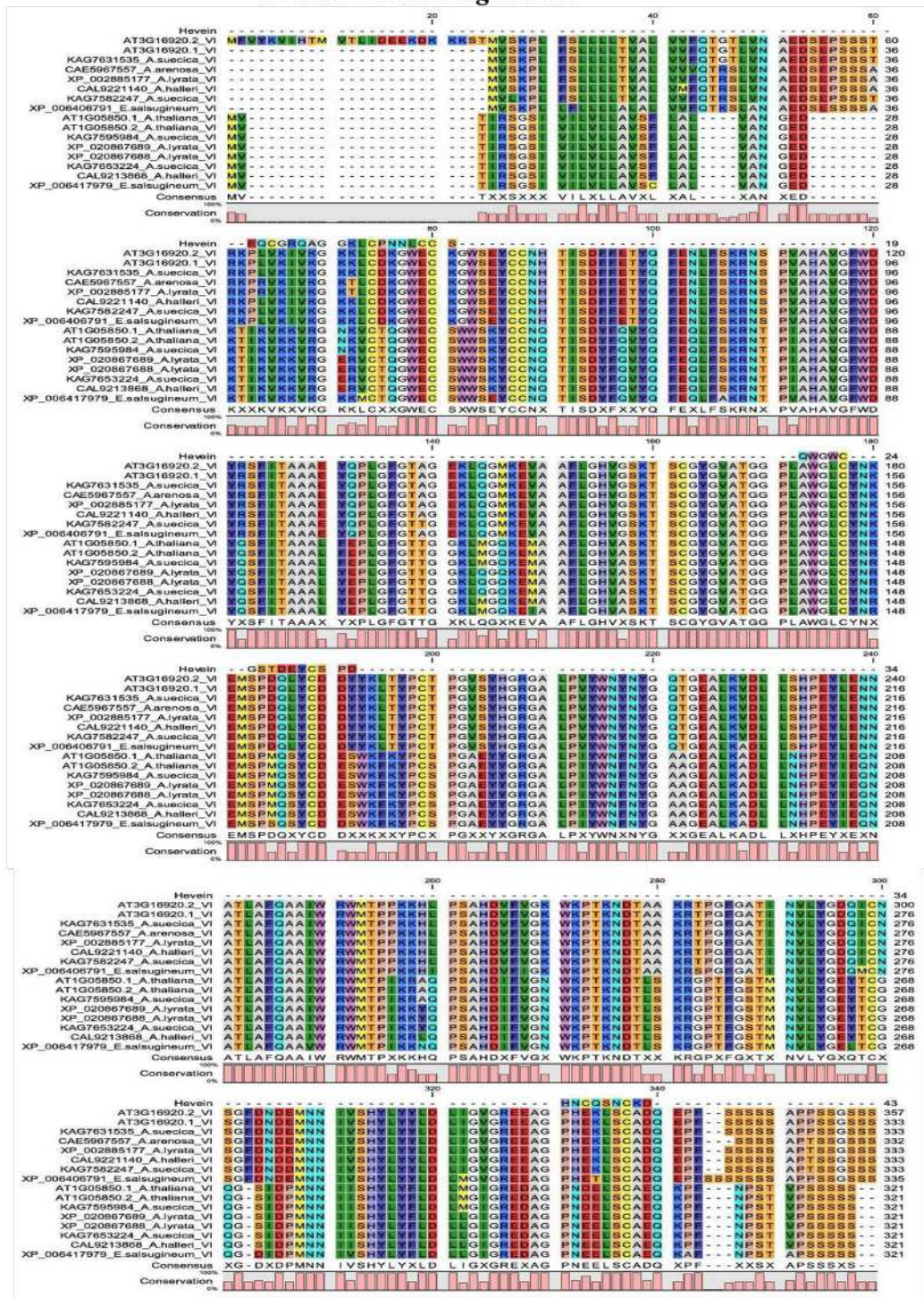






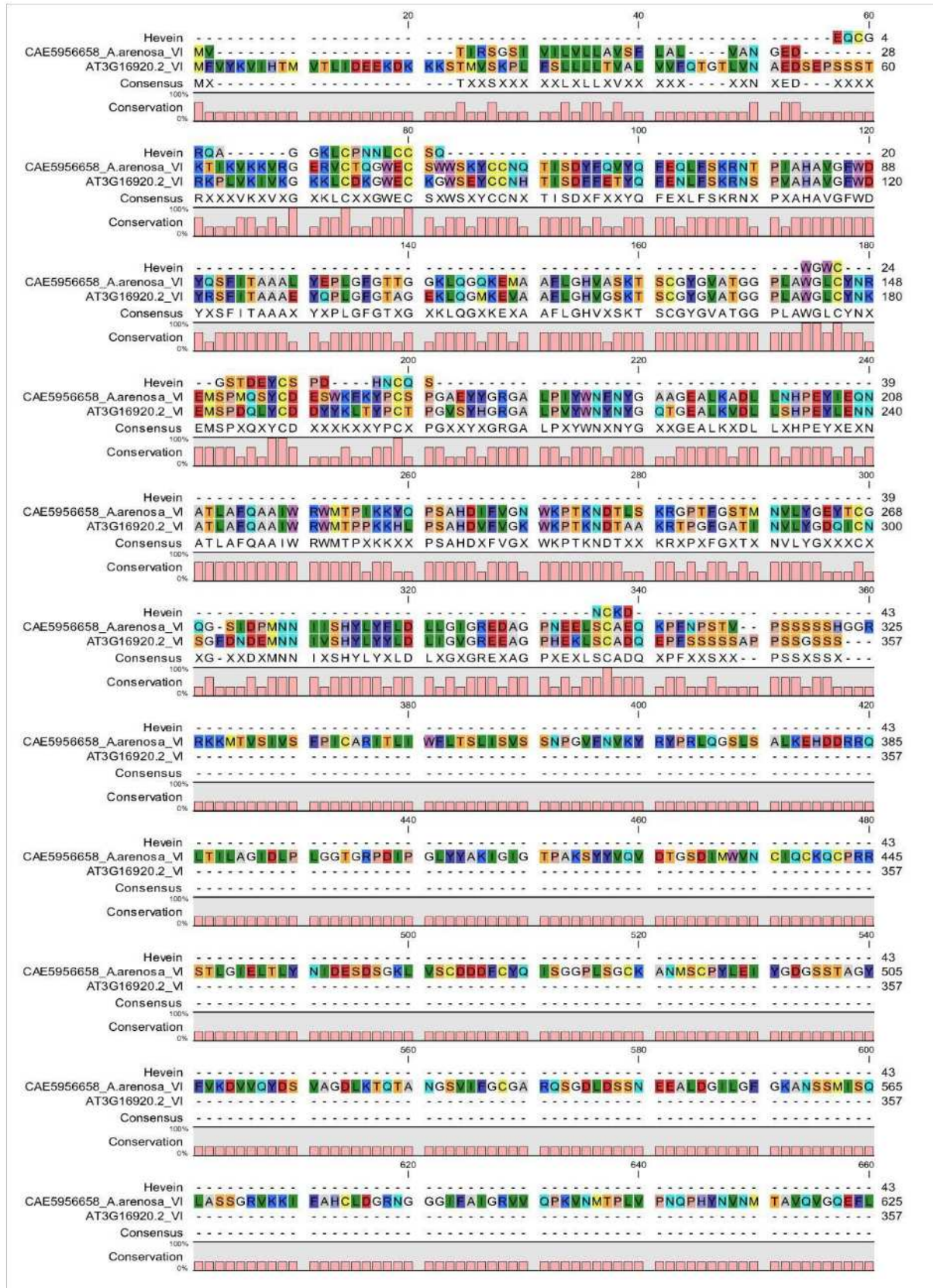


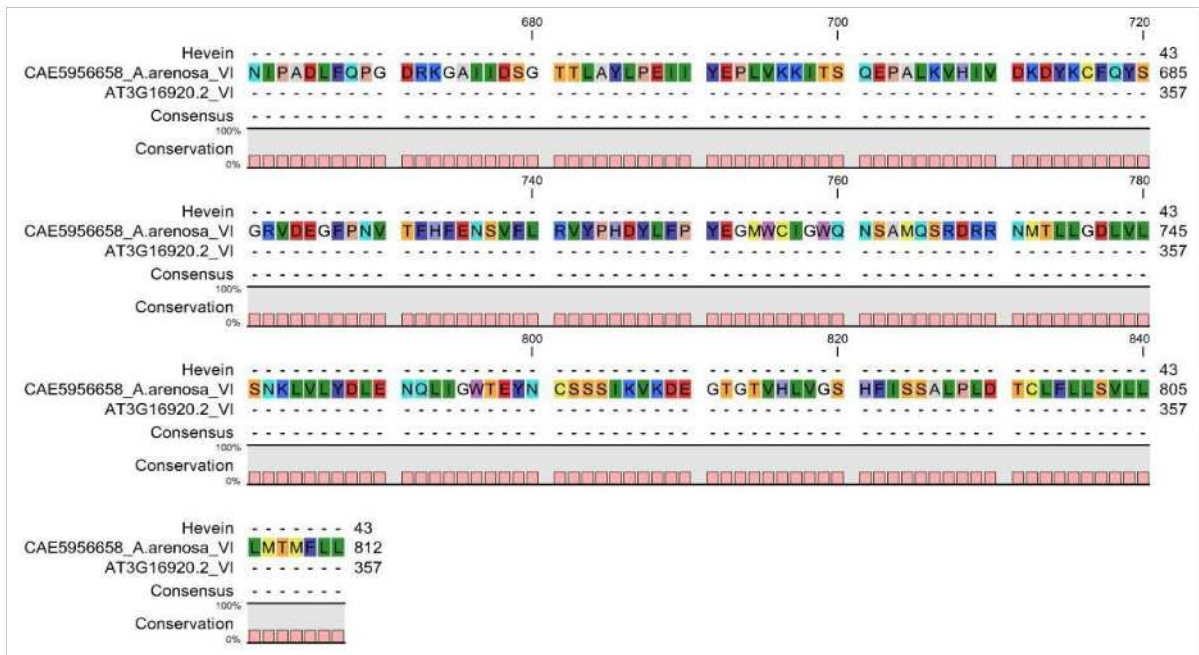
# APÊNDICE G – ALINHAMENTO DO ChtBD DE HEVEÍNA DE *Hevea brasiliensis* COM QUITINASES DE CLASSE VI DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. suecica*, *A. arenosa* E *E. salsugineum*.



Fonte: Autoral. Imagem gerada via CLC Sequence Viewer 8. Sequências de *A. thaliana* possuem apenas o identificador do transcrito do projeto TAIR10 (Swarbreck et al., 2008).

## APÊNDICE H – ALINHAMENTO DO ChtBD DE HEVEÍNA DE *Hevea brasiliensis* COM QUITINASES DE CLASSE VI COM DOMÍNIOS ADICIONAIS DE *A. thaliana* E *A. arenosa*.

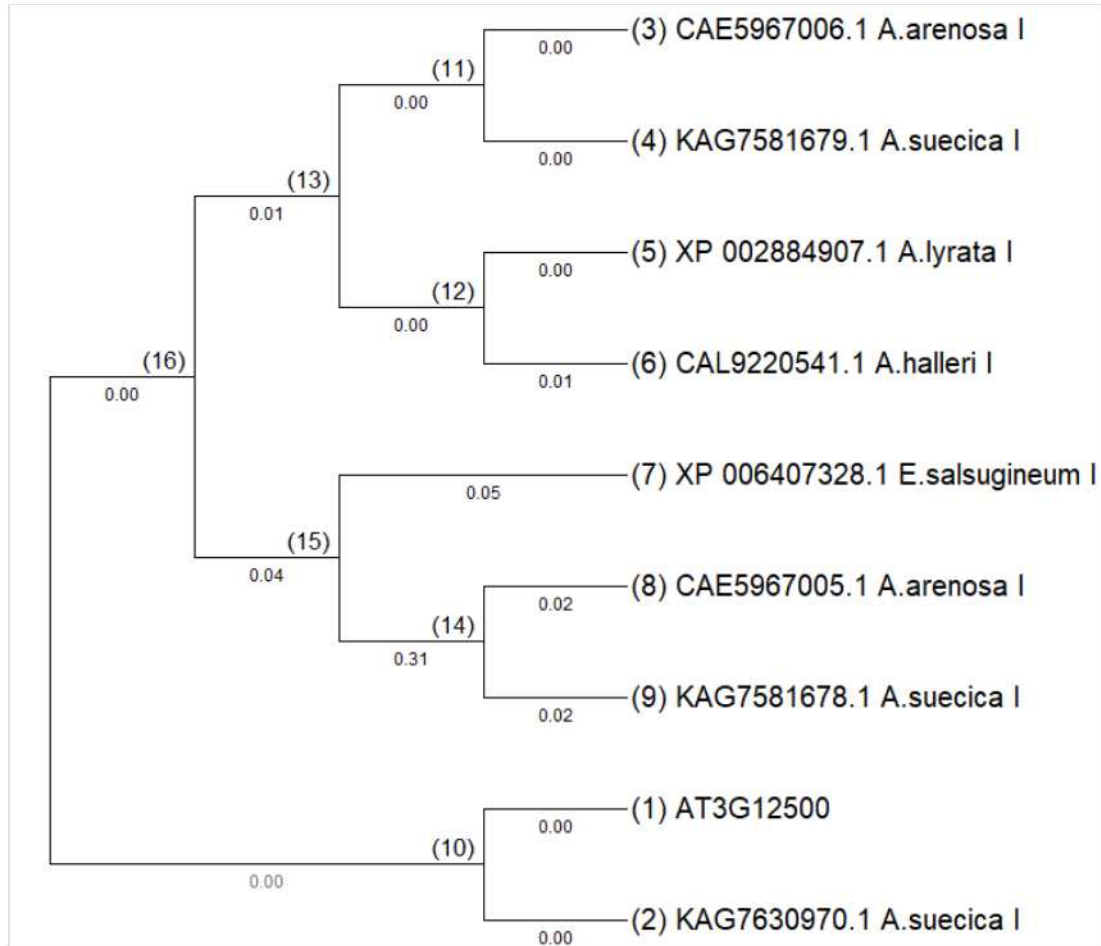




Fonte: Autoral. Imagem gerada via CLC Sequence Viewer 8. Sequências de *A. thaliana* possuem apenas o identificador do transcrito do projeto TAIR10 (Swarbreck et al., 2008) e foi utilizada de modelo comparativo nesse alinhamento.

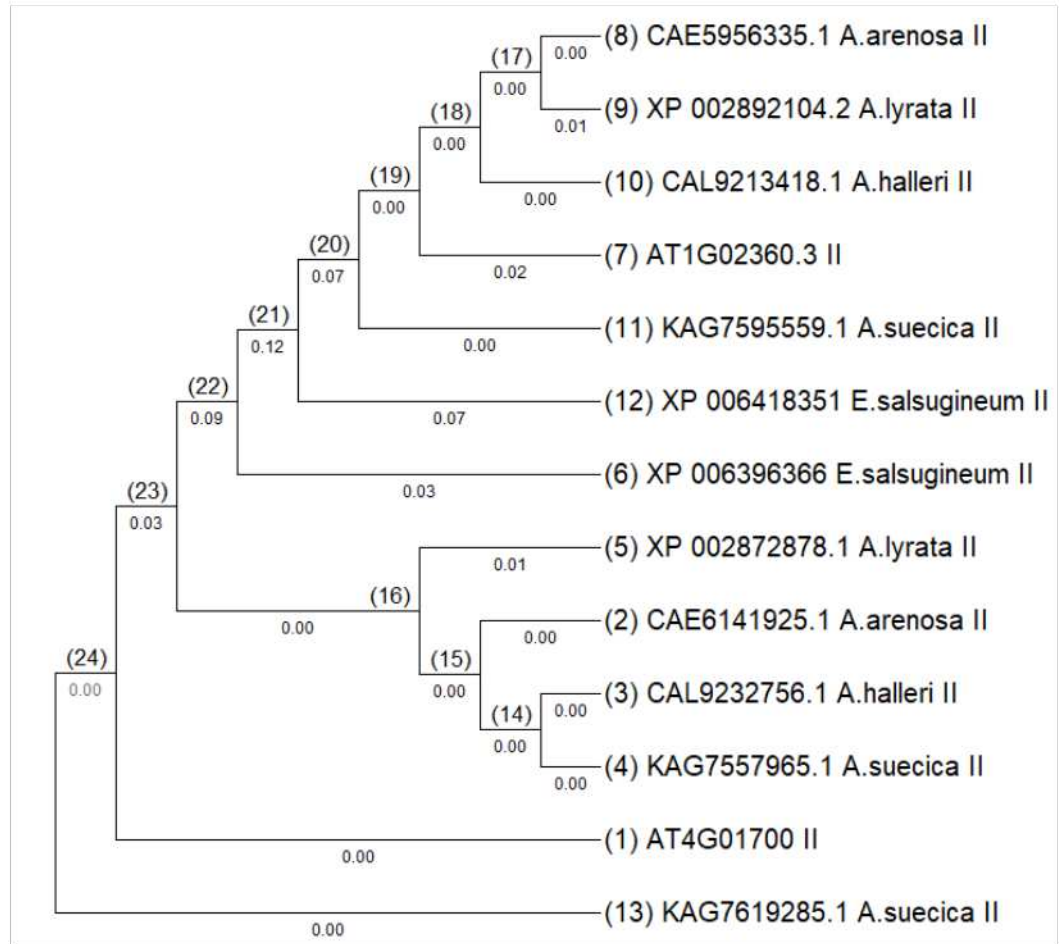


**APÊNDICE J – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE I DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. arenosa* E *E. salsugineum***



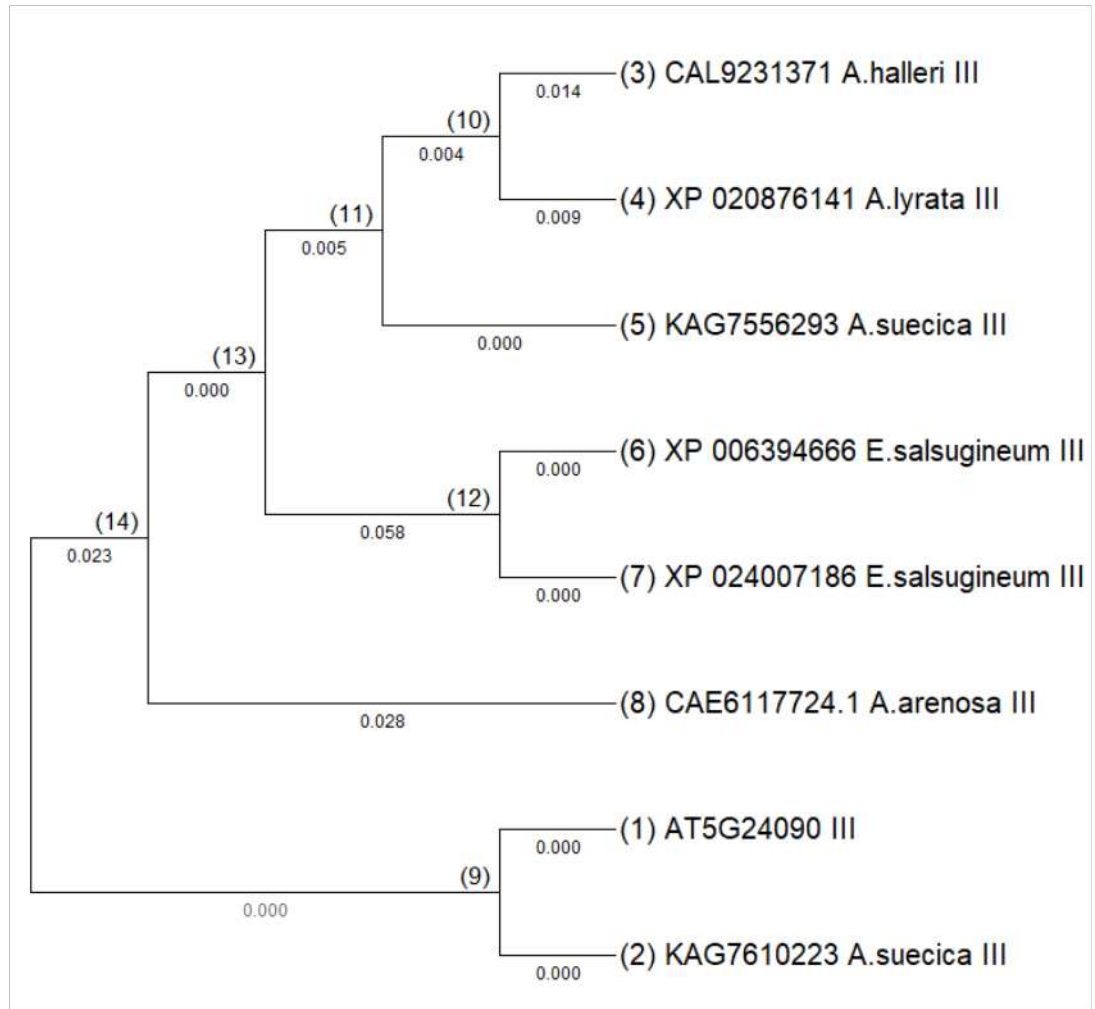
Fonte: Autoral. Gerada via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025). Cada número em parêntese representa o identificador nó. Árvore gerada somente com o CatD de cada sequência.

**APÊNDICE K – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE II DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. arenosa* E *E. salsugineum***



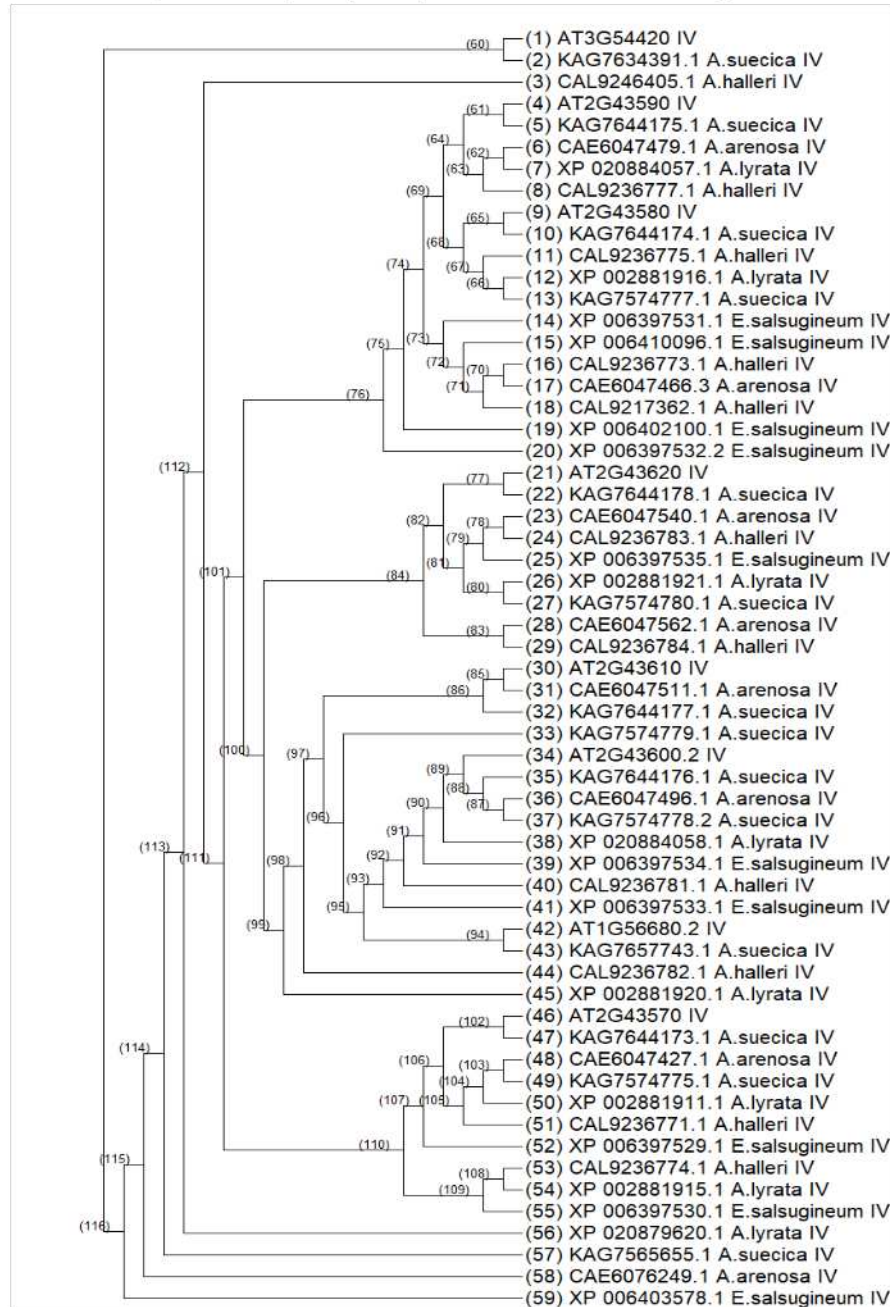
Fonte: Autoral. Gerada via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025). Cada número em parêntese representa o identificador nó. Árvore gerada somente com o CatD de cada sequência.

**APÊNDICE L – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE III DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. arenosa* E *E. salsugineum***



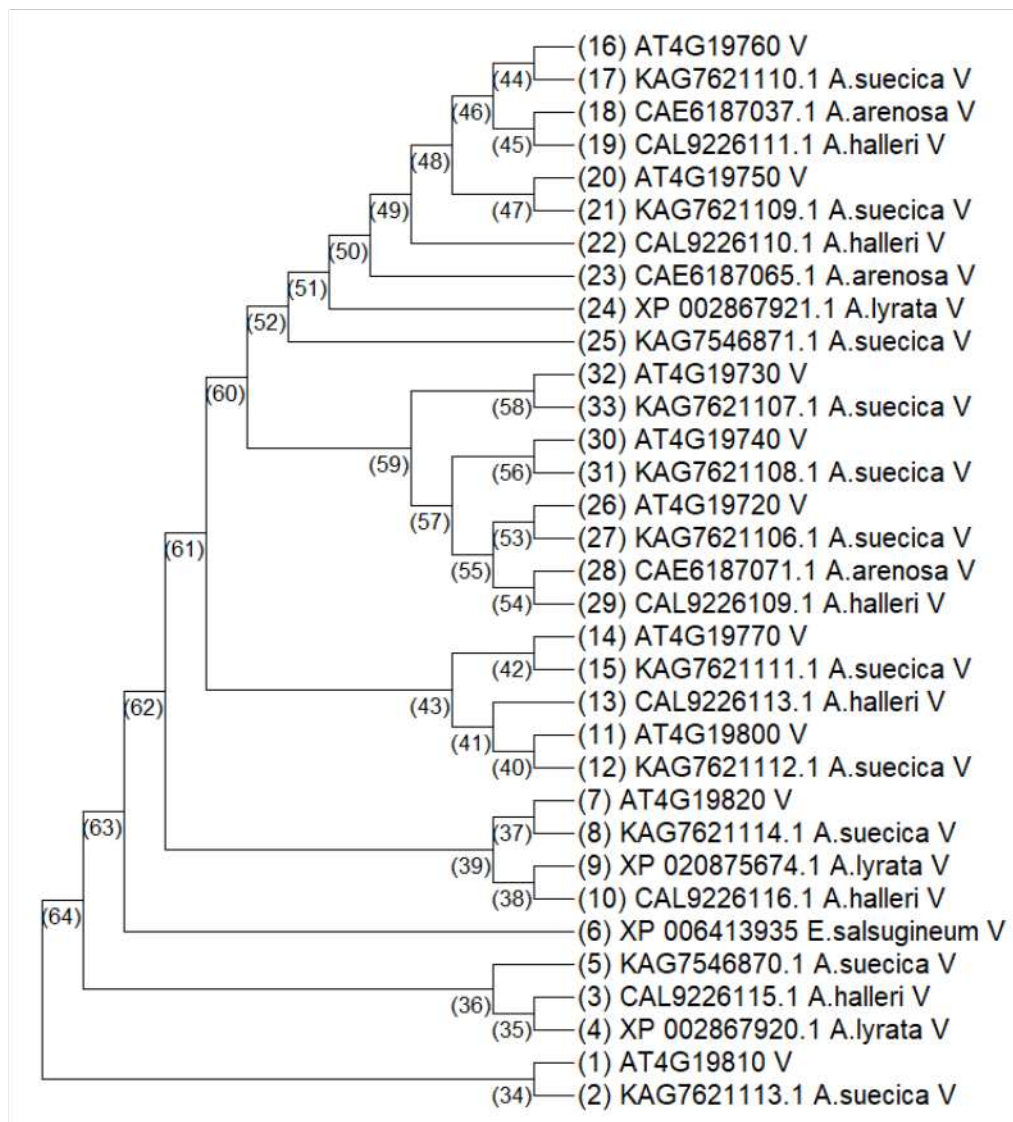
Fonte: Autoral. Gerada via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025). Cada número em parêntese representa o identificador nó. Árvore gerada somente com o CatD de cada sequência.

**APÊNDICE M – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE IV DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. arenosa* E *E. salsugineum***



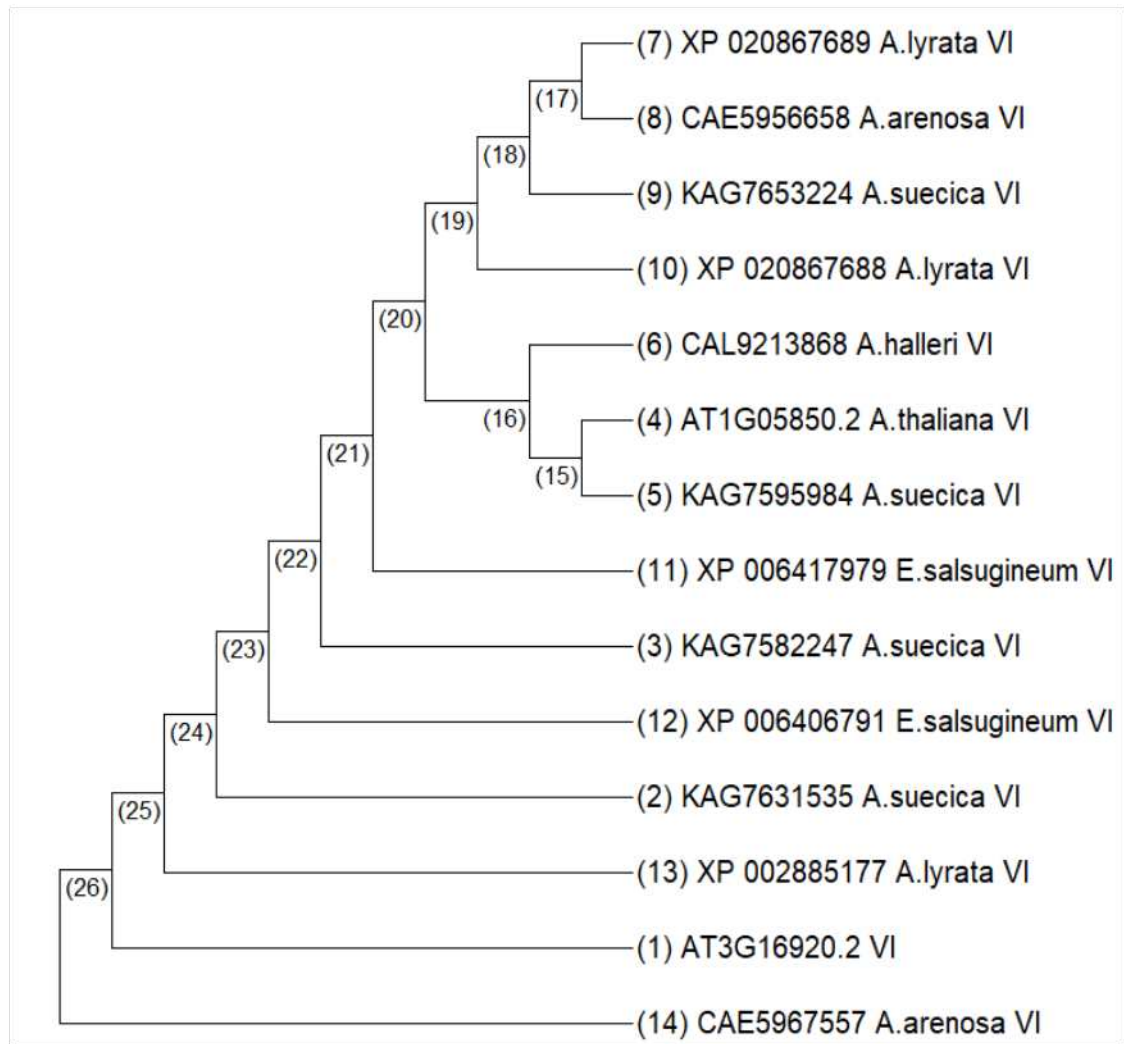
Fonte: Autoral. Gerada via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025). Cada número em parêntese representa o identificador nó. Árvore gerada somente com o CatD de cada sequência.

**APÊNDICE N – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE V DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. suecica*, *A. arenosa* E *E. salsugineum***



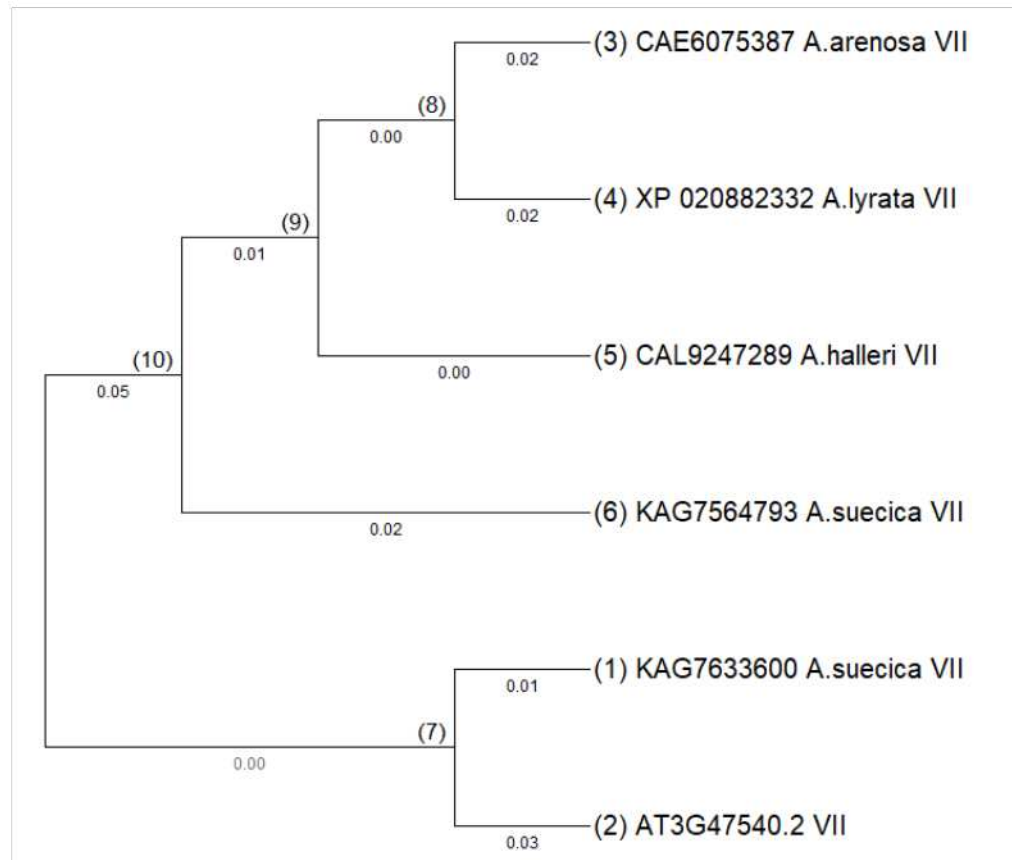
Fonte: Autoral. Gerada via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025). Cada número em parêntese representa o identificador nó. Árvore gerada somente com o CatD de cada sequência.

**APÊNDICE O – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE VI DE *A. thaliana*, *A. halleri*, *A. lyrata*, *A. suecica*, *A. arenosa* E *E. salsugineum***



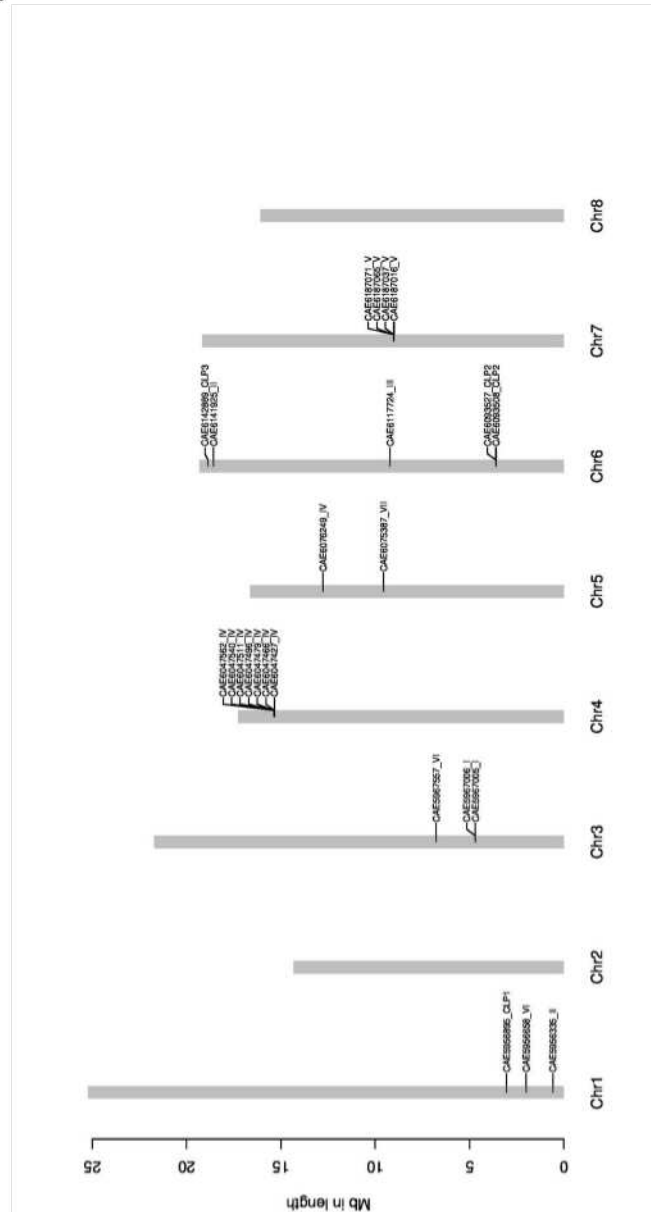
Fonte: Autoral. Gerada via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025). Cada número em parêntese representa o identificador nó. Árvore gerada somente com o CatD de cada sequência.

**APÊNDICE P – ÁRVORE FILOGENÉTICA DE QUITINASES DE CLASSE VII DE  
*A. thaliana*, *A. halleri*, *A. lyrata*, *A. suecica* E *A. arenosa***



Fonte: Autoral. Gerada via software IQ-TREE (Minh et al., 2020) (<https://iqtree.github.io/about/>), sendo LG+G+F o modelo de substituição de aminoácidos pelo método de Máxima Verossimilhança (*Maximum likelihood*; ML). As figuras de árvores foram obtidas via software MEGA 12 (Stecher et al., 2025). Cada número em parêntese representa o identificador nó. Árvore gerada somente com o CatD de cada sequência.

**APÊNDICE Q – MAPA CROMOSSÔMICO COM GENES DE QUITINASES DE *Arabidopsis arenosa***



Fonte: Autoral. Imagem gerada com o pacote Genes\_on\_chr ([https://github.com/xukai/Genes\\_on\\_Chr/tree/master](https://github.com/xukai/Genes_on_Chr/tree/master))

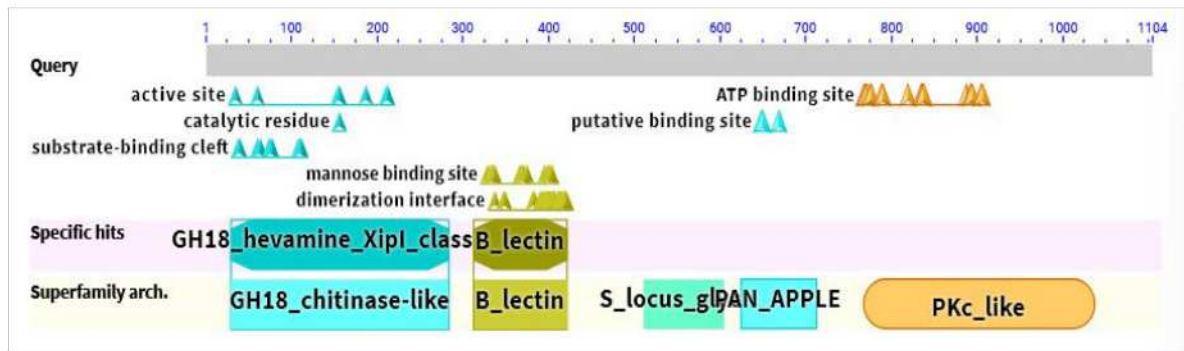


**APÊNDICE S – QUANTIDADE DE GENES QUE CODIFICAM CADA UMA DAS  
CLASSES DE  
QUITINASES EM *Arabidopsis thaliana*, *Arabidopsis suecica*, *Arabidopsis arenosa*,  
*Arabidopsis lyrata*, *Arabidopsis halleri* e *Eutrema salsugineum***

<b>Classe</b>	<b><i>A.</i> <i>thaliana</i></b>	<b><i>A. suecica</i></b>	<b><i>A. arenosa</i></b>	<b><i>A.</i> <i>lyrata</i></b>	<b><i>A.</i> <i>halleri</i></b>	<b><i>E.</i> <i>salsugineum</i></b>
I	1	3	2	1	1	1
II	2	3	2	2	2	2
III	1	2	1	1	1	2
IV	8	14	8	8	11	10
V	9	11	3	3	6	1
VI	2	4	1	3	2	2
VII	1	2	1	1	1	0
<b>Total:</b>	24	39	18	19	24	18

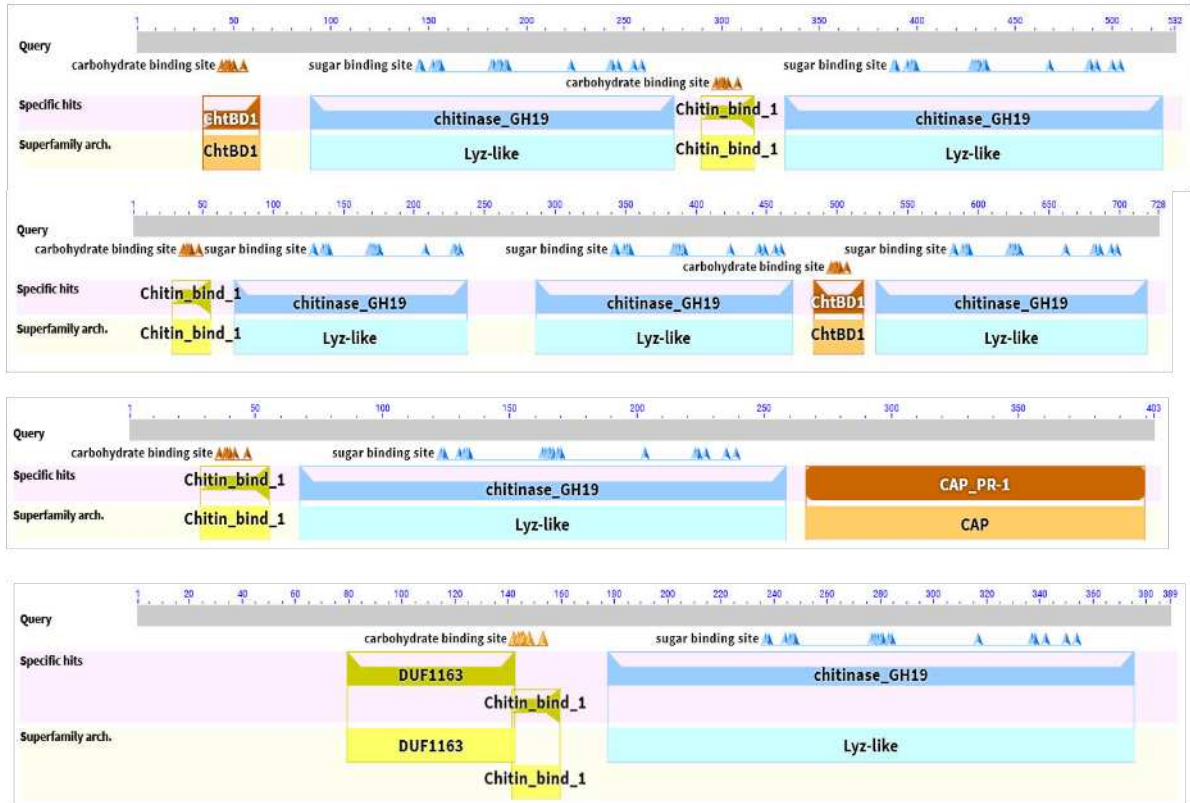
Fonte: Autoral (2025).

## ANEXO A – DOMÍNIOS CONSERVADOS DA PROTEÍNA CAE6117724\_A.arenosa\_III



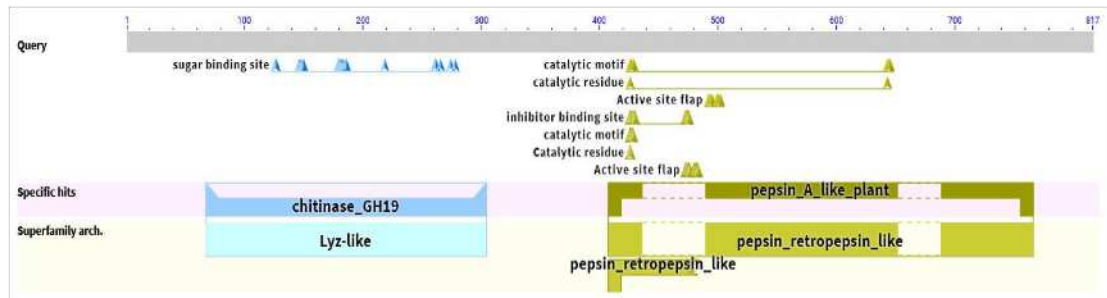
Fonte: Conserved Domains Database (CDD) (<https://www.ncbi.nlm.nih.gov/Structure/cdd>)

**ANEXO B – DOMÍNIOS CONSERVADOS DAS PROTEÍNAS  
KAG7574778.2\_A.suecica\_class\_IV, CAE6047466.3\_A.arenosa\_class\_IV,  
CAL9217362.1\_A.halleri\_class\_IV e CAE6076249.1\_A.arenosa\_class\_IV,  
respectivamente**



Fonte: Conserved Domains Database (CDD) (<https://www.ncbi.nlm.nih.gov/Structure/cdd>)

## ANEXO C – DOMÍNIOS CONSERVADOS DA PROTEÍNA CAE5956658\_A.arenosa\_VI



Fonte: Conserved Domains Database (CDD) (<https://www.ncbi.nlm.nih.gov/Structure/cdd>)