

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS

THAYNÁ CRISTINA STOFEL ANDRADE

ANÁLISE DE GENES DE RESISTÊNCIA NO TRANSCRIPTOMA DE *Panicum
maximum* CV. TAMANI

Dourados, MS

2018

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS

Thayná Cristina Stofel Andrade

**ANÁLISE DE GENES DE RESISTÊNCIA NO TRANSCRIPTOMA DE
Panicum maximum CV. TAMANI**

Trabalho apresentado ao Curso de Biotecnologia da Universidade Federal da Grande Dourados como parte dos requisitos para obtenção do título de Bacharel em Biotecnologia.

Orientador: Prof. Dr. Rodrigo Matheus Pereira

Dourados, MS

2018

Dados Internacionais de Catalogação na Publicação (CIP).

A554a Andrade, Thayná Cristina Stofel
ANÁLISE DE GENES DE RESISTÊNCIA NO TRANSCRIPTOMA DE
Panicum maximum CV. TAMANI / Thayná Cristina Stofel Andrade --
Dourados: UFGD, 2018.
87f. : il. ; 30 cm.

Orientador: Rodrigo Matheus Pereira

TCC (Graduação em Biotecnologia) - Faculdade de Ciências
Biológicas e Ambientais, Universidade Federal da Grande Dourados.
Inclui bibliografia

1. Forrageira. 2. Estresse hídrico. 3. Capim-mombaça. I. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

©Direitos reservados. Permitido a reprodução parcial desde que citada a fonte.

Thayná Cristina Stofel Andrade

**ANÁLISE DE GENES DE RESISTÊNCIA NO TRANSCRIPTOMA DE
Panicum maximum CV. TAMANI**

Aprovado em 19 de julho de 2018.

BANCA EXAMINADORA

Prof. Dr. Rodrigo Matheus Pereira

Universidade Federal da Grande Dourados

Profa. Dra. Liliam Silvia Candido

Universidade Federal da Grande Dourados

Profa. Dra. Alexeia Barufatti Grisolia

Universidade Federal da Grande Dourados

Dedico este trabalho à
minha família e ao meu
namorado Mateus Costa
Brasil que me ajudaram
chegar até aqui.

“Que aproveita ao homem ganhar o mundo inteiro e perder a sua alma?”

Marcos 8.36

AGRADECIMENTOS

Agradeço primeiramente à Deus que fez a criação, e me deu a vida e inteligência para que pudesse explorar e aprender sobre Sua criação. Agradeço, pois me permitiu completar mais uma etapa em minha vida e porque Ele foi e é minha força, esperança e abrigo. Que este trabalho seja para a honra e glória do nome do Senhor Jesus Cristo.

Agradeço à minha família por me apoiar, me incentivar, orar por mim, cuidar de mim mesmo não estando perto.

Agradeço ao meu namorado Mateus Costa Brasil por estar presente em todos os momentos me dando suporte, principalmente nos difíceis em que precisei dele.

Agradeço ao meu orientador Prof. Dr. Rodrigo Matheus Pereira pela oportunidade, pelo aprendizado, por ter acreditado em mim e sempre estar disposto à me atender.

Agradeço a pesquisadora Dra. Lucimara Chiari que através da Embrapa Gado de Corte permitiu o acesso ao transcriptoma da planta estudada nesse trabalho.

Agradeço aos membros da Igreja Presbiteriana de Dourados por me acolherem e me ajudarem em tantos momentos.

Agradeço aos colegas do grupo de bioinformática pelo companheirismo, principalmente à Fátima Termos por me ajudar no desenvolvimento deste trabalho.

À todos sou eternamente grata.

RESUMO

As espécies do gênero *Panicum* englobam importantes forrageiras que se destacam no Brasil, devido a alta produtividade, capacidade de adaptação ao pastejo, a condições ambientais e ao manejo, principalmente em áreas cultivadas. Essa alta capacidade de adaptação é de grande importância para a comunidade científica, principalmente em condições de estresse, pois revela genes que podem ser usados para transformação genética. O estresse hídrico pode apresentar impactos significativos no crescimento e desenvolvimento das plantas. Os genes que são expressos nessas condições, podem ser analisados através da bioinformática, que é uma ferramenta que possibilita diferentes análises como a transcriptômica. Com o presente trabalho, objetivou-se buscar genes de resistência a doenças em *P. maximum* através da transcriptômica, levantando os domínios e classes de proteínas de resistência e fazendo a modelagem 3D de proteínas com os domínios encontrados. Isso foi feito realizando um alinhamento local dos transcritos, organizando os resultados em um banco de dados. Em seguida, foi realizado um levantamento de domínios e classes de resistência, seguido da modelagem 3D. A partir da condição fisiológica de estresse hídrico de *Panicum maximum* foi verificada a presença de 8 diferentes domínios de proteínas relacionadas à resistência, sendo os domínios mais expressivos Serina-Treonina-Quinase, NBS e LRR. Observou-se a expressão de 10 diferentes classes de proteínas relacionadas à resistência, sendo as classes mais expressivas RLP, NL e CNL; e foi observado similaridade das sequências de resistência entre *Panicum maximum* e 13 diferentes espécies, sendo as espécies com mais sequências similares *Setaria italica*, *Panicum virgatum* e *Sorghum bicolor*. A modelagem 3D de proteínas de cada domínio observado possibilitou uma compreensão maior das funções de cada uma das proteínas modeladas, tendo como resultado modelos com um alto padrão de qualidade.

Palavras-chave: forrageira; estresse hídrico; capim-mombaça.

ABSTRACT

Species of the genus *Panicum* include important forages that stand out in Brazil, due to high productivity, adaptability to grazing, environmental conditions and management, especially in cultivated areas. This high adaptability is of great importance to the scientific community as it reveals genes that can be used for genetic transformation. Hydric stress can have significant impacts on plant growth and development. The genes that are expressed under these conditions can be analyzed through bioinformatics, which is a tool that enables different analyzes such as transcriptomics. With the present work, the objective was to search for resistance genes to *P. maximum* diseases through transcriptomy, raising the domains and classes of resistance proteins and doing 3D modeling of proteins with the domains found. This was done by performing a local alignment of the transcripts, organizing the results into a database. Next, a survey of domains and resistance classes was carried out, followed by 3D modeling. From the physiological condition of *Panicum maximum* water stress, the presence of 8 different resistance-related protein domains was verified, being the most expressive domains Serine-Threonine-Kinase, NBS and LRR. The expression of 10 different classes of resistance-related proteins was observed, the most expressive classes being RLP, NL and CNL; and similarity of resistance sequences between *Panicum maximum* and 13 different species was observed, being the species with more sequences similar *Setaria italica*, *Panicum virgatum* and *Sorghum bicolor*. The 3D modeling of proteins from each observed domain allowed a greater understanding of the functions of each of the modeled proteins, resulting in models with a high quality standard.

Keywords: forage; hydric stress; Mombasa grass.

LISTA DE FIGURAS

Figura 1. Um resumo esquemático dos mecanismos de detecção do patógeno inato da imunidade e da resposta imune das plantas, caracterizado pelos patossistemas. PTI, desencadeada por imunidade PAMP; PAMPs, padrões moleculares associados a patógenos; PRRs, receptores de reconhecimento de padrões; ROS, espécies que reagem ao oxigênio; ETI, imunidade desencadeada por efetores; NLR, receptor do tipo domínio de oligomerização de ligação a nucleotídeos; HR, resposta hipersensível; JA, ácido jasmônico; ET, etileno; SA, ácido salicílico; COI1, proteína insensível à coronária 1; EIN2, proteína insensível a etileno 2 (MILLER et al.; 2017).	24
Figura 2. Arquitetura de domínio de proteínas NLR e STAND de plantas e animais representativas (BENTHAM et al.; 2016).	25
Figura 3. Predição da estrutura tridimensional de uma proteína (VERLI; 2014).	38
Figura 4. Etapas de predição de estrutura tridimensional de proteínas usando o método de Modelagem Comparativa (VERLI; 2014).	39
Figura 5. Fluxograma mostrando a sequência de materiais e métodos usados. A – Embrapa Gado de Corte – MS; B – <i>Panicum maximum</i> ; C – Extração de RNA; D – Sequenciamento Illumina; E – FASTX-Toolkit; F – FastQC; G – Trinity; H – Blastn; I – PRGdb; J – MySQL; K – Protein Data Bank; L – SWISS-Model; M – Phyre2.	45
Figura 6. Modelagem 3D da (A) proteína do domínio NBS e seu (B) modelo transmembrana.	64
Figura 7. Modelagem 3D da (A) proteína do domínio LRR e seu (B) modelo transmembrana.	65
Figura 8. Modelagem 3D da (A) proteína do domínio MLO e seu (B) modelo transmembrana.	66
Figura 9. Modelagem 3D da proteína do domínio TIR.	67
Figura 10. Modelagem 3D da (A) proteína do domínio Receptor-Like Kinase e seu (B) modelo transmembrana.	68

Figura 11. Modelagem 3D da (A) proteína do domínio Ser-Thr Kinase e seu (B) modelo transmembrana.	69
Figura 12. Modelagem 3D da (A) proteína do domínio RPW8 e seu (B) modelo transmembrana.	72
Figura 13. Modelagem 3D da (A) proteína do domínio GNK2 e seu (B) modelo transmembrana.	71

LISTA DE TABELAS

Tabela 1. Classes de proteínas de resistência presentes em <i>Panicum maximum</i>	47
Tabela 2. Domínios de resistência simples presentes em <i>Panicum maximum</i>	49
Tabela 3. Domínios de resistência conjugados presentes em <i>Panicum maximum</i>	50
Tabela 4. Espécies relacionadas à <i>Panicum maximum</i>	61

LISTA DE ABREVIações

ARC: Região de conexão entre NBS e LRR

ATP: Adenosina trifosfato

AVR: Avirulência

BLAST: Ferramenta básica de busca por alinhamento local

CaM: Calmodulina

CaMBD: CaM C-terminal

CC: Bobina enrolada

cDNA: DNA complementar

CHS: Chalcona sintase

CNL: Classe CC-NBS-LRR

CNV: Variação do número de cópias

CRISPR: Repetições palindrômicas curtas agrupadas e regularmente interespaçadas

DNA: Ácido desoxirribonucléico

EST: Tags de sequência expressa

ETI: Imunidade desencadeada por efetor

Genes Avr: Genes de virulência

Genes R: Genes de resistência

GNK2: Domínio de *Ginkgo biloba*

GTP: Guanina trifosfato

HR: Resposta hipersensível

HRGP: Glicoproteínas ricas em hidroxiprolina

IL: Interleucina

IRO: Intermediários reativos de oxigênio

LOX: Lipoxigenase

LRR: Repetições ricas em leucina

LysM: Lisina

MLO: Mutação induzida

NB-ARC: Domínio NBS e ARC

NBS: Sítio de ligação de nucleotídeos

NGS: Sequenciamento de nova geração

NLR: Classe de resistência dos domínios LRR e NBS

NOD: Domínio de oligomerização de ligação de nucleotídeos

OREST: Fase de leitura aberta EST

ORF: Fase de leitura aberta

PAL: Fenilalanina amônia liase

PAMP: Padrão molecular associado à patógenos

PCR: Reação em cadeia da polimerase

PRGdb: Banco de dados de genes de resistência de plantas

PRR: Receptores de reconhecimento de padrões

QDR: Resistência quantitativa à doença

QRL: Locus de resistência quantitativa

RGA: Análogos de genes de resistência

RLK: Receptor tipo quinase

RLP: Receptores associados à membrana

RNA: Ácido ribonucléico

RNA-seq: Sequenciamento de mRNA

RPW8: Resistência ao oídio 8

Ser-Thr Kinase: Serina-Treonina quinase

SNARE: Receptor de proteína de ligação à proteína de fusão N-etilmaleimida solúvel

SNP: Polimorfismo de nucleotídeo único

TALEN: Nucleases efectoras de ativação de transcrição

TIR: Receptor Toll/Interleucina-1

TLR: Receptores tipo Toll

TM: Região transmembrana

TNL: TIR-NBS-LRR

SUMÁRIO

1.	Introdução.....	16
1.1.	Objetivos.....	17
2.	Revisão bibliográfica.....	18
2.1.	<i>Panicum maximum</i> cv. Tamani.....	18
2.2.	Estresse à fatores bióticos e abióticos em plantas.....	19
2.3.	Mecanismos de defesa das plantas.....	22
2.4.	Bioinformática.....	28
2.5.	Transcriptômica.....	33
2.6.	Modelagem 3D de proteínas.....	36
3.	Material e métodos.....	41
3.1.	Amostras.....	41
3.2.	Condições de estresse fisiológico.....	41
3.3.	Extração de RNA, biblioteca de cDNA e sequenciamento.....	41
3.4.	Controle de qualidade e <i>assembly de novo</i>	42
3.5.	Alinhamento local.....	43
3.6.	Levantamento das classes e domínios de resistência.....	43
3.7.	Modelagem 3D de proteínas.....	44
4.	Resultados e discussão.....	46
4.1.	Classes de resistência.....	47
4.2.	Domínios de resistência.....	48
4.2.1.	Domínio de resistência Ser-Thr Kinase.....	50
4.2.2.	Domínio de resistência LRR.....	51
4.2.3.	Domínio de resistência NBS.....	53
4.2.4.	Domínio de resistência GNK2.....	54
4.2.5.	Domínio de resistência MLO.....	55
4.2.6.	Domínio de resistência Receptor-Like Kinase.....	56
4.2.7.	Domínio de resistência TIR.....	58
4.2.8.	Domínio de resistência RPW8.....	59
4.3.	Espécies relacionadas.....	60
4.4.	Modelagem 3D de proteínas.....	63
5.	Conclusões.....	73

6.	Considerações Finais.....	74
7.	Referências.....	75

1. INTRODUÇÃO

As espécies do gênero *Panicum* englobam importantes forrageiras que se destacam no Brasil, devido à sua alta produtividade, capacidade de adaptação ao pastejo, a condições ambientais e ao manejo, principalmente em áreas cultivadas. O uso dessas forrageiras, como opção de produção de massa para o plantio direto ou em integração lavoura-pecuária, é uma realidade nas regiões produtoras de grãos, sobretudo no Brasil central pecuário (QUEIROZ et al.; 2014).

Devido ao seu comportamento sésstil, as plantas são continuamente expostas a uma ampla gama de estresses. Os principais estresses abióticos que afetam plantas e culturas no campo estão sendo amplamente estudados, constituem: a seca, a salinidade, o calor, o frio, o arrefecimento, o congelamento, o nutriente, a intensidade luminosa alta, o ozônio (O₃) e as tensões anaeróbicas (SUZUKI et al.; 2014).

Para defender-se dos estresses sofridos, as plantas desenvolveram uma grande variedade de mecanismos de resistência a doenças, que incluem síntese de fitoalexinas, antibióticos, inibidores de protease; deposição de material de parede celular e acumulação de enzimas hidrolíticas como as quitinases. Aparentemente, a resistência da planta depende da habilidade do hospedeiro em reconhecer o patógeno, e rapidamente induzir os mecanismos de defesa a tempo de limitar a ação do patógeno (CHRISTOPOULOU et al.; 2015). Desta forma, o conhecimento dos genes que possuem expressão alterada em resposta a determinado fator ambiental pode revelar grande parte das bases genéticas determinantes ou envolvidas neste processo biológico (VINECKY et al, 2005).

O transcriptoma é o conjunto completo de transcrições em uma célula e sua quantidade, para um estágio de desenvolvimento específico ou condição fisiológica (WANG et al.; 2009). Através da análise do transcriptoma é possível determinar quando e onde cada gene é expresso em determinado tecido do organismo e, ainda, é possível quantificar essa expressão (SANGI & BIANCHETTI.; 2017).

Sendo assim, esta pesquisa se mostra de grande interesse, visto que o conhecimento dos genes de resistência no capim *Panicum maximum* pode

proporcionar informações adicionais à comunidade científica e também ser usado em trabalhos futuros para o melhoramento da espécie, ou de outras espécies, como para superexpressar genes de resistência encontrados ou transferência de genes de resistência para outra planta. Cabe ressaltar que esse é o primeiro trabalho realizado envolvendo o transcriptoma desta planta.

1.1 Objetivos

Objetivo geral: prospectar genes de resistência à doença em *Panicum maximum* utilizando as ferramentas da bioinformática

Objetivos específicos:

- Encontrar os domínios de proteínas de resistência à doença presentes em *P. maximum*;
- Verificar as classes de proteínas relacionadas à resistência à doença em *P. maximum*;
- Levantamento das espécies relacionadas à *P. maximum*;
- Predizer a estrutura tridimensional dos principais genes de resistência à doença encontrados.

2. REVISÃO BIBLIOGRÁFICA

2.1 *Panicum maximum* cv. Tamani

Dentre as principais forrageiras cultivadas no Brasil, a espécie *Panicum maximum* destaca-se pelo grande potencial de produção de matéria seca, fácil adaptação e boa qualidade da forrageira tanto nas características agronômicas quanto em valores nutricionais (NONATO, et al.; 2016).

Lançada em março de 2015 (BRAGA et al.; 2015), a cultivar BRS Tamani é o primeiro híbrido de *Panicum maximum* da Embrapa, sendo resultado do cruzamento entre a planta sexual S12 e o acesso apomítico T60 (BRA-007234) (PEREIRA et al.; 2016) em 1992. Apresenta como características ser cespitosa de porte ereto e baixo (até 1,3 m) com folhas verde-escuras, longas, finas (até 1,9 cm) e arqueadas (PEREIRA et al.; 2016), decumbentes de alta qualidade (JANK et al.; 2017), com elevada produção de folhas de alto valor nutritivo (elevados teores de proteína bruta e digestibilidade), produtividade e vigor, sendo de fácil manejo (ZANELA & DERETI; 2017) e possui florescimento precoce (PEREIRA et al.; 2016).

A BRS Tamani tem boa resistência às cigarrinha-das-pastagens e resistência intermediária à mancha foliar (*Bipolaris maydis*). Não é recomendada para solos de baixa fertilidade; baixa tolerância a solos encharcados; suscetível ao nematoide das lesões radiculares (*Pratylenchus brachyurus*) (PEREIRA et al.; 2016). Apresenta maior ganho por animal que a cultivar 'Massai', devido ao seu alto perfilhamento e folhas finas e decumbentes, e também apresenta a maior facilidade de manejo de todas as cultivares da espécie (JANK et al.; 2017).

Entretanto, não é indicada para áreas sujeitas a alagamentos, mesmo que temporários, por apresentar baixa tolerância ao encharcamento do solo. Em condições de baixas temperaturas, apresenta maior persistência que as cultivares Massai e Tanzânia, sendo semelhante à cultivar Mombaça (ZANELA & DERETI; 2017). É recomendada para solos de média a alta fertilidade. Avaliada em pastejo alternado por dois anos, suportou 3,2 e 1,6 uA/ha durante o período das águas e da seca, respectivamente, apresentando ganho médio de 681 kg de peso vivo por hectare ao ano. Recomenda-se o uso de 3 a 4 kg/ha de sementes puras viáveis, semeadas na profundidade de 2,5 a 5 cm e

incorporadas com grade niveladora. Deve ser manejada, preferencialmente, sob lotação rotacionada, com altura de resíduo de pelo menos 20-25 cm e períodos de descanso não superiores a 28 dias, na época chuvosa (PEREIRA et al.; 2016).

2.2 Estresse à fatores bióticos e abióticos em plantas

Devido ao seu comportamento sésstil, as plantas são continuamente expostas a uma ampla gama de estresses. Os principais estresses abióticos que afetam plantas e culturas no campo estão sendo amplamente estudados. Incluem a seca, a salinidade, o calor, o frio, o arrefecimento, o congelamento, o nutriente, a intensidade luminosa alta, o ozônio (O₃) e as tensões anaeróbicas (SUZUKI et al.; 2014).

Em condições naturais, combinações de dois ou mais estresses, como a seca e salinidade, salinidade e calor, e combinações de seca com temperatura extrema ou alta intensidade luminosa são comuns a muitas áreas agrícolas em todo o mundo e podem afetar a produtividade das culturas (SUZUKI et al.; 2014).

O alumínio (Al) é um dos principais fatores limitantes para o cultivo de culturas em solos ácidos. Em condições ácidas do solo, o Al é ionizado em alumínio trivalente (Al³⁺), que está prontamente disponível e altamente reativo no solo. O Al pode interagir com uma série de estruturas extracelulares e intracelulares, inibindo o crescimento da planta e reduzindo a produtividade e o rendimento (SALGADO et al.; 2017).

O déficit hídrico é observado em muitas culturas podendo apresentar impactos significativos no crescimento e desenvolvimento das plantas provocando alterações no comportamento, cuja tolerância vai depender do genótipo, da duração, da severidade e do estágio de desenvolvimento da planta (BRITO et al.; 2012).

Os índices fisiológicos, como a taxa de assimilação de CO₂, transpiração, condutância estomática e concentração interna de CO₂, também são afetados por fatores ambientais, como luz, temperatura, concentração de CO₂ e disponibilidade de água e nutrientes; assim, pode-se conhecer, com base no

estudo das interações desses parâmetros com cada fator ambiental, em particular o estado hídrico da planta, a eficiência do crescimento e a habilidade de adaptação às condições ambientais de uma espécie ou variedade (BRITO et al.; 2012).

As plantas sujeitas ao alagamento possuem sua condutância estomática afetada e reduz as taxas de fotossíntese e de crescimento, resultando em queda da produção da parte aérea e de raízes. Contudo, as plantas desenvolvem certos mecanismos de adaptação metabólica, que envolve a respiração anaeróbica, o desenvolvimento de aerênquima e formação de raízes adventícias (BONFIM-SILVA et al.; 2011).

Entre todos os tipos de estresses abióticos, a desidratação e os extremos de temperatura são os que mais afetam o desenvolvimento vegetal. As plantas respondem ao estresse hídrico e às altas e baixas temperaturas com diversas mudanças relacionadas ao seu desenvolvimento e a sua fisiologia. Uma das respostas das plantas aos estresses é a expressão de um grande número de genes, cujos produtos podem estar envolvidos em diversas funções adaptativas em condições de estresse. Genes induzidos durante os estresses abióticos codificam proteínas que não funcionam apenas na proteção das células contra o estresse, mas também na expressão gênica e transdução de sinal em resposta ao estresse (ITO et al.; 2013).

Em 2002 e 2004, revelou-se que a resposta molecular das plantas a uma combinação de seca e estresse térmico é única e não pode ser extrapolada diretamente da resposta das plantas à seca ou ao estresse térmico aplicado individualmente. Desde então, muitos estudos envolvendo outras combinações de estresses demonstraram que, apesar de um certo grau de sobreposição, cada condição de estresse exigia um mecanismo de resposta única, adaptado às necessidades específicas da planta, e que cada combinação de dois ou mais estresses diferentes também pode exigir uma resposta específica (SUZUKI et al.; 2014).

A salinidade também tem sido um problema, particularmente nas regiões tropicais onde prevalecem condições climáticas adversas como evapotranspiração e temperaturas elevadas. Ela causa grandes distúrbios no

metabolismo das plantas, acarretando restrição de crescimento e perda de produtividade (SILVEIRA et al.; 2010).

Uma comparação de todos os principais desastres climáticos dos EUA que excederam um bilhão de dólares cada, entre 1980 e 2012, indica que uma combinação de seca e estresse térmico causou grandes perdas agrícolas na ordem de \$ 200 bilhões. Em contrapartida, durante o mesmo período, apenas a seca, atuando de forma isolada causou \$ 50 bilhões em danos à produção agrícola (SUZUKI et al.; 2014).

Além disso, os modelos atuais de previsão climática indicam um aumento gradual da temperatura ambiente e um aumento na frequência e amplitude do estresse térmico no futuro próximo. E, as altas temperaturas serão acompanhadas por outros desastres climáticos, como extensões prolongadas, que poderiam afetar drasticamente a produção de culturas em todo o mundo. Por conseguinte, existe uma necessidade urgente de gerar culturas com tolerância melhorada às combinações de estresse (SUZUKI et al.; 2014).

Além dos estresses abióticos, em condições naturais, as plantas enfrentam a ameaça de infecção por agentes patogênicos (incluindo bactérias, fungos, vírus e nematóides) e ataques por pragas herbívoros (SUZUKI et al.; 2014), ambos conhecidos como estresse biótico. Muitas plantas passam por mudanças bioquímicas significativas em resposta ao ataque de estressores bióticos, como fitopatógenos ou insetos herbívoros (NASCIMENTO & BARRIGOSI; 2014).

Quando o ataque é realizado por patógenos, as plantas se defendem com mecanismos de defesa passiva ou pré-existente e ativa ou induzida. A suscetibilidade e resistência das plantas estão relacionadas com a eficiência desta resposta de defesa ao ataque de fitopatógenos. A resistência a doenças em plantas é conhecida por ser controlada geneticamente conferindo resistência a fitopatógenos com um gene avirulento para um reconhecido evento específico (NASCIMENTO & BARRIGOSI; 2014).

O alcance do habitat de pragas e patógenos pode ser influenciado pelas mudanças climáticas, por exemplo, facilitando a disseminação de patógenos. Além disso, muitas condições de estresse abiótico demonstraram enfraquecer

os mecanismos de defesa das plantas e aumentar sua susceptibilidade à infecção por patógenos (SUZUKI et al.; 2014).

Além do mais, a ocorrência simultânea de diferentes estresses bióticos e abióticos resultou em um alto grau de complexidade nas respostas das plantas, uma vez que as respostas a esses estresses combinados são amplamente controladas por diferentes caminhos de sinalização que podem interagir e inibir um ao outro. As vias metabólicas e de sinalização envolvidas na resposta das plantas à combinação de estresse incluíram fatores de transcrição, fotossíntese, mecanismos antioxidantes, respostas de agentes patogênicos, sinalização hormonal e síntese de osmólitos. No entanto, a maioria dos mecanismos subjacentes à tolerância das plantas às combinações de estresse ainda são desconhecidas e são necessários mais estudos para abordá-las (SUZUKI et al.; 2014).

2.3 Mecanismos de defesa das plantas

A imunidade da planta pode ser definida simplesmente como a capacidade de uma fábrica para prevenir ou suportar ataque biológico por patógenos (MILLER et al.; 2017). Para defender-se dos estresses sofridos, as plantas desenvolveram uma grande variedade de mecanismos de resistência a doenças, que incluem síntese de fitoalexinas, antibióticos, inibidores de protease; deposição de material de parede celular e acumulação de enzimas hidrolíticas como as quitinases. Aparentemente, a resistência da planta depende da habilidade do hospedeiro em reconhecer o patógeno, e rapidamente induzir os mecanismos de defesa a tempo de limitar a ação do patógeno (CHRISTOPOULOU et al.; 2015).

O sistema de defesa da planta consiste em várias camadas de defesa. Os receptores de reconhecimento de padrões, localizados na membrana plasmática, podem detectar padrões moleculares de patógenos conservados, como a quitina ou flagelina, e desencadear a resistência basal de amplo espectro (CHRISTOPOULOU et al.; 2015) Os agentes patogênicos desenvolveram mecanismos para suprimir as defesas basais que incluem exclusão completa, inativação ou regulação do gene AVR (gene de avirulência), ou mutações

pontuais que permitem o reconhecimento para enganar, mantendo a função de virulência da proteína AVR (proteína de avirulência) (PETIT-HOUDENOT & FUDAL; 2017).

Em resposta, as plantas desenvolveram mecanismos que levam à virulência, receptores NLR (ligação de nucleotídeos e proteínas de repetição ricas em leucina) intracelulares que podem detectar as moléculas efetoras diretamente ou por meio dos compostos produzidos à partir delas e obter respostas de defesa potentes (imunidade desencadeada por efetor [ETI]) (Figura 1), que muitas vezes envolve a morte celular programada caracterizada pela resposta hipersensível (HR) (CHRISTOPOULOU et al.; 2015). E a percepção de patógenos na matriz extracelular por PRRs transmembranares (comumente chamados de imunidade desencadeada por PAMP – padrões moleculares associados à patógenos- [PTI]) (Figura 1) (BENTHAM et al.; 2016).

Os NLRs normalmente têm uma estrutura modular de tri-domínio, com papéis para cada um dos três domínios (Figura 2). A família é definida pela região central NOD pertencente à família ATPase; este domínio geralmente possui funções de oligomerização e regulamentação. O domínio C-terminal em plantas NLRs normalmente contém LRRs, mas podem consistir em outras repetições formadoras da superestrutura, e motifs WD/WD40, HEAT, ankirina ou TPR (tetratricopeptídeo); este domínio geralmente tem funções de detecção e auto-regulação. O domínio N-terminal geralmente tem um função de sinalização. Nas plantas, na maioria das vezes é um TIR (Toll-like/interleucina-1 receptor/resistência) ou um domínio de bobina enrolada (CC) (BENTHAM et al.; 2016).

As plantas contêm múltiplos genes de resistência fenotípica (R) que proporcionam resistência a diversos agentes patogênicos e pragas. A maioria dos genes envolvidos no reconhecimento de patógenos identificados até o momento codificam a ligação de nucleotídeos e proteínas de repetição ricas em leucina (NLR). A presença de motivos característicos compartilhados entre os genes R clonados facilitou a identificação de genes de resistência de candidatos numerosos (CHRISTOPOULOU et al.; 2015).

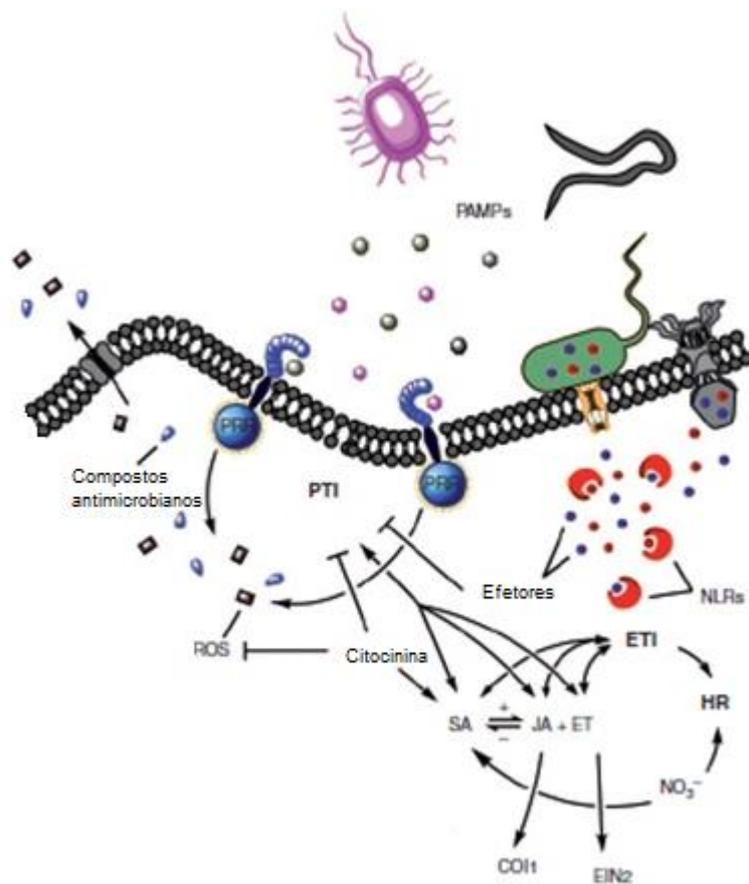


Figura 1. Um resumo esquemático dos mecanismos de detecção do patógeno inato da imunidade e da resposta imune das plantas, caracterizado pelos patossistemas. PTI, desencadeada por imunidade PAMP; PAMPs, padrões moleculares associados a patógenos; PRRs, receptores de reconhecimento de padrões; ROS, espécies que reagem ao oxigênio; ETI, imunidade desencadeada por efetores; NLR, receptor do tipo domínio de oligomerização de ligação a nucleotídeos; HR, resposta hipersensível; JA, ácido jasmônico; ET, etileno; SA, ácido salicílico; COI1, proteína insensível à coronária 1; EIN2, proteína insensível a etileno 2. (MILLER et al.; 2017).

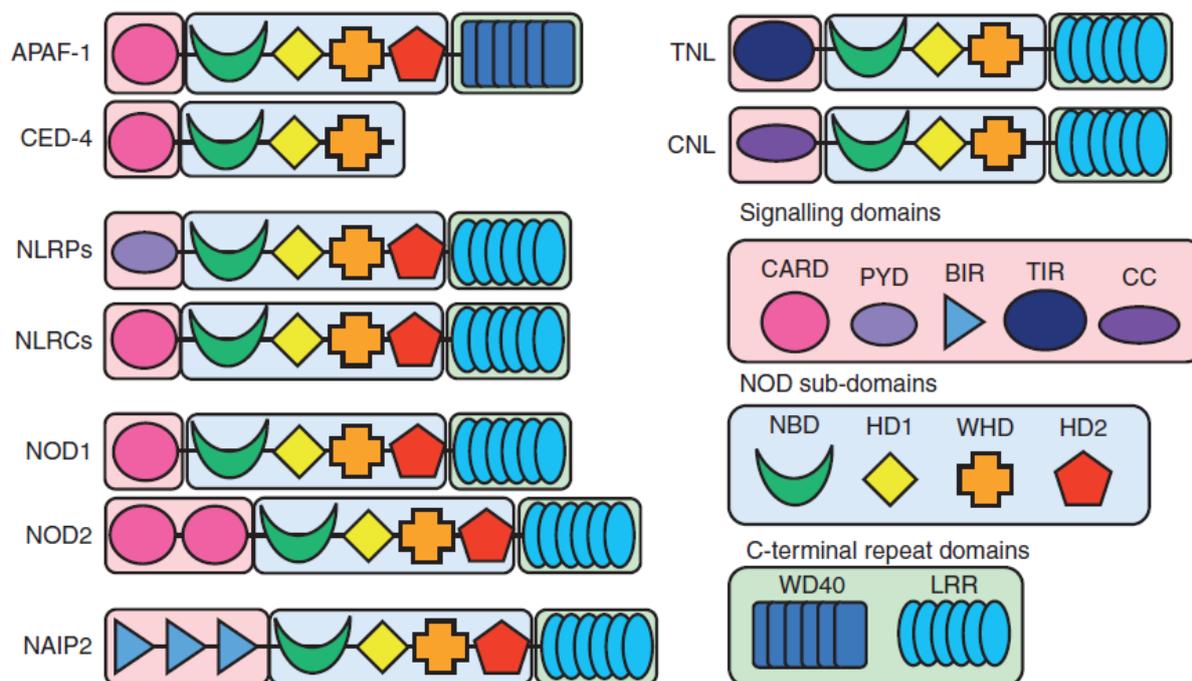


Figura 2. Arquitetura de domínio de proteínas NLR e STAND de plantas e animais representativas (BENTHAM et al.; 2016).

As sequências com similaridade aos genes R clonados geralmente são agrupadas em genomas de plantas e, muitas vezes, cosegregadas com múltiplas especificidades de resistência. A clonagem de genes R não só promove a compreensão das bases moleculares da resistência à doença, mas também tem implicações práticas significativas para os programas de reprodução, fornecendo marcadores robustos para a seleção assistida por marcadores e possibilitando a piramidação de genes R (combinação de dois ou mais transgenes em um organismo) (CHRISTOPOULOU et al.; 2015).

Os genes envolvidos na defesa vegetal podem ser divididos em três grupos. O primeiro é chamado de proteínas antinutricionais ou genes de defesa: codificam proteínas de defesa, tais como inibidores de protease, enzimas da biossíntese de compostos metabólicos secundários, um exemplo destas enzimas é a fenilalanina amônia liase (PAL), que produz intermediários reativos de oxigênio (IRO) que atuam no reforçamento da parede celular (SILVESTRINI et al.; 2005).

O segundo grupo é chamado genes de sinalização: produzem compostos voláteis, para repelir os insetos, comunicar o ataque entre partes das plantas e entre plantas. São genes que codificam enzimas envolvidas na síntese de flavonóides, que produzem compostos fenólicos tóxicos e proteínas inibidoras de proteinases. E o terceiro grupo, genes envolvidos no redirecionamento metabólico: genes que ativam a síntese de enzimas que produzem IRO, que podem atuar também na transdução de sinais, ativando a lipoxigenase (LOX), que participa da produção de ácido jasmônico (SILVESTRINI et al.; 2005).

Os domínios de resistência podem ser encontrados sozinhos nas proteínas, ou conjugados à outros domínios. Quando são encontrados juntos, são denominados classes, e essas podem ser agrupadas funcionalmente em cinco principais classes de resistência à doença distintas com base na presença de domínios específicos: a classe CNL compreende genes de resistência que codificam proteínas com pelo menos um domínio de bobina enrolada, um local de ligação de nucleotídeo e uma repetição rica em leucina (CC-NB-LRR). A classe TNL inclui aqueles com um domínio do receptor de interleucina Toll, um local de ligação a nucleotídeos e uma repetição rica em leucina (TIR-NB-LRR) (SANSEVERINO et al.; 2009).

A classe RLP, sigla para a proteína semelhante a um receptor, agrupa aqueles com um domínio receptor de serina-treonina quinase receptor e uma repetição extracelular rica em leucina (ser/thr-LRR). A classe RLK contém aqueles com um domínio quinase e uma repetição extracelular rica em leucina (Kin-LRR). A classe "Outros" inclui todos os outros genes que foram descritos como conferindo resistência através de diferentes mecanismos moleculares (SANSEVERINO et al.; 2009).

A resistência é geralmente reconhecida como vertical ou horizontal. A resistência vertical, relacionada à expressão de número reduzido de natureza dominante, é mais fácil de ser trabalhada em programas de melhoramento. Por outro lado, ela tem caráter temporário, ocorrendo frequentes quebras da resistência, devido a adaptações do patógeno ao genótipo (BOTELHO et al.; 2015). A resistência horizontal é efetiva contra várias raças do patógeno. É herdada de forma quantitativa e controlada por vários genes, cada um com um efeito menor (FILIPPI et al.; 2015).

A piramidização de genes R horizontais em um único genoma geralmente melhora seu fenótipo de resistência. Os genes de resistência podem ser combinados para produzir novos fenótipos horizontais ou verticais. A piramidização de genes R pode melhorar o espectro de resistência, portanto, os genes com espectro de resistência complementar podem ser selecionados de tal forma que as pirâmides de genes proporcionam resistência a um amplo conjunto de raças de patógenos (NELSON et al.; 2018).

A durabilidade da resistência depende de muitos fatores, incluindo a biologia, a genética e a evolução do patógeno ao qual confere resistência. A resistência vertical baseada em genes R é muitas vezes facilmente superada pela evolução das populações de patógenos, particularmente quando plantas resistentes são implantadas em grandes monoculturas. A estabilidade de um gene R está assim relacionada ao papel fisiológico e ao potencial evolutivo do efetor que ele reconhece (NELSON et al.; 2018).

Assim, o influxo constante de novas estratégias de reprodução e biotecnologia é indispensável para aumentar a imunidade das plantas e conferir resistência duradoura e de amplo espectro às plantas cultivadas contra patógenos microbianos. Algumas técnicas novas para aumentar a imunidade inata da planta e conferir resistência duradoura à doenças, são: Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas (CRISPR)/Cas e Nucleases Efetoras de Ativação de Transcrição (TALENs) (NEJAT et al.; 2017).

Os genes que desempenham papéis em resistência ou susceptibilidade a agentes patogênicos também podem afetar outros traços importantes, por exemplo, rendimento ou resposta a fatores abióticos como o estresse de água ou nutrientes. Além disso, a resistência a uma doença pode estar associada a resistência ou susceptibilidade a outra (NELSON et al.; 2018).

Além dos genes de resistência, existem os análogos de genes de resistência (RGAs) que apresentam possível resistência. Esses genes possuem similaridade com genes de resistência conhecidos. Alguns destes RGAs estão geneticamente ligados a locos de genes de resistência ou parte dos mesmos (GUIMARÃES et al.; 2005).

Os RGAs compreendem Receptores de Reconhecimento de Padrões (PRRs) e genes R e, a maioria tem domínios e motifs conservados. A maioria dos PRRs caracterizados são proteínas-quinases semelhantes a receptores localizados em superfície (RLKs) ou proteínas semelhantes a receptores associados à membrana (RLPs). Os RLKs possuem um domínio de detecção extracelular, uma região transmembrana (TM) e uma proteína quinase intracelular, contendo dois tipos de acordo com a estrutura do domínio, tipo de repetição rica em leucina (LRR), como FLS2, EFR e XA21, e motif de lisina (LysM), como CERK1. Os RLPs têm arquitetura de domínio semelhante a RLKs, exceto pela ausência de um domínio de quinase na região intracelular, como Cf-9 (tipo LRR), Eix1 e Eix2 (tipo LRR), CEBiP (tipo LysM) (LI et al.; 2016).

2.4 Bioinformática

Um modo de se estudar os mecanismos de resistência de plantas aos estresses bióticos e abióticos é o uso da bioinformática, que se refere ao emprego de ferramentas computacionais no estudo de problemas e questões biológicas. Neste caminho, da sequência de nucleotídeos até estruturas proteicas, alcançando por fim fármacos e mais produtos, diversas áreas do conhecimento estão envolvidas, como: biologia molecular, biologia celular, bioquímica, química, física e computação (VERLI; 2014).

Buscando tratar dados biológicos brutos, a bioinformática consolida-se como uma nova área do conhecimento, graças à crescente necessidade de desenvolver programas computacionais que permitam reconhecer sequências de genes; prever a configuração tridimensional de proteínas; identificar inibidores de enzimas; organizar e relacionar informação biológica; agrupar proteínas homólogas; estabelecer árvores filogenéticas; analisar experimentos de expressão gênica, entre outros (ARAÚJO et al.; 2017).

Vários recursos são disponibilizados pela bioinformática, como: mineração de textos e dados; sequenciamento do DNA e a Genômica; filogenia; RNAs e a transcriptômica; proteômica e estrutura de proteínas; metabolômica; farmacogenômica; biotecnologia (ESPINDOLA et al.; 2010).

Podemos traçar como momento chave da bioinformática o início da década de 1950, quando o trabalho clássico sobre a estrutura em hélice da molécula de DNA por James Watson e Francis Crick foi publicado. Além deste trabalho, outros trabalhos também foram muito importantes para esse momento, dentre eles destacam-se os trabalhos de Linus Pauling e Robert Corey, no início da década de 1950, e de Gopalasamudram N. Ramachandran, no início da década de 1960, que ofereceram as bases para a compreensão da estrutura tridimensional de proteínas (VERLI; 2014).

Com o desenvolvimento de computadores mais poderosos e com o avanço no entendimento dos determinantes da estrutura e da dinâmica proteica, tornam-se possíveis os primeiros estudos acerca da dinâmica e do enovelamento de proteínas por simulações de dinâmica molecular por Michael Levitt e Arieh Warshel, nos anos de 1970. A partir dos trabalhos destes e de outros pesquisadores, diversos avanços foram feitos progressivamente nos anos que se seguiram, tanto no entendimento de biomoléculas quanto no emprego de técnicas computacionais para retroalimentar este entendimento (VERLI; 2014).

O domínio da bioinformática cresceu em paralelo e com a internet, em que a rápida análise de dados e a troca de informações sobre os códigos biológicos e computacionais estão em convivência harmônica, por meio de suas múltiplas ramificações, gerenciando e integrando bancos de dados aplicáveis, e construindo sistemas *in silico* para simulação de formas naturais e modificadas de produtos específicos (ESPINDOLA et al.; 2010).

Os dados gerados pelo sequenciamento dos genomas de diferentes organismos transformaram a biologia. A integração de várias áreas do conhecimento permitiu avançar os estudos em relação à genômica, os processos de transcrição das informações contidas nos genes, a transcriptômica, bem como a compreensão do conjunto dos produtos destes genes pela proteômica. Neste contexto, promoveu-se o desenvolvimento e o aperfeiçoamento das técnicas que permitiram os avanços destas novas ciências ômicas, como a transcriptômica, proteômica e metabolômica, com o objetivo de isolar e caracterizar o RNA, as proteínas e os metabólitos, respectivamente; sendo possível devido também ao desenvolvimento da bioinformática (ESPINDOLA et al.; 2010).

Um grande desafio veio à tona com o advento das novas tecnologias de sequenciamento, ou sequenciamento de nova geração, do inglês *Next Generation Sequencing* (NGS), em termos de capacidade de armazenamento e processamento de dados, assim como a necessidade do desenvolvimento de novas ferramentas de análise (MASSRUHÁ & HIGA; 2014).

Dificuldades à parte, sem dúvidas o surgimento do sequenciamento de nova geração tem transformado várias áreas da pesquisa biológica, incluindo a agricultura. Os estudos genéticos foram largamente beneficiados com o avanço na obtenção de informações genômicas que podem ser aplicadas no pré-melhoramento e melhoramento genético de espécies animais e vegetais de interesse econômico, na caracterização de recursos genéticos com vistas à prospecção gênica e na descoberta de ativos biotecnológicos (MASSRUHÁ & HIGA; 2014).

As ferramentas desenvolvidas para a análise dos dados gerados pelos sequenciadores de nova geração incluem aquelas que permitem o alinhamento das sequências produzidas contra um genoma referência ou a montagem de novo das sequências geradas, resultando em um genoma “montado”, a detecção de polimorfismos do tipo SNP (Single Nucleotide Polymorphism) e ainda outras variações estruturais como deleções, inserções, rearranjos e variações no número de cópias de trechos do genoma - as CNVs (Copy Number Variation), a análise de transcriptomas em grande profundidade e a análise de metagenomas, a partir de comunidades microbianas (MASSRUHÁ & HIGA; 2014).

Cerca de 1.000 projetos já foram finalizados no mundo ou estão em fase de conclusão, sendo responsáveis pela disponibilização de uma vasta quantidade de dados biológicos, tais como: identificação, sequenciamento e descrição da estrutura tridimensional de proteínas e das similaridades dos genes dentro de um genoma e entre diferentes espécies (ARAÚJO et al.; 2017).

No Brasil, esse fenômeno pode ser acompanhado de perto pela efetiva execução e andamento de diversos projetos do Programa Genoma, acabando por consolidar uma nova concepção do desenho institucional da pesquisa no país. Além disso, tais iniciativas têm contribuído para o estabelecimento de uma cultura de interação e de integração entre as universidades e as empresas,

essencial quando se quer enfrentar o desafio necessário e impositivo da transformação do conhecimento no desenvolvimento de patentes e em riqueza (ARAÚJO et al.; 2017).

Uma busca de semelhança de sequência fornece frequentemente a primeira informação sobre uma nova sequência de DNA ou proteína. Uma pesquisa permite aos cientistas inferir a função de uma sequência de sequências semelhantes. Existem muitas maneiras de realizar uma busca de similaridade de sequência, mas provavelmente o método mais popular é a "Ferramenta básica de pesquisa de alinhamento local" (BLAST). O BLAST usa heurísticas para produzir resultados rapidamente. Ele também calcula um "valor esperado" que estima quantos jogos teriam ocorrido em uma determinada pontuação por chance, o que pode ajudar um usuário a julgar a confiança necessária em um alinhamento. Como o nome indica, o BLAST executa alinhamentos "locais" (MADDEN; 2013).

O BLAST pode ser aplicado na comparação de transcriptomas, como já foi feito por Basante et al. (2013) com transcriptoma de feijão-caupi; por Salgado et al. (2017) com transcriptoma de *Uruchloa decumbens*; por Souza et al. (2016) em *Sorghum bicolor*, *Zea mays* e *teosinto*; e por Wang et al. (2009) em *Raphanus sativus*.

Existem algumas ferramentas da bioinformática feitas para a análise de genes de resistência em plantas, entre elas a RGAugury, que é uma ferramenta de *pipeline* eficiente e integrativa para identificar todos os tipos conhecidos de candidatos RGA a partir de conjuntos de dados em escala de genoma, com uma interface amigável e que seria totalmente integrada do conjunto de dados à sumarização (LI et al.; 2016).

Além dessa, há também a DRPPP, que é uma ferramenta de máquina de aprendizagem para predição de proteínas de resistência a doenças em plantas (PAL et al.; 2016). E a NBSPred, que é uma pipeline de alta produtividade baseada em máquina de vetor de suporte para predição de proteína de resistência de plantas NBSLRR (KUSHWAHA et al.; 2015).

A bioinformática contemporânea engloba uma crescente e ampla gama de atividades relacionadas ao gerenciamento, análise e visualização da

informação biotecnológica. Dentre a plethora de programas e abordagens de bioinformática aplicadas às análises de proteômica, genômica, microarranjos, metabolômica e biologia de sistemas, o desenvolvimento e utilização de banco de dados tem crescido em importância e mostrado sua fundamentalidade não só como essência dos grandes repositórios de dados biológicos como o DDBJ (DNA Data Bank of Japan), EMBL Nucleotide Sequence DB (European Molecular Biology Laboratory) e GenBank (National Center for Biotechnology Information), mas também no cotidiano científico dos laboratórios de pesquisa envolvidos em projetos que empreguem alguma abordagem de sequenciamento de nova geração (MOREIRA; 2015).

Na ciência computacional, banco de dados é definido como uma coleção estruturada de dados. Os dados armazenados em um banco de dados são organizados de forma a permitir agilidade na busca e na recuperação por um computador, ou seja, não há nada além de uma simples coleção de itens. Diferentes tipos de bancos de dados – hierárquico, rede, relacional e orientado para objetos – usam diferentes modelos para organizar os dados. As gravações em bancos de dados hierárquicos, por exemplo, são organizadas em uma estrutura de árvores. Os bancos de dados orientados para objetos armazenam estruturas de dados complexas, chamadas de “objetos”, e que são organizadas em classes hierárquicas que podem herdar propriedades de classes mais altas na cadeia (MANOVICH; 2015).

O banco de dados de genes de resistência de plantas (PRGdb), é o primeiro recurso abrangente de bioinformática dedicado a genes de resistência a doenças de plantas conhecidos e previstos. Este recurso visa proporcionar aos cientistas que trabalham neste campo de pesquisa uma coleção abrangente e atualizada de genes R curados manualmente extraídos da literatura (SANSEVERINO et al.; 2009), que em sua atualização foi expandido mais de 6 vezes para incluir informações biológicas úteis sobre um total de 104 459 genes R (dos quais 90 000 recém- anotados) de 233 espécies de plantas. Destas entradas, 112 são genes R curados manualmente descritos na literatura para conferir resistência a 122 patógenos diferentes (SANSEVERINO et al.; 2012).

2.5 Transcriptômica

O transcriptoma é o conjunto completo de transcrições em uma célula e sua quantidade, para um estágio de desenvolvimento específico ou condição fisiológica. Compreender o transcriptoma é essencial para interpretar os elementos funcionais do genoma e revelar os constituintes moleculares das células e tecidos e também para entender o desenvolvimento e a patologia. Os principais objetivos da transcriptômica são: catalogar todas as espécies de transcrição, incluindo mRNAs, RNA não codificantes e pequenos RNAs; para determinar a estrutura transcricional dos genes, em termos de seus locais de início, extremidades 5' e 3', padrões de empalme e outras modificações pós-transcrição; e para quantificar os níveis de expressão em mudança de cada transcrição durante o desenvolvimento e sob diferentes condições (WANG et al.; 2009).

A molécula de DNA é estática e está presente, com a mesma constituição, em todas as células do organismo. Decifrar esse conteúdo estático de DNA é tarefa da genômica. Já o conteúdo de RNA de uma determinada célula depende do tempo e das condições à qual ela está sendo submetida (ALMEIDA & FERNANDES; 2013). O desenvolvimento de tecnologias aplicadas à biologia e genética molecular vem propiciando uma produção enorme de informações na área genômica, através do sequenciamento de diferentes organismos. A utilização da tecnologia de arranjos de DNA para se analisar simultaneamente a expressão de vários genes têm proporcionado uma gama enorme de informações acerca do perfil da expressão gênica em plantas submetidas a condições de estresses abióticos (VINECKY, et al.; 2005).

Os resultados destas análises, começam a revelar a complexidade dos mecanismos moleculares envolvidos nas respostas dos vegetais aos estresses abióticos. Esses mecanismos moleculares são baseados na ativação e regulação da expressão de vários genes, cujos produtos proteicos podem estar envolvidos tanto na resposta (percepção, tradução do sinal e controle da expressão gênica/fatores de transcrição) como na tolerância ao estresse (antioxidantes, protetores osmóticos, transporte iônico, etc) (VINECKY, et al.; 2005). Através da análise do transcriptoma é possível determinar quando e onde

cada gene é expresso em determinado tecido do organismo e, ainda, é possível quantificar essa expressão (SANGI & BIANCHETTI.; 2017).

O estudo do transcriptoma de cada organismo é de grande importância para a identificação de genes, mas também incorpora informações sobre o funcionamento do seu genoma. Para se estudar o transcriptoma não é necessário sequenciar completamente todos os genes de um tecido ou organismo. Grande parte dos genes podem ser identificados através da análise de pequenas sequências que funcionam como etiquetas (SANTOS & ORTEGA; 2003). Um dos motivadores do uso do genoma funcional para a identificação de genes de interesse é o custo do sequenciamento dos transcritos, que tem decrescido ano a ano, permitindo o sequenciamento de mRNA (RNA-Seq) em larga escala (LANNA et al.; 2013).

Existem vários métodos para a análise em larga escala da expressão de um genoma, classificados em três categorias: baseado em sequenciamento, como por exemplo o sequenciamento *shotgun* de bibliotecas de cDNA (EST/ORESTES), sequenciamento do RNA (RNA-seq) e bibliotecas SAGE; métodos baseados em amplificação, com o por exemplo PCR array; métodos baseados em hibridização, tais como os microarranjos (MOREIRA; 2015).

Entender a resposta de uma planta a um estresse exigirá uma análise global de genes responsivos ao estresse. Uma visão significativa sobre a porção expressa de uma planta com um tamanho grande e genoma desconhecido é através da geração em grande escala e análise de tags de sequência expressa (ESTs). Bibliotecas de cDNA preparadas a partir de diferentes tecidos expostos a várias condições de estresse são ferramentas valiosas para obter a porção expressa do genoma. Isso não apenas leva à identificação de genes que têm funções conhecidas ou putativas, mas também à descoberta de novos genes putativos, até então desconhecidos (LIU et al.; 2013).

Todavia, uma outra tecnologia permite o sequenciamento da região central dos mRNAs. A tecnologia, denominada ORESTES, de Open Reading frames ESTs baseia-se na amplificação de cDNAs por PCR aleatório cujos produtos são utilizados para gerar uma biblioteca. O sequenciamento desta biblioteca, contendo fragmentos aleatórios derivados de diferentes regiões de

cada mRNA, favorece o reconhecimento da função do transcrito por pesquisa de homologia, pois incorpora mais frequentemente a ORF (fase de leitura aberta) no transcrito do que as ESTs convencionais (SANTOS & ORTEGA; 2003).

A técnica SuperSAGE, que representa uma das mais eficientes para análises de transcriptomas. O emprego dessa técnica, na identificação de genes de resistência responsivos proporciona melhor entendimento dos mecanismos envolvidos, com possibilidades de aplicações no melhoramento da espécie (BASANTE et al.; 2013).

A análise de transcriptomas por meio da técnica de RNA-Seq revolucionou os estudos na área, por apresentar várias vantagens quando comparada a outras. O RNA-Seq fornece uma medida mais precisa do nível de transcritos, assim como suas isoformas, e promove uma acurada quantificação da expressão diferencial, capaz de gerar novos conhecimentos sobre mecanismos moleculares de uma dada característica de interesse, em adição à identificação de novos transcritos. O RNA-Seq permite também a identificação sistemática de SNPs em regiões transcritas, os quais podem ser utilizados na detecção de expressão alelo-específica e também como marcadores moleculares (MASSRUHÁ & HIGA; 2014).

Uma execução de RNA-seq pode produzir mais de 200 milhões de leituras curtas, cada uma com 100–150 pares de bases (PERTEA et al.; 2015). Essa é uma metodologia que utiliza o sequenciamento de nova geração na análise de transcriptomas, e tem sido utilizado em estudos da expressão gênica de várias plantas de importância econômica, como o arroz, milho, cevada, laranja doce e cana-de-açúcar, e animais de produção, como bovinos de corte e leite, suínos, aves, caprinos e ovinos (MASSRUHÁ & HIGA; 2014).

O RNA-Seq também tem sido aplicado à diversas gramíneas, como para verificar expressão de genes referentes à tolerância ao alumínio (SALGADO et al.; 2017), tolerância ao calor (LI et al.; 2013), produção de biomassa (TOLEDO-SILVA et al.; 2013), regulação de processos fisiológicos (YUE et al.; 2016).

Existem algumas ferramentas que permitem a análise dos dados de transcriptoma, uma delas é o FASTX-Toolkit, que é uma coleção de ferramentas de linha de comando para o pré-processamento de arquivos Short-Reads

FASTA/FASTQ que são gerados pelas máquinas de sequenciamento de nova geração (GORDON & HANNON; 2009).

Uma outra ferramenta muito utilizada é o FastQC, que visa fornecer uma maneira simples de fazer algumas verificações de controle de qualidade em dados de sequência bruta provenientes de *pipelines* de sequenciamento de alto rendimento. Ele fornece um conjunto modular de análises para dar uma rápida impressão dos problemas presentes nos dados (ANDREWS; 2010).

Há também o programa Trinity, que faz a reconstrução *de novo* de transcriptomas, consistindo de três módulos de software: Inchworm, Chrysalis e Butterfly, aplicados sequencialmente para processar grandes volumes de leituras de RNA-Seq (GRABHERR et al.; 2011).

2.6 Modelagem 3D de proteínas

Proteínas são polímeros que consistem em aminoácidos unidos por ligações peptídicas covalentes. Várias conformações diferentes (estruturas tridimensionais) são possíveis para uma molécula tão grande quanto a proteína. Dessas diversas estruturas, uma ou algumas têm atividade biológica; estas são chamadas conformações nativas. Muitas proteínas não têm estrutura repetitiva regular óbvia. Esta conformação é necessária para seu funcionamento correto, e devido a esta complexidade, as proteínas são definidas em quatro níveis de estrutura (CAMPBELL & FARRELL; 2016).

A estrutura primária é a ordem na qual os aminoácidos são ligados de forma co- valente. A estrutura secundária é o arranjo espacial dos átomos do esqueleto peptídico. A estrutura terciária inclui o arranjo tridimensional de todos os átomos da proteína, inclusive nas cadeias laterais ou em qualquer grupo prostético (grupos de átomos que não são aminoácidos). Uma proteína pode consistir em múltiplas cadeias polipeptídicas chamadas subunidades. O arranjo das subunidades em relação às outras é a estrutura quaternária. A interação entre as subunidades é mediada por interações não covalentes, como ligações de hidrogênio, atrações eletrostáticas e interações hidrofóbicas (CAMPBELL & FARRELL; 2016).

As proteínas também podem ser de membrana, que são associadas a uma matriz de bicamada lipídica. As proteínas de membrana executam os processos dinâmicos associados às membranas e, por isso, proteínas específicas ocorrem somente em determinadas membranas. Elas podem ser: A) integrais - estão firmemente ligadas às membranas por interações hidrofóbicas, podendo também atravessar a membrana (transmembrana), sendo ela unipasso, quando atravessa uma vez a membrana, ou multipasso, quando atravessa mais de uma vez a membrana; B) periféricas - dissociadas das membranas por meio de procedimentos relativamente brandos, que deixam a membrana intacta (VOET & VOET; 2013).

Cadeias polipeptídicas são descritas começando-se pelo terminal amino (conhecido como N-terminal) e designando sequencialmente cada resíduo até o terminal carboxila (o C-terminal). As abreviaturas das cadeias polipeptídicas são sempre escritas de forma que a extremidade N-terminal da cadeia do polipeptídeo esteja para a esquerda e a extremidade C-terminal esteja para a direita. E essa descrição é importante pois quando se altera a sequência de aminoácidos, altera-se também a estrutura da proteína (VOET & VOET; 2013).

A função de uma proteína está intimamente associada à sua estrutura tridimensional. Por isso a modelagem 3D de proteínas ou predição de proteínas é tão importante. Diante disso, atualmente existem 2 formas de modelagem. Métodos independentes de estruturas molde (predição *ab initio* e a predição *de novo*) e métodos baseados em estruturas molde (*threading* e modelagem comparativa). A escolha do método está intrinsecamente relacionada com a taxa de identidade obtida a partir do alinhamento entre a sequência alvo e possíveis candidatos a molde (Figura 3) (VERLI; 2014).

Os métodos ditos *de novo* são aqueles que utilizam algum tipo de informação estrutural, tais como fragmentos de proteínas, predição de estrutura secundária e potenciais estatísticos, oriundas de proteínas não homólogas à sequência alvo. No método de modelagem comparativa, a proteína de interesse terá sua estrutura 3D predita usando como referência a estrutura 3D de outra proteína similar (Figura 4). O método de predição do enovelamento ou *threading* parte da ideia de observações de que a estrutura 3D é mais conservada que a sequência, de forma que mesmo sequências com pouca similaridade podem

possuir estruturas muito semelhantes, o que limita o número de enovelamentos que proteínas podem assumir (VERLI; 2014).

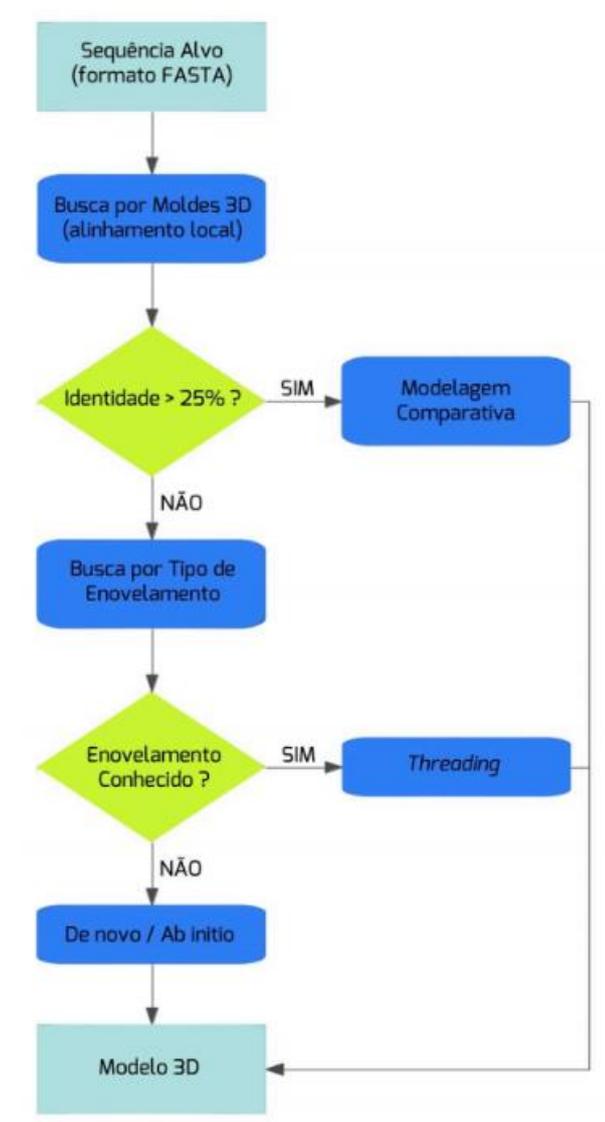


Figura 3. Predição da estrutura tridimensional de uma proteína (VERLI; 2014).

Para avaliar a qualidade das modelagens recomenda-se as seguintes análises: a) estereoquímica que consiste em analisar os aspectos tridimensionais de uma molécula, a fim de se verificar a estabilidade conformacional da mesma; b) energia que são métodos baseados em minimização de funções de energia; e funcional que envolve a comparação do modelo obtido com aspectos funcionais ou mesmo estruturais (sem resolução atomística) determinados por métodos experimentais (VERLI; 2014).

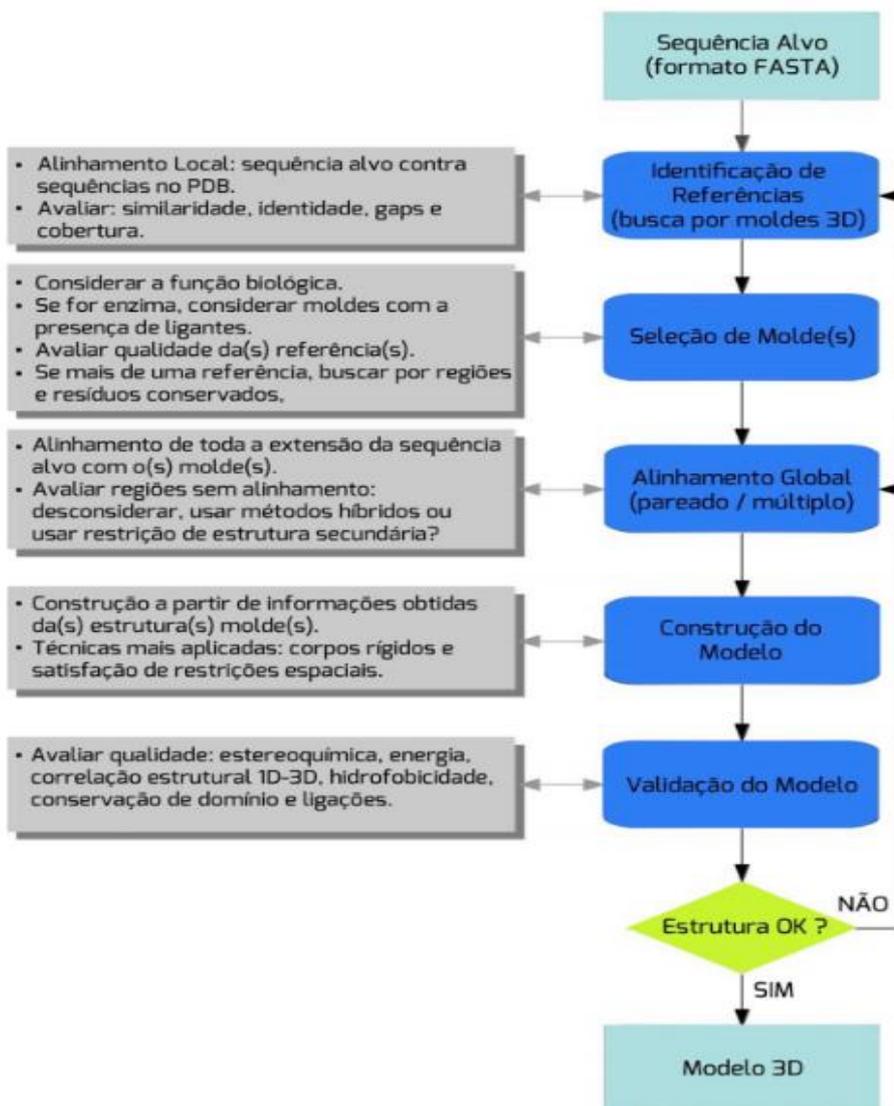


Figura 4. Etapas de predição de estrutura tridimensional de proteínas usando o método de Modelagem Comparativa (VERLI; 2014).

Para a modelagem 3D de proteínas existem algumas ferramentas da bioinformática online entre elas o SWISS-MODEL, que é um servidor para modelagem comparativa automatizada de estruturas proteicas tridimensionais (3D) (SCHWEDE et al.; 2003). Além dessa, há também o Phyre2, que é um conjunto de ferramentas disponíveis na Web para prever e analisar a estrutura, função e mutações da proteína, o foco do Phyre2 é fornecer uma interface simples e intuitiva para as ferramentas de bioinformática de proteínas de última geração. (KELLEY et al.; 2015).

O método central de Phyre2 para gerar um modelo 3D de uma sequência de proteínas é composto por quatro estágios técnicos subjacentes. Há também um modo intensivo opcional que tenta criar um modelo completo de uma sequência através de uma combinação de modelagem de múltiplos modelos e simulação simplificada de dobragem *ab initio*. O primeiro estágio é determinar um perfil evolucionário para a consulta que captura as preferências de resíduos em cada posição ao longo de seu comprimento. O perfil calculado no estágio 1, juntamente com a estrutura secundária prevista, é convertido em um modelo oculto de Markov (HMM) (KELLEY et al.; 2015).

Os indels são manuseados usando uma biblioteca de fragmentos de estruturas proteicas conhecidas, com comprimentos de 2-15 aminoácidos. Quando se usa o modo intensivo, se cria um modelo completo da proteína de consulta, mesmo quando diferentes regiões ou domínios forem modelados por modelos separados, ou quando não houver modelos (modelagem *ab initio*). Por fim, a conexão da cadeia lateral ao carbono central é realizada usando o protocolo R320, que envolve uma técnica rápida baseada em gráficos e uma biblioteca de rotâmeros de cadeia lateral para colocar as cadeias laterais em seu rotamer mais provável, evitando choques estéricos (KELLEY et al.; 2015).

3. MATERIAL E MÉTODOS

3.1 Amostras

Os transcritos únicos de *Panicum maximum* foram obtidos pelo grupo de pesquisa da Dra. Lucimara Chiari com a EMBRAPA Gado de Corte em Campo Grande – MS (Figura 3A), que realizou a extração de RNA das plantas submetidas a stress hídrico.

3.2 Condições de estresse fisiológico

Foram coletadas amostras de folhas em três repetições (Figura 3B), em dois níveis de disponibilidade hídrica: controle com 80% de Volume Total de Poros (VTP) e estresse com 30% de VTP. Os vasos em estresse severo foram mantidos sob estas condições por oito dias. As plantas foram amostradas em três pontos cronológicos, 1 dia, 4 dias e 8 dias após indução do estresse.

Em cada ponto de coleta foram utilizadas plantas que não foram amostradas no ponto anterior, para fins de se descartar uma possível interferência fisiológica/genética causada pela resposta da planta ao corte da folha. Foi utilizada uma planta tolerante à seca e uma sensível.

3.3 Extração de RNA, biblioteca de cDNA e sequenciamento

O RNA foi extraído das folhas (Figura 3C), realizado pelo grupo de pesquisa “Desenvolvimento de novas cultivares forrageiras”. O sequenciamento do RNA foi realizado na Unicamp pelo laboratório LaCTAD. Foi utilizado sequenciamento de nova geração pelo método Illumina HiSeq2500, de acordo com as instruções do fabricante (Illumina, San Diego, CA, EUA) (Figura 3D). Sendo sequenciados 2 genótipos de *Panicum maximum* x 3 tempos de coleta sob estresse e um antes de iniciar o estresse (controle) x 3 repetições.

Resumidamente, isolou-se RNA poli (A) a partir de 2 de RNA total utilizando Esferas Magnicas Oligo (dT) (Illumina). Para evitar viés de iniciação ao sintetizar o cDNA, o mRNA purificado foi primeiro fragmentado em pequenos pedaços (faixa de 180 a 300pb). Em seguida, o cDNA de cadeia dupla foi

sintetizado utilizando o estojo SuperScript Double-Stranded cDNA Synthesis (Invitrogen) com iniciadores de hexâmeros aleatórios (N6) (Illumina). O ADNc sintetizado foi submetido a reparação de fim e fosforilação utilizando polimerase de ADN de T4 (Illumina).

Os fragmentos de cDNA reparados na extremidade foram 3' adenilados utilizando Klenow Exo- (3' a 5' exo menos, Illumina). Os adaptadores de extremidade pareada da Illumina foram ligados às extremidades dos fragmentos de cDNA 3'-adenilados. Para selecionar um intervalo de tamanho de modelo para o enriquecimento a jusante, os produtos da reação de ligação foram selecionados pela técnica de seleção de tamanho AMPure XP, resultando em uma faixa de fragmentos de cDNA de ~ 300pb.

Foram realizados quinze ciclos de amplificação por reação em cadeia da polimerase (PCR) para enriquecer o molde de adaptador de ADNc purificado utilizando o iniciador de PCR PE 1.0 e o PE 2.0 (Illumina) com a polimerase de ADN de fusão. Finalmente, após a validação em um Agilent Technologies 2100 Bioanalyzer usando o kit de chip Agilent DNA 1000 e quantificado com o kit Kappa qPCR (KAPPA), a biblioteca de cDNA foi marcada e sequenciada em uma célula de fluxo PE usando uma Illumina HiSeq 2500.

O fluxo de trabalho foi o seguinte: hibridização molde, amplificação isotérmica, linearização, bloqueio, hibridização de sequenciamento de primers e sequenciamento no sequenciador para 100pb em Read 1. Após a conclusão da primeira leitura, os modelos foram regenerados in situ para permitir um segundo 100 -pb lidos a partir da extremidade oposta dos fragmentos, ie, os novos cordões sequenciados foram removidos, e as cadeias complementares foram amplificadas em ponte para formar aglomerados. Uma vez que os moldes originais foram clivados e removidos, os filamentos reversos foram submetidos a sequenciamento por síntese.

3.4 Controle de qualidade e *assembly de novo*

O total de reads brutos obtidos foi de 814.286.058 sequências. Posteriormente as sequências brutas foram submetidas a limpeza “trimming”

pela ferramenta FASTX-Toolkit (GORDON & HANNON; 2010) (Figura 3E), onde as sequências menores do que 50 nucleotídeos e com qualidade phred menor do que 20 foram removidas na limpeza. Em seguida a qualidade final das sequências foi avaliada no programa FastQC (Figura 3F) (BIOINFORMATICS; 2011). Em seguida o *assembly de novo* foi realizado utilizando o programa Trinity (Figura 3G) (GRABHERR et al.; 2011). Após a limpeza e *assembly de novo* o total de contigs final foi de 61.575 sequências.

3.5 Alinhamento local

Após o *assembly* dos transcritos, foi realizado um alinhamento local, no dia 11 de julho de 2017, dos genes transcritos sequenciados de *Panicum maximum* contra o banco de dados de genes de resistência de plantas PRGdb (<http://prgdb.crg.eu/wiki/MainPage>) que possui 104.459 genes de resistência armazenados (SANSEVERINO; 2012). A ferramenta BLAST (Figura 3H) (ALTSCHUL, et al; 1990) foi utilizada para fazer o alinhamento local entre os genes transcritos sequenciados de *Panicum maximum* contra o banco de dados de genes de resistência de plantas PRGdb (Figura 3I) instalado em computador local. Somente as sequências que apresentaram valor de e-value igual à 0.0 foram usadas nesse estudo, para que somente as sequências com alinhamento perfeito fossem utilizadas, ou seja, com mais confiança. O e-value é um parâmetro usado pela ferramenta de alinhamento local - BLAST – que mostra a probabilidade de um alinhamento ter ocorrido ao acaso.

3.6 Levantamento das classes e domínios de resistência

Os resultados deste alinhamento foram organizados em um banco de dados usando o MySQL (Figura 3J). Em seguida, foi realizado um levantamento dos domínios e classes de resistência, e espécies relacionadas. Os genes de resistência encontrados foram utilizados para modelagem 3D de proteínas.

3.7 Modelagem 3D de proteínas

Para a modelagem 3D de proteínas, foi usada uma sequência escolhida aleatoriamente de cada domínio encontrado. Procurou-se no banco de dados de proteínas PDB (<https://www.rcsb.org/>) (Figura 3K) todas as sequências escolhidas aleatoriamente, e, não tendo nenhuma proteína semelhante, prosseguiu-se a modelagem.

Para escolher qual programa seria utilizado para a modelagem, foi usada a sequência do domínio NBS, comparando assim dois programas: Swiss-model (Figura 3L) e Phyre2 (Figura 3M). O Programa escolhido para a modelagem de todas as proteínas foi o Phyre2. As sequências foram submetidas ao programa online no modo intensivo, obtendo-se resultados com melhores cobertura de sequência e confiabilidade.

O método central de Phyre2 para gerar um modelo 3D de uma sequência de proteínas é composto por quatro estágios técnicos subjacentes. Há também um modo intensivo opcional que tenta criar um modelo completo de uma sequência. através de uma combinação de modelagem de múltiplos modelos e simulação simplificada de dobragem *ab initio*. O primeiro estágio é determinar um perfil evolucionário para a consulta que captura as preferências de resíduos em cada posição ao longo de seu comprimento. O perfil calculado no estágio 1, juntamente com a estrutura secundária prevista, é convertido em um modelo oculto de Markov (HMM) (KELLEY et al.; 2015).

Os indels são manuseados usando uma biblioteca de fragmentos de estruturas proteicas conhecidas, com comprimentos de 2-15 aminoácidos. Quando se usa o modo intensivo, se cria um modelo completo da proteína de consulta, mesmo quando diferentes regiões ou domínios forem modelados por modelos separados, ou quando não houver modelos (modelagem *ab initio*). Por fim, a conexão da cadeia lateral ao carbono central é realizada usando o protocolo R320, que envolve uma técnica rápida baseada em gráficos e uma biblioteca de rotâmeros de cadeia lateral para colocar as cadeias laterais em seu rotamer mais provável, evitando choques estéricos (KELLEY et al.; 2015).

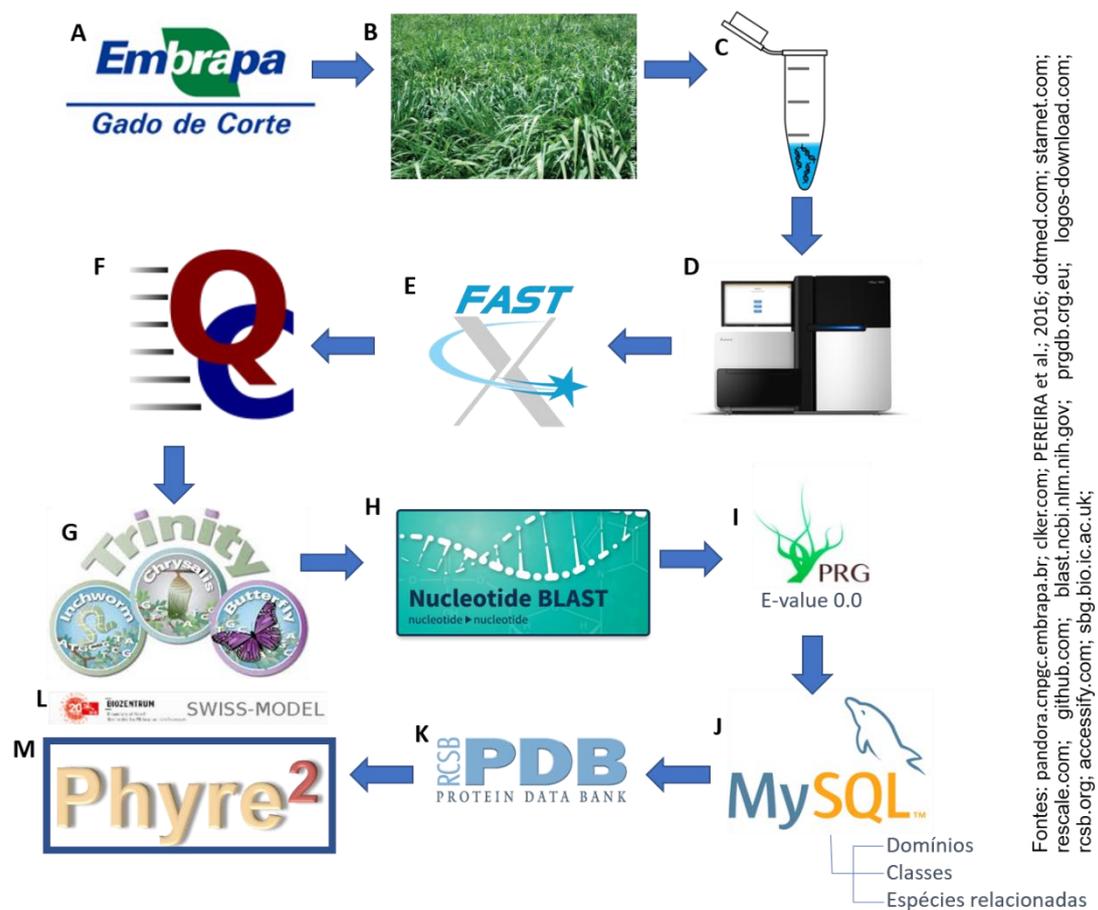


Figura 5. Fluxograma mostrando a sequência de materiais e métodos usados. A – Embrapa Gado de Corte – MS; B – *Panicum maximum*; C – Extração de RNA; D – Sequenciamento Illumina; E – FASTX-Toolkit; F – FastQC; G – Trinity; H – Blastn; I – PRGdb; J – MySQL; K – Protein Data Bank; L – SWISS-Model; M – Phyre2.

4. RESULTADOS E DISCUSSÃO

Dos 61575 genes transcritos e sequenciados de *Panicum maximum*, foi observado uma semelhança em 2115 deles, quando foram contrastados com o banco de dados de genes de resistência em plantas PRGdb (SANSEVERINO; 2009) através do BLAST (ALTSCHUL, et al; 1990). Dos 2115 alinhamentos resultantes, 1592 diferem entre si, sendo estes genes de resistência em plantas, e o restante (523 alinhamentos resultantes) genes expressos repetidamente.

Esse número de genes de resistência encontrado após a planta ter sido exposta a estresse hídrico, é devido ao fato de que os mecanismos de adaptação das plantas submetidas a estresses abióticos (como o estresse hídrico) variam expressivamente em função dos genes envolvidos na rota de resistência porém foi reportado que plantas tolerantes desengatilham naturalmente processos morfológicos, fisiológicos e bioquímicos, como proteínas de resistência, na tentativa de minimizar os efeitos causados pelo estresse (DUARTE et al.; 2013).

Dados empíricos demonstram que o comportamento fenológico das espécies está diretamente associado ao tipo de estratégia em resposta à seca e a forma de utilização da água (SOUZA et al.; 2015). A resposta das plantas ao déficit hídrico está relacionada a uma rede complexa de mudanças morfológicas e fisiológicas (MORALES et al.; 2015).

As alterações morfológicas podem ser de simples aferição, como redução do tamanho da planta e da folha, abscisão foliar e redução da produtividade; mas podem ser também mais complexas, como o aumento do volume do sistema radicular e redução da massa fresca e seca das diferentes partes da planta (parte aérea, frutos e caule). Essas alterações ocorrem, indiretamente, devido a uma série de mudanças fisiológicas, como redução do conteúdo de água na folha, fechamento dos estômatos, deposição de cera na cutícula, redução da fotossíntese e produção de solutos orgânicos que protegem as plantas de estresses por meio de diferentes rotas metabólicas (MORALES et al.; 2015).

4.1 Classes de resistência

Classe é um conjunto de domínios, e a partir dos genes de resistência encontrados, pode-se observar que as plantas de *Panicum maximum* cv. Tamani expressaram 10 diferentes classes de proteínas relacionadas à resistência (Tabela 1). Sendo as classes mais expressivas RLP, NL e CNL.

Tabela 1. Classes de proteínas de resistência presentes em *Panicum maximum*.

Classes	Genes transcritos
Outros	1097
RLP	359
NL	277
CNL	197
N	87
RLK-GNK2	48
CN	36
MLO-Like	12
LRR	1
RLK	1
Total	2115

A classe RLP foi a que mais se destacou, tendo sido encontrada 359 vezes (Tabela 1). Denominada proteína tipo receptor (RLP), possui um domínio transmembrana (TM) com o motif LRR no domínio extracelular e uma cauda C-terminal curta sem motifs evidentes no domínio intracelular (NI et al.; 2014). Descobriu-se que genes que codificam receptores imunes do tipo RLP atuam na defesa de patógenos em várias espécies de plantas. A interação entre

Cladosporium fulvum e tomate que transporta genes de resistência Cf tem sido amplamente explorada para estudar a sinalização de resistência mediada por RLP (ZHANG et al.; 2013).

A classe CNL foi encontrada em 197 sequências (Tabela 1), e compreende genes de resistência que codificam proteínas com pelo menos um domínio de filamento helicoidal, um nucleotídeo local de ligação e uma repetição rica em leucina (CC-NB-LRR) (SOUZA et al.; 2016). Os genes CC-NBS-LRR foram relatados como proteínas receptoras citoplasmáticas, que geralmente reconhecem proteínas efetoras que desencadeiam ETI (DALIO et al.; 2017).

Se a região N-terminal contém uma estrutura de espiral helicoidal prevista (CC), as proteínas NB-LRR n-TIR são coletivamente referidas como CC-NB-LRR ou CNL. Se a região N-terminal não houver essa estrutura, este grupo de proteínas R é coletivamente referido como proteínas NB-LRR ou NL (OSUNA-CRUZ et al.; 2017), que dessa forma foi encontrado 277 vezes (Tabela 1).

4.2 Domínios de resistência

Baseado nos genes de resistência encontrados de *Panicum maximum* foi verificado a presença de 8 diferentes domínios de proteínas relacionadas à resistência, distribuídos em diferentes genes (Tabelas 2), ou combinados a outros domínios em um mesmo gene (Tabelas 3). Sendo os domínios mais expressivos Serina-Treonina-Quinase, NBS e LRR.

Tabela 2. Domínios de resistência simples presentes em *Panicum maximum*.

Domínios	Genes transcritos
Ser-Thr Kinase	841
LRR	144
NBS	96
MLO	12
Receptor-Like Kinase	1
GNK2	0
RPW8	0
TIR	0
Total	1094

Tabela 3. Domínios de resistência conjugados presentes em *Panicum maximum*.

Domínios	Genes transcritos
NBS, LRR, Outros	413
Ser-Thr Kinase, LRR	341
NBS, LRR	59
Ser-Thr Kinase, GNK2	48
NBS, Ser-Thr Kinase	21
Ser-Thr Kinase, NBS	20
LRR, Ser-Thr Kinase	19
NBS, Outros	3
Ser-Thr Kinase, Outros	2
LRR, NBS, Outros	1
MLO, NBS	1
NBS, LRR, RPW8, Outros	1
NBS, Ser-Thr Kinase, Outros	1
NBS, TIR, LRR	1
Ser-Thr Kinase, Receptor-Like Kinase	1
Total	2115

4.2.1 Domínio de resistência Ser-Thr Kinase

O domínio de resistência Serina-Treonina Quinase foi encontrado 1294 vezes, sendo 841 vezes sozinho (Tabela 2), e 453 vezes em conjunto à outros domínios de resistência (Tabela 3), sendo o mais encontrado. As proteínas

quinases são enzimas que catalisam a fosforilação de proteínas através da transferência de um grupo fosforila de ATP e, em casos excepcionais, de GTP, para treonina, serina (quinase específica para Ser/Thr) ou resíduos de tirosina (quinase específica para Tyr). A fosforilação destes resíduos é responsável por estímulos extracelulares e intracelulares, que fornecem um mecanismo altamente eficiente para o controle da atividade de proteínas (SILVA et al.; 2009).

As proteínas quinases são a maior família de proteínas em eucariotos e são a chave central da comunicação no controle intracelular, regulação e transdução de sinais. O mecanismo regulador inclui vários fenômenos que vão desde alterações químicas e estruturais da proteína até o controle transcricional (SILVA et al.; 2009). Os eucariotos superiores utilizam principalmente modificações de Ser / Thr ou Tyr relativamente estáveis, as Ser-Thr quinases eucarióticas (STPK) frequentemente possuem muitos substratos, incluindo quinases adicionais, que atuam para espalhar sinais (BAER et al.; 2014).

As quinases Ser / Thr, juntamente com as suas fosfatases cognatas, desempenham um papel importante na sinalização celular. As enzimas ou processos encontrados para serem regulados por STKP incluem: fosfoglicosamina mutase, GlmM, responsável pela síntese do intermediário inicial da biossíntese da parede celular, N-acetilglucosamina, proteína ribossômica S1 e subunidade alfa da RNA polimerase; Proteína de divisão celular DivIVA; FtsZ, proteína de divisão celular; enzimas da biossíntese da pirimidina da proteína de divisão celular, reparação do DNA, absorção de ferro, resposta ao estresse oxidativo e competência; proteína de ligação à colina A (CbpA); e, RR06/HK06, um dos doze sistemas de transdução de sinal de dois componentes (FALK & WEISBLUM; 2013).

4.2.2 Domínio de resistência LRR

O domínio Repetições Ricas em Leucina (LRR) foi encontrado 979 vezes, sendo 144 vezes sozinho (Tabela 2), e 835 vezes em conjunto à outros domínios (Tabela 3), e é muito variável (GUIMARÃES et al.; 2005). LRRs são domínios estruturais altamente adaptáveis que são dedicados a interações proteína-proteína, e estes podem evoluir especificidades de ligação muito diferentes. Os

LRRs estão sob diversificação da seleção, especialmente no nível dos resíduos expostos a solventes previstos (MARONE et al.; 2013).

A maioria das proteínas codificantes que carregam um motif estrutural com um padrão de repetição de 20 a 30 aminoácidos chamado de repetição rica em leucina (LRR). Os genes R contendo LRR podem ser subdivididos em duas classes amplas - aquelas em que o produto do gene previsto contém uma LRR extracelular N-terminal e uma âncora de membrana, e aquelas em que se prevê que o produto do gene R seja citoplasmático (MILLIGAN et al.; 1998).

Na estrutura do domínio LRR as repetições de leucina supostamente se ajustam e formam uma folha β paralela exposta ao solvente. Como é exposta ao solvente, a superfície hipervariável poderia facilitar a interação da proteína R com o seu fator Avr (elicitor) relacionado e poderia fornecer diferentes especificidades de reconhecimento para fatores Avr alterados. De fato, em vários sistemas planta patógeno, a variação da sequência na região LRR, particularmente na estrutura fita β /volta β , tem mostrado ser responsável por diferentes especificidades de reconhecimento ou de resistência (ALVARENGA et al.; 2009).

A taxa de conversão gênica parece ter sido maior para domínios LRR do que para regiões NBS, que parecem estar sujeitas a seleção purificadora. A alta variabilidade dos domínios LRR está relacionada ao seu papel no reconhecimento de produtos específicos do gene Avr, para promover a resistência das plantas aos patógenos. De fato, razões elevadas de substituições não-sinônimas de nucleotídeos sinônimos foram encontradas em domínios LRR; isto indica que a seleção de diversificação atua nos domínios LRR para manter as variações nos resíduos expostos ao solvente. Para ter uma ideia das variações ao nível dos domínios LRR, existem, em média, 14 LRRs por proteína, e frequentemente cinco a 10 variantes de sequências para cada repetição (MARONE et al.; 2013).

Também há evidências indicando que proteínas da superfície celular contendo um domínio de repetição rico em leucina extracelular (LRR) são os principais organizadores de sinapses excitatórias e inibitórias no SNC. Proteínas LRR distintas são expressas em diferentes tipos de células, e são geralmente

localizadas na membrana pós-sináptica, de onde exercem uma forte influência no desenvolvimento de conexões sinápticas (WIT & GHOSH; 2014).

As proteínas LRR interagem com componentes-chave da maquinaria pós-sináptica e trans-sinapticamente acoplam-se a receptores pré-sinápticos essenciais, o que os coloca em uma posição ideal para coordenar a diferenciação pré e pós-sináptica durante a formação do circuito. Há indícios que a função das proteínas LRR se estende além da formação inicial dos contatos sinápticos e aponta para um papel na regulação das propriedades funcionais e da plasticidade dependente de atividade das sinapses. Estes resultados indicam que as proteínas LRR desempenham um papel crítico na organização e função dos circuitos neurais (WIT & GHOSH; 2014).

4.2.3 Domínio de resistência NBS

O domínio de resistência de Ligação de Nucleotídeos (NBS) foi encontrado 617 vezes, sendo 96 vezes sozinho (Tabela 2), e 521 vezes combinado à outros domínios de resistência (Tabela 3). O maior grupo de genes de resistência de plantas já clonados codifica para proteínas com um domínio de ligação a nucleotídeos (NBS) na região N-terminal, e para uma região com repetições ricas em leucina (LRR) no C-terminal. Genes desta classe conferem resistência a diversos patógenos incluindo vírus, bactérias, fungos e nematóides. Enquanto a região LRR é muito variável, a região NBS apresenta “motivos conservados” (GUIMARÃES et al.; 2005).

Foi demonstrado que o domínio NBS altamente conservado nas proteínas R é capaz de se ligar e hidrolisar ATP e GTP e este domínio também é encontrado em outras proteínas procarióticas e eucarióticas (YANG et al.; 2008). Os domínios NBS estão envolvidos na sinalização, e incluem vários motivos altamente conservados e estritamente ordenados, como os motivos P-loop, kinase-2 e Gly-Leu-Pro-Leu (MARONE et al.; 2013).

Os domínios NBS também são encontrados em muitas famílias de proteínas, incluindo as do grupo RAS, ATPases, fatores de alongação e proteínas G. Essas proteínas servem como chaves moleculares e são importantes para muitos eventos celulares eucarióticos fundamentais, como

crescimento celular, diferenciação, organização do citoesqueleto, transporte de vesículas, apoptose e defesa (ALVARENGA et al.; 2009).

O conjunto de genes NBS-LRR está presente nas proteínas de *Arabidopsis* RPP8 (reconhecimento de *Peronospora parasitica* 1), HRT e RCY1 que são codificadas por diferentes alelos do mesmo locus, e estes conferem resistência a um oomiceto e dois diferentes vírus, respectivamente. Da mesma forma, as proteínas Rx e Gpa2 (*Globodera pallida* 2) da batata são altamente similares, mas reconhecem um vírus e um nematoide, respectivamente (MARONE et al.; 2013). Demonstrou-se que estas proteínas funcionam como receptores imunes intracelulares que reconhecem, direta ou indiretamente, espectros de vetores de patógenos codificados por genes de avirulência (Avr) (YANG et al.; 2008).

4.2.4 Domínio de resistência GNK2

O domínio de resistência Gnk2 foi encontrado 48 vezes combinado à outros domínios (Tabela 3). Ginkbilobina-2 (Gnk2) é uma proteína antifúngica, no endosperma de sementes de Ginkgo. Gnk2 consiste em 108 aminoácidos como uma proteína madura e inibe o crescimento de fungos fitopatogênicos, como *Fusarium oxysporum*. Esta proteína antifúngica não apresenta similaridade com outras antifúngicas. Por outro lado, o Gnk2 possui considerável homologia (85%) em proteínas abundantes em bromo (EAP) das gimnospermas *Picea abies* e *P. glauca*, o que sugere que as proteínas tipo Gnk2 são amplamente conservadas em gimnospermas. As EAPs vegetais são expressas na fase de maturação das sementes e estão envolvidas na proteção contra estresses ambientais, como a seca. No entanto, não há relatos sugerindo que essas proteínas ajudem na defesa contra patógenos fúngicos (MIYAKAWA et al.; 2009).

É relatado que Gnk2 tem a sequência N-terminal. A proteína exerce atividade antifúngica potente e atividade antibacteriana moderada. Além disso, inibe a transcriptase reversa do HIV-1 e a proliferação de esplenócitos murinos (WANG & NG; 2000).

Os homólogos de angiospermas de ginkbilobina partilham características de quinases receptoras, tais como o peptídeo sinal N-terminal, um domínio

extracelular, uma região transmembranar e um domínio C-terminal assinaturas eucarióticas da quinase. A ginkbilobina foi purificada em duas versões - a proteína de comprimento total contém um peptídeo sinal e é denominada ginkbilobina-2, enquanto a primeira ginkbilobina-1 descoberta é um fragmento sem o peptídeo sinal e o domínio DUF26 abrigando uma assinatura característica de cisteína, que é conservada entre todos os homólogos de ginkbilobina de plantas incluindo as quinases semelhantes a receptores de angiospermas (GAO et al.; 2016).

4.2.5 Domínio de resistência MLO

O domínio de resistência de Mutações Induzidas (MLO) foi encontrado 13 vezes, sendo 12 vezes sozinho (Tabela 2), e 1 acompanhado do domínio de resistência NBS (Tabela 3). Este domínio confere resistência à oídio. E também é um domínio particular de genes de proteínas de ligação a CaM (calmodulina) específicos de plantas. O gene MLO foi isolado pela primeira vez em cevada e, desde então, os genes que codificam MLOs foram identificados em várias espécies de plantas, incluindo 15 MLOs em *Arabidopsis thaliana*, sete em trigo e 12 do arroz (SHEN et al.; 2012).

A acumulação de resultados indica que as proteínas MLO têm uma estrutura característica que compreende sete domínios transmembranares (TM) e um domínio de ligação de CaM C-terminal (CaMBD). A ligação de CaM a MLO medeia uma reação de defesa patogênica dependente de Ca^{2+} . Assim, o gene MLO é conhecido como um gene responsivo a Ca^{2+} precoce (SHEN et al.; 2012).

Em cevada, homozigotos recessivos alelos mutantes do gene MLO pode conferir durável e de amplo espectro de agentes patogênicos resistência ao fungo biotrófico do oídio. Outro papel principal do MLO nas plantas é a regulação da morte celular; por exemplo, plantas de MLO mutantes exibem morte celular espontânea de mesofilo associada à senescência foliar parcialmente acelerada. Os MLOs também tem funções adicionais em resposta a estímulos abióticos, como em resposta a ferimentos, tratamento do herbicida paraquat, elicitor de hidratos de carbono oídio derivado do trigo, estresse salino e tratamento de manitol. (SHEN et al.; 2012).

O gene MLO codifica o fundador de uma família de proteínas de membrana integrais específicas de plantas com topologia heptaélica e a atividade bioquímica ainda desconhecida. Os genes MLO estão organizados em famílias de pequeno a médio porte, por espécies de plantas superiores (REINSTÄDLER et al.; 2010).

A proteína MLO da cevada é uma proteína integral localizada na membrana plasmática, possuindo sete hélices hidrofóbicas que atravessam a membrana. O terminal C citoplasmático abriga uma hélice α anfifílica que serve como um domínio de ligação a calmodulina. A ligação de calmodulina dependente de cálcio a este domínio peptídico foi previamente demonstrada tanto in vitro como in vivo e contribui com cerca de metade da atividade conferida pela susceptibilidade de MLO. Evidências preliminares sugerem que a MLO pode formar homo-oligômeros e requer interação de domínios citoplasmáticos para funcionalidade (PANSTRUGA; 2005).

Um gene Ror2 (requerido para a resistência a MLO 2) quando mutado, a resistência mediada por supressor de MLO codifica uma sintaxina residente na membrana plasmática. As sintaxinas pertencem à superfamília das proteínas SNARE (receptor de proteína de ligação à proteína de fusão N-etilmaleimida solúvel) que representam mediadores chave de eventos de fusão de membrana durante o tráfego de vesículas em leveduras e células animais. A falta de ROR2 do tipo selvagem compromete parcialmente a resistência à penetração em genótipos de MLO, sugerindo que a atividade da sintaxina é necessária para a resistência efetiva ao MLO (PANSTRUGA; 2005).

4.2.6 Domínio de resistência Receptor-Like Kinase

Os receptores semelhantes à quinases (RLK) foram encontrados 2 vezes, sendo 1 vez sozinho (Tabela 2), e 1 vez combinado à outro domínio de resistência (Tabela 3). Os RLKs são definidos pela presença de uma sequência sinal, um domínio amino-terminal com uma região transmembranar e um domínio quinase carboxi-terminal (SHIU & BLEECKER; 2001). Embora os papéis para a maioria dos RLKs sejam desconhecidos, muitos estão implicados no crescimento e desenvolvimento das plantas, enquanto outros atuam na defesa

de plantas como os chamados receptores de reconhecimento de padrões (PRRs) (HEESE et al.; 2007).

Os RLKs participam de uma gama diversa de processos, incluindo o reconhecimento próprio e o não-reconhecimento próprio, a resistência a doenças, a regulação do desenvolvimento e a percepção hormonal. Além disso, vários componentes que interagem a montante ou a jusante com as proteínas quinases receptoras também foram identificados. A possibilidade de que as plantas usem centenas de receptores como pontos de entrada em vias de sinalização tem amplas implicações para o desenvolvimento e respostas ambientais (SHIU & BLEECKER; 2001).

Os domínios de quinase de RLKs de plantas pertencem à mesma família de genes que os de quinases de *Drosophila melanogaster* Pelle e de receptor de interleucina de mamífero. Semelhante às contrapartes animais, as funções biológicas dos membros da família de plantas RLK podem ser classificadas em duas grandes categorias. A primeira categoria inclui RLKs que controlam o crescimento e desenvolvimento das plantas, por exemplo, *Arabidopsis thaliana* ERECTA na determinação da forma de órgãos, CLAVATA1 na manutenção do meristema, BRI1 na regulação do crescimento celular, milho CRINKLY4 no controle da morfogênese e diferenciação celular, e PSKR de cenoura (*Daucus carota*) no controle da proliferação celular (SHIU et al.; 2004).

A segunda categoria inclui RLKs envolvidos em interações planta-micróbio e respostas de defesa. Nessa categoria, alguns RLKs estão envolvidos em interações planta-patógeno, como arroz (*Oryza sativa*) Xa21 na resistência ao patógeno bacteriano, *Arabidopsis* FLS2 na percepção da flagelina, e tomate (*Lycopersicon esculentum*) SR160 na sinalização sistêmica, enquanto o restante é crucial para interações com simbioses de plantas, incluindo NORX/SYMRK e HAR1 na simbiose fúngica e/ou bacteriana e RLKs contendo Lys motif envolvido em estágios iniciais de nodulação e percepção do fator Nod (SHIU et al.; 2004).

4.2.7 Domínio de resistência TIR

Receptores Toll-like (TIRs) são receptores codificados na linha germinativa e desempenham um papel importante na imunidade inata de insetos a mamíferos. Este domínio foi encontrado 1 vez, associado à outros domínios (Tabela 3). Os TIRs reconhecem padrões moleculares associados ao patógeno (PAMPs), tais como componentes microbianos. A estimulação de TIRs por ligantes ativa uma cascata de sinalização, que culmina na produção de citocinas pró-inflamatórias e subsequentes respostas imunes. A cascata de sinalização a jusante do TLR é gerada a partir do seu terminal Toll / receptor de interleucina-1 (TIR) (YAMAMOTO et al.; 2004). O domínio de sinalização chave, que é exclusivo do sistema TLR, é o domínio receptor de Toll / interleucina-1 (IL-1) (TIR), que está localizado na face citosólica de cada TLR, e também nos adaptadores (O'NEILL et al.; 2007).

Os TIRs consistem em um domínio extracelular que contém múltiplas repetições ricas em leucina, que se acredita estarem envolvidas na ligação do ligante, e um domínio intracelular necessário para sinalização, também encontrado no receptor da interleucina 1, daí o nome Toll-IL- Domain IR (TIR). Os domínios TIR são de fato módulos proteicos antigos e evolutivos encontrados em proteínas relacionadas à defesa em plantas e animais. Nos animais, estão presentes tanto nos receptores da superfície celular como numa classe de proteínas adaptadoras necessárias para a transdução de sinal. Por exemplo, em moscas, a proteína DmMyD88 do domínio TIR interage com Toll e Toll-9 e, como resultado, está envolvida na ativação de vários alvos transcricionais, incluindo o gene que codifica o peptídeo antifúngico drosomicina (COUILLAUD et al.; 2004).

O domínio TIR é um antigo domínio de interação proteína-proteína que ocorre em um grande grupo de proteínas associadas à defesa do hospedeiro de diversas espécies, incluindo vertebrados, invertebrados e plantas. As proteases contendo domínios TIR no genoma humano compreendem dez membros da família TLR, oito membros do grupo receptor IL-1 / IL-18 (IL-1R / IL-18R) e cinco proteases citosólicas identificadas ou previstas que funcionam como adaptador. proteínas que ligam os receptores Toll-like ou interleucina às vias de sinalização (MEIJER et al.; 2004).

4.2.8 Domínio de resistência RPW8

O domínio de resistência RPW8 foi encontrado 1 vez combinado com outros domínios (Tabela 3). Os produtos previstos dos genes R RPW8 (do inglês Resistance to Powdery Mildew 8) de *Arabidopsis* são pequenas proteínas básicas com um domínio putativo transmembranar N-terminal e um domínio coiled coil. Eles não possuem o sítio de ligação de nucleotídeos e repetições ricas em leucina que caracterizam os produtos dos outros genes de *Arabidopsis* R. O RPW8 confere resistência a todos os isolados testados das quatro espécies de patógenos do oídio de *Arabidopsis*. Em contraste, a maioria dos outros genes R conferem resistência a apenas um ou poucos isolados de uma espécie patogênica que contém os genes Avr correspondentes (XIAO et al.; 2003).

Apesar dessas diferenças, a resistência mediada por RPW8 é caracterizada por uma HR envolvendo a formação de H₂O₂ e a indução de transcritos de PR (XIAO et al.; 2003). Análises anteriores mostraram que a transcrição dos genes RPW8 é induzida a níveis mais altos pelo oídio. A sobre-expressão de RPW8 a partir dos promotores nativos resulta em lesões espontâneas semelhantes à HR (SHL) e indução de genes PR; e os genes RPW8 parecem desencadear HR e resistência através de uma via que envolve os componentes de sinalização SA, PAD4, EDS1, EDS5 e NPR1 que também são recrutados pelos genes NBSLRR contendo um receptor de interleucina 1 - domínio de codificação. Curiosamente, esses componentes de sinalização também estão implicados na resistência basal contra patógenos biotróficos ou hemibiotróficos virulentos (WANG et al.; 2007).

Como os patógenos do oídio causam doenças e como o RPW8 detecta os patógenos e induz resistência em *A. thaliana* não são claros. No entanto, RPW8 parece ativar a defesa através de uma via de sinalização conservada que também é utilizada por um subconjunto de genes NBSLRR (ORGIL et al.; 2007).

Para entender como a função da proteína RPW8 é regulada, um quadro de dois híbridos de levedura (Y2H) foi realizado usando RPW8.2 como isca. Uma particular isoforma 14-3-3 (At5g10450), GF14k, foi identificada. Os 14-3-3s dentro de um organismo constituem um conjunto de diversos membros de uma família de proteínas que é exclusivamente eucariótica. O genoma de *Arabidopsis*

contém 15 genes que codificam as isoformas de 14-3-3, dos quais 13 são expressos (YANG et al.; 2009).

Numerosos estudos têm sugerido que os 14-3-3s funcionam como reguladores de muitos processos celulares importantes, incluindo controle do ciclo celular, apoptose e metabolismo de carbono e nitrogênio. Cada monômero 14-3-3 consiste em nove hélices anti-paralelas e um sulco anfipático ao qual o segmento fosforilado do parceiro de ligação está ligado. Dois monômeros 14-3-3 formam um homo ou heterodímero através das porções do terminal N da proteína, e a estrutura dimerizada é capaz de se ligar simultaneamente a dois alvos ou a duas áreas do mesmo alvo no sulco de ligação criado em cada monômero (YANG et al.; 2009).

Essa ligação geralmente, mas nem sempre, ocorre de maneira dependente da fosforilação. Mais de 100 proteínas foram rigorosamente identificadas como alvos de ligação 14-3-3, incluindo um número de proteínas proeminentes que estão envolvidas em diversas funções celulares. Estes incluem a fosfatase mitótica Cdc25, a proteína pró-apoptótica BAD (um antagonista Bcl-2/Bcl-XL que causa a morte celular), a MAPKK cinase Raf-1 e redutase do nitrato da planta (YANG et al.; 2009).

4.3 Espécies relacionadas

Foi observado similaridade das sequências de resistência entre *Panicum maximum* e 13 diferentes espécies (Tabela 4). Sendo as espécies com mais sequências similares Painço (*Setaria itálica*), Switchgrass (*Panicum virgatum*) e Sorgo (*Sorghum bicolor*). Essa relação pode ser atribuída ao fato de que as 3 espécies mais relacionadas e *Panicum maximum* são gramíneas, e tem algumas características semelhantes, como metabolismo fotossintético C4, alto rendimento de biomassa e fácil adaptação. Também é importante ressaltar que essa proporção de similaridade pode sofrer variação no futuro, visto que o banco de dados biológicos PRGdb pode passar por atualizações ao longo do tempo.

Tabela 4. Espécies relacionadas à *Panicum maximum*.

Espécie	Nº de sequências
<i>Setaria itálica</i>	1245
<i>Panicum virgatum</i>	626
<i>Sorghum bicolor</i>	129
<i>Zea mays</i>	53
<i>Oryza sativa</i> Japonica Group	15
<i>Ricinus communis</i>	14
<i>Oryza sativa</i>	13
<i>Saccharum hybrid</i> cultivar	8
<i>Hordeum vulgare</i>	5
<i>Triticum aestivum</i>	3
<i>Saccharum officinarum</i>	1
<i>Brachypodium distachyon</i>	1
<i>Hordeum vulgare</i> subsp. <i>vulgare</i>	1
<i>Oryza sativa</i> Indica Group	1
Total	2115

Foi observada similaridade entre *Panicum maximum* e *Setaria itálica* 1245 vezes (Tabela 4). Painço (*Setaria itálica*) é um membro da família das gramíneas Poaceae. Essa espécie é distribuída em ecossistemas naturais e agrícolas em todo o mundo, e é cultivada como uma cultura de cereais para humanos alimentos e forragem. Com excelentes características como baixa estatura, curto período de crescimento, autógama, resistência à seca e alta eficiência luminosa, painço é favorecida como uma cultura importante na área tropical semiárida, e

considerado como um ótimo modelo para todos os tipos de pesquisa de plantas, especialmente os estudos da fotossíntese C4. Além disso, o painço exibe rica diversidade genética (6000 variedades), que deve ser útil para caracterização genética e genômica (YOU et al.; 2015).

A espécie *Panicum virgatum* foi encontrada em similaridade com *Panicum maximum* 626 vezes (Tabela 4). O Switchgrass (*Panicum virgatum*) é considerado um dos principais produtos renováveis e culturas de matérias-primas para a produção de biocombustíveis. É uma espécie de gramíneas perenes com alto rendimento de biomassa acima do solo e alta produtividade quando cultivada em terras marginais, que é em parte devido às suas características fisiológicas benéficas, como a produção de raízes profundas e a utilização do metabolismo fotossintético de C4. Além disso, quando comparado com culturas de biocombustíveis de primeira geração, tem se mostrado extremamente eficiente em seu consumo de água e nutrientes com baixos insumos de fertilizantes durante o desenvolvimento (ZHONG et al.; 2015).

A espécie *Sorghum bicolor* foi encontrada em similaridade com *Panicum maximum* 129 vezes (Tabela 4). Sorgo (*Sorghum bicolor*) é uma gramínea diplóide C4 de polinização cruzada, que compete com milho, trigo, arroz e cevada em termos de área cultivada. O sorgo é uma das poucas culturas resilientes que pode adaptar-se bem às futuras condições de mudança do clima, particularmente seca e altas temperaturas (REDDY et al.; 2015).

O sorgo, pertencente à família Poaceae, é amplamente utilizado como forrageira, e é considerado como uma grande planta de energia promissora, devido ao seu crescimento rápido e alta produção de biomassa. Alguns estudos mostraram que *S. bicolor* tem a capacidade de acumular metais pesados, mostrando a maior remoção de Pb pelas folhas e a maior remoção de Cd e Zn pelos caules. O cultivo de plantas de *S. bicolor* facilitado por práticas agrônômicas pode ser uma técnica sustentável para a descontaminação parcial de solos contaminados com metais pesados (AL CHAMI et al.; 2015).

4.4 Modelagem 3D de proteínas

Os resultados das modelagens 3D de proteínas podem ser vistos nas figuras de 5 à 12. Segundo o programa Phyre 2, onde as sequências foram submetidas, sempre que a confiança (probabilidade - de 0 a 100 - de que a correspondência entre sua sequência e esse modelo seja uma verdadeira homologia) do modelo obtido se dá acima de 90%, o valor de RMSD (valor médio para o desvio médio dos átomos de uma estrutura X, relativamente a uma segunda estrutura Y) está entre 2 e 4 Å (Angstrom) que é o recomendável por Kelley et al. (2015). Todas as proteínas estão coloridas em forma de arco-íris, do N-terminal para o C-terminal, sendo que cores mais quentes (vermelho) estão próximas ao N-terminal, e cores mais frias (azul) mais próximas ao C-terminal.

A proteína do domínio NBS (Figura 6A) apresentou 94% de cobertura dos 1292 resíduos, ou seja, 94% dos 1292 resíduos da sequência usada foram modelados, e >90% de confiança. Foi utilizada a sequência [EVGmRNA000350t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função de hidrolase e/ou transporte de proteínas e proteína de membrana. A predição revelou que a proteína tem grande possibilidade de ser transmembrana multipasso (Figura 6B), atravessando 14 vezes a bicamada lipídica tendo sua parte N-terminal e C-terminal posicionada no meio extracelular.

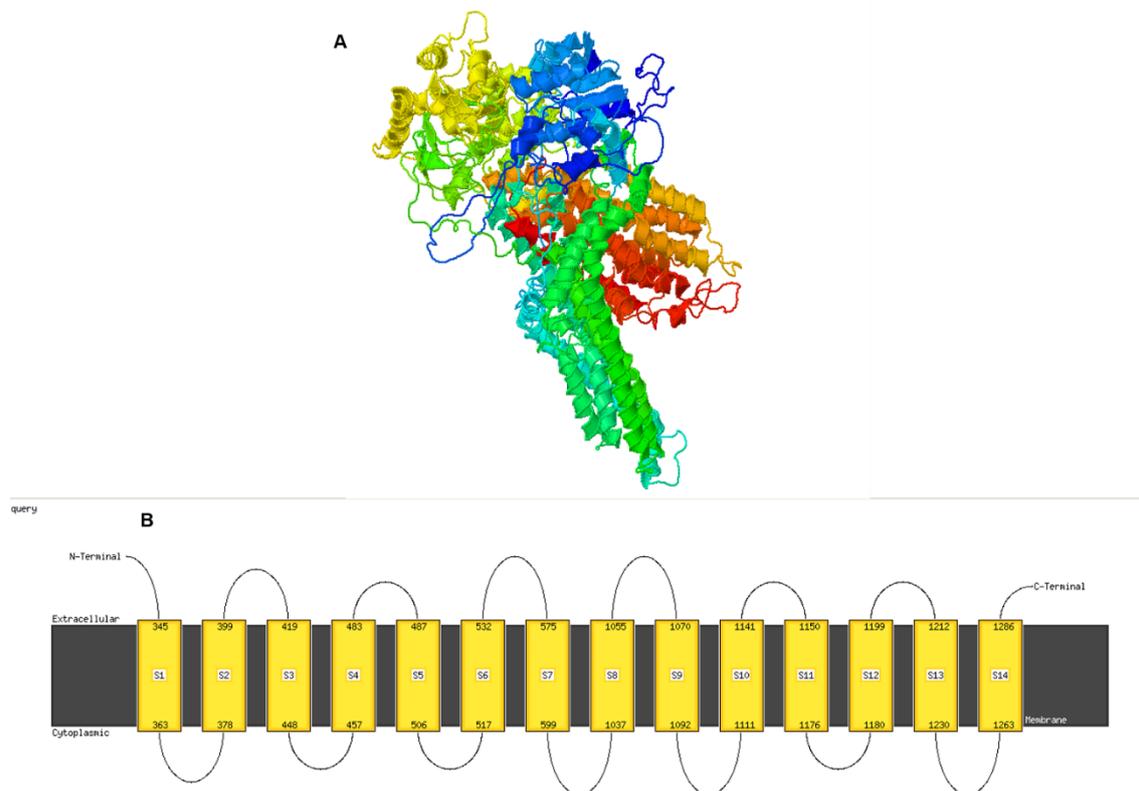


Figura 6. Modelagem 3D da (A) proteína do domínio NBS e seu (B) modelo transmembrana.

A proteína do domínio LRR (Figura 7A) apresentou 91% de cobertura dos 990 resíduos e >90% de confiança. Foi utilizada a sequência [EVGmRNA001068t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função de transferase, inibidor de transferase, sinalização de proteínas, participam do sistema imune, receptor de transferase. A predição revelou que a proteína tem grande possibilidade de ser transmembrana unipasso (Figura 7B), tendo sua parte N-terminal posicionada no meio extracelular, juntamente com o seu peptídeo de sinal, e a parte C-terminal no meio intracelular.

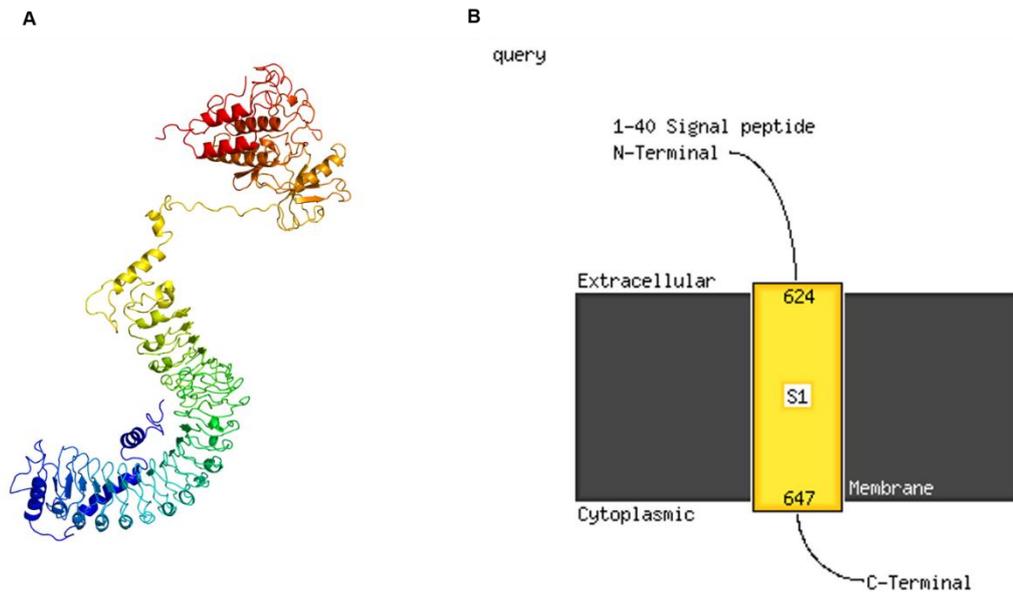


Figura 7. Modelagem 3D da (A) proteína do domínio LRR e seu (B) modelo transmembrana.

A proteína do domínio MLO (Figura 8A) apresentou 27% de cobertura dos 466 resíduos e >90% de confiança. Foi utilizada a sequência [EVGmRNA009140t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função estrutural contrátil e anticongelamento. A predição revelou que a proteína tem grande possibilidade de ser transmembrana multipasso (Figura 8B), atravessando 5 vezes a bicamada lipídica tendo sua parte N-terminal posicionada no meio extracelular, e a parte C-terminal no meio intracelular.

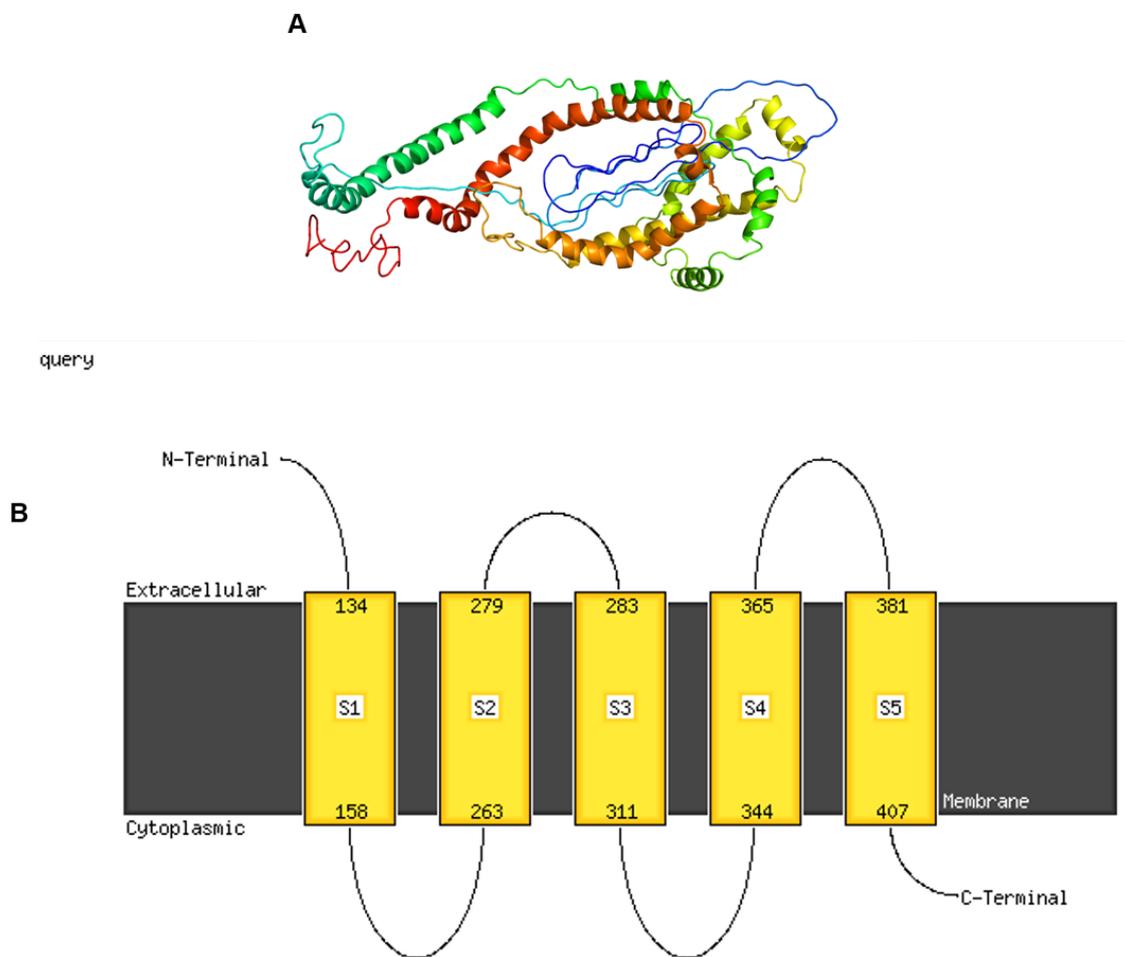


Figura 8. Modelagem 3D da (A) proteína do domínio MLO e seu (B) modelo transmembrana.

A proteína do domínio TIR (Figura 9) apresentou 72% de cobertura dos 1002 resíduos e >90% de confiança. Foi utilizada a sequência [EVGmRNA001018t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função de apoptose, participam do sistema imune, são proteínas de sinalização, proteína de resistência da doença do Linho, transferase, transcrição, proteína motora, proteína de resistência da doença da *Arabidopsis thaliana*. A predição revelou que a proteína não é transmembrana.

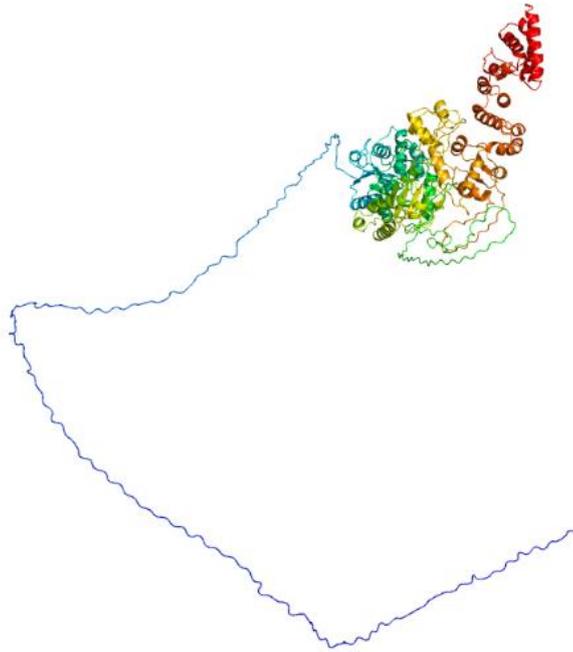


Figura 9. Modelagem 3D da proteína do domínio TIR e seu modelo transmembrana.

A proteína do domínio Receptor-Like Kinase (Figura 10A) apresentou 80% de cobertura dos 584 resíduos e >90% de confiança. Foi utilizada a sequência [EVGmRNA005321t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função de hidrolase, isomerase, receptor de hidrolase e proteína de sinalização. A predição revelou que a proteína tem grande possibilidade de ser transmembrana unipasso (Figura 10B), tendo sua parte N-terminal posicionada no meio extracelular, juntamente com o seu peptídeo de sinal, e a parte C-terminal no meio intracelular.

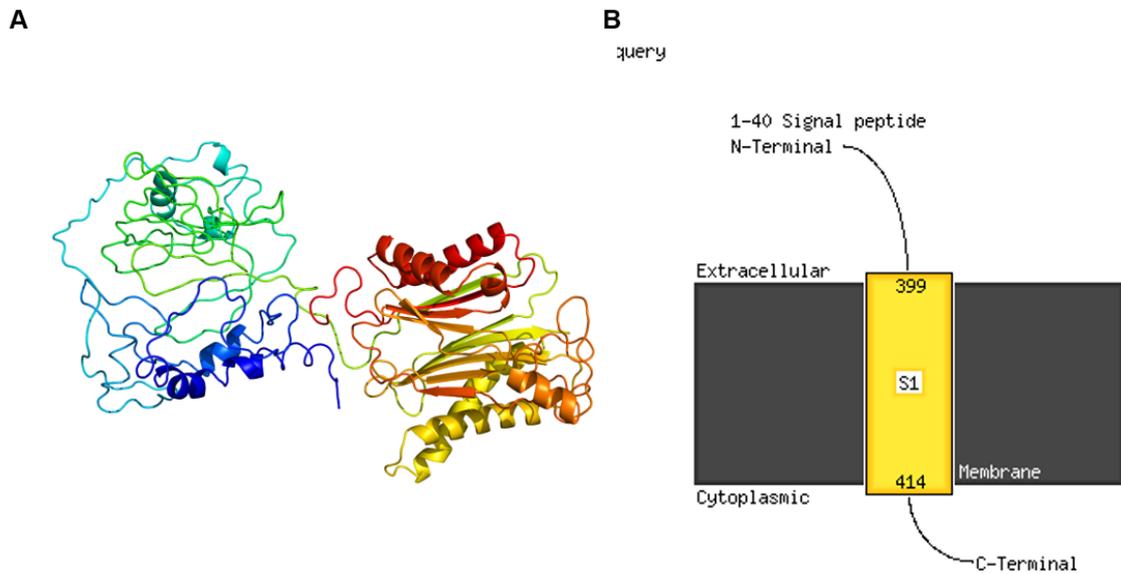
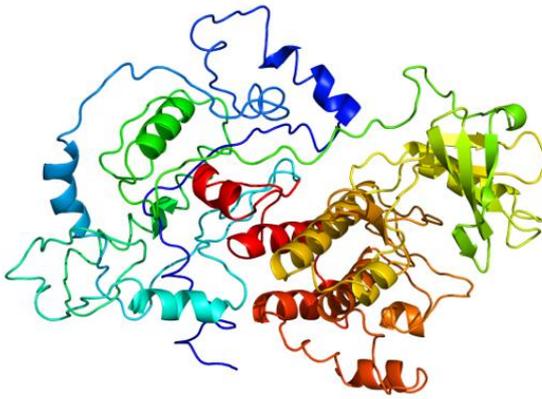


Figura 10. Modelagem 3D da (A) proteína do domínio Receptor-Like Kinase e seu (B) modelo transmembrana.

A proteína do domínio Ser-Thr Kinase (Figura 11A) apresentou 58% de cobertura dos 546 resíduos e >90% de confiança. Foi utilizada a sequência [EVGmRNA006268t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função de transferase e inibidor de transferase. A predição revelou que a proteína tem grande possibilidade de ser transmembrana unipasso (Figura 11B), tendo sua parte N-terminal posicionada no meio intracelular, e a parte C-terminal no meio extracelular.

A**B**

query

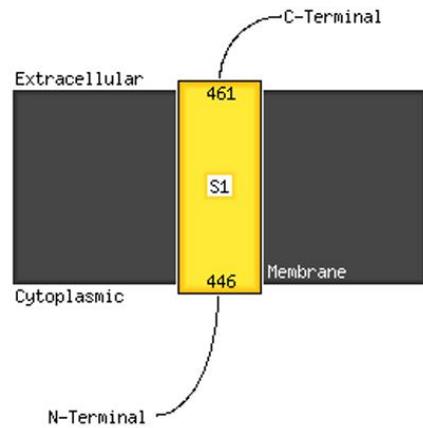
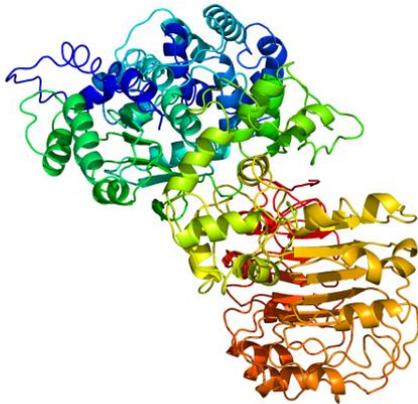


Figura 11. Modelagem 3D da (A) proteína do domínio Ser-Thr Kinase e seu (B) modelo transmembrana.

A proteína do domínio RPW8 (Figura 12A) apresentou 92% de cobertura dos 865 resíduos e >90% de confiança. Foi utilizada a sequência [EVGmRNA001846t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função de apoptose ou participam do sistema imune. A predição revelou que a proteína tem grande possibilidade de ser transmembrana unipasso (Figura 12B), tendo sua parte N-terminal posicionada no meio extracelular, juntamente com o seu peptídeo de sinal, e a parte C-terminal no meio intracelular.

A**B**

query

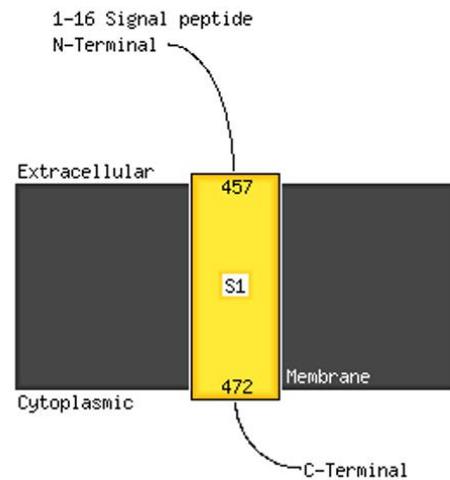


Figura 12. Modelagem 3D da (A) proteína do domínio RPW8 e seu (B) modelo transmembrana.

A proteína do domínio GNK2 (Figura 13A) apresentou 54% de cobertura dos 679 resíduos e >90% de confiança. Foi utilizada a sequência [EVGmRNA004010t1](#) para fazer a modelagem, sendo esta escolhida aleatoriamente. As proteínas usadas para sua modelagem têm função de transferase, inibidor de transferase, participam do sistema imune, proteínas de sinalização. A predição revelou que a proteína tem grande possibilidade de ser transmembrana unipasso (Figura 13B), tendo sua parte N-terminal posicionada no meio extracelular, juntamente com o seu peptídeo de sinal, e a parte C-terminal no meio intracelular.

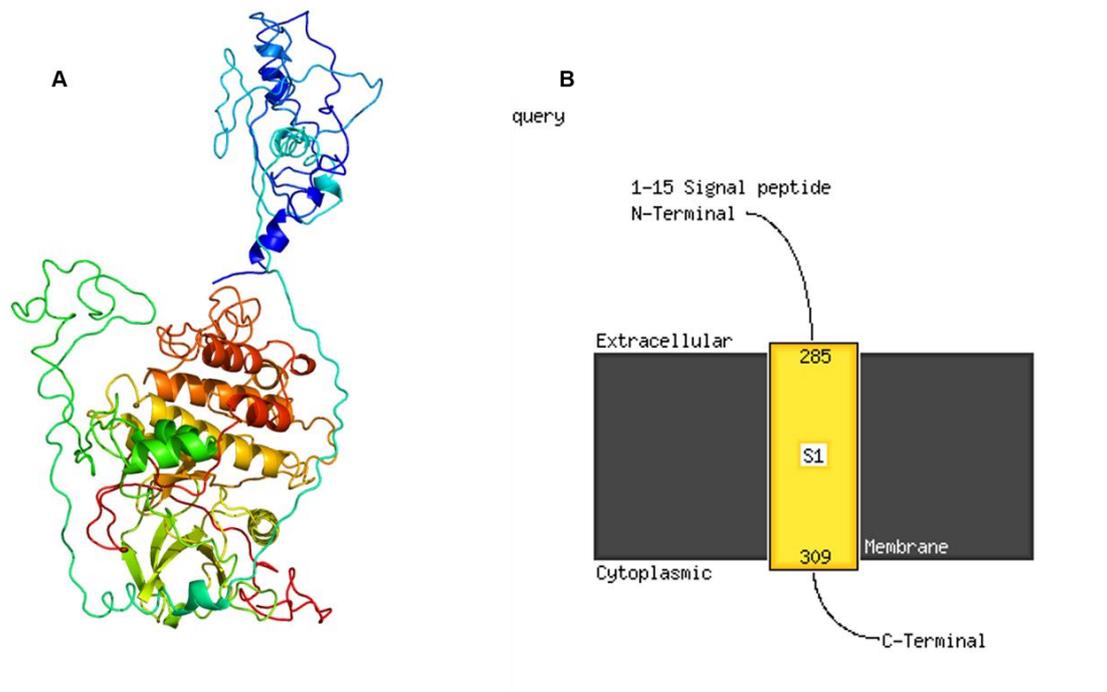


Figura 13. Modelagem 3D da (A) proteína do domínio GNK2 e seu (B) modelo transmembrana.

Pode-se observar que a maioria das proteínas usadas na formação de todos os modelos, tem alguma função de resistência, como de apoptose e sinalização. Mostrando assim que essas proteínas têm grandes chances de serem de resistência.

As proteínas de membrana integrais desempenham um papel fundamental na detecção e transporte de sinais externos para as células, permitindo que elas interajam e respondam ao seu ambiente de maneira específica. Eles formam os principais nós na sinalização hormonal e neuronal. Há importantes grupos de proteínas de membrana, como os receptores acoplados à proteína G (GPCR), aquaporinas, canais iônicos, ATPases (ALMÉN et al.; 2009).

É possível visualizar que 4 das 8 proteínas apresentaram em suas proteínas moldes, função de transferase. Plantas que possuem glutathiona s-transferase apresentam resistência à herbicida (HU; 2014). Algumas também apresentaram função de hidrolase, que podem hidrolisar a parede celular de muitos fungos que contém em sua parede celular β -1,3-glucana e quitina, como maiores componentes (RODRIGUES et al.; 2006).

A modelagem também revelou semelhança entre as proteínas do *Panicum maximum* e proteínas de resistência como: anticongelamento, resistência da doença do Linho, resistência da doença da *Arabidopsis thaliana*. Também revelou que 4 das 8 proteínas possuem peptídeo sinal, sendo estes tipicamente produzidos pela infecção de patógenos ou por enzimas endógenas nas plantas que são sintetizadas em resposta à infecção de patógenos, servindo como sinalização (TRIVILIN; 2008).

5. CONCLUSÕES

Com o alinhamento realizado, foi possível aferir que a planta *Panicum maximum* cv. Tamani expressou, na condição de estresse hídrico, 2115 genes de resistência à doença.

Sendo assim, observou-se expressão de 10 diferentes classes de proteínas relacionadas à resistência à doença, sendo as classes mais expressivas RLP, NL e CNL.

Foi verificado a presença de 8 diferentes domínios de proteínas relacionadas à resistência à doença, sendo os domínios mais expressivos Serina-Treonina-Quinase, NBS e LRR.

Verificou-se similaridade das sequências de resistência entre *Panicum maximum* e 13 diferentes espécies, sendo as espécies com mais sequências similares Painço (*Setaria itálica*), Switchgrass (*Panicum virgatum*) e Sorgo (*Sorghum bicolor*).

A modelagem 3D de proteínas de cada domínio observado possibilitou uma compreensão maior das funções de cada uma das proteínas modeladas, tendo como resultado modelos com um alto padrão de qualidade. Também possibilitou observar semelhança entre as proteínas modeladas e outras proteínas de resistência já descritas na literatura.

6. CONSIDERAÇÕES FINAIS

Esta pesquisa proporcionou informações adicionais e inéditas à comunidade científica sobre genes de resistência à doença no capim *Panicum maximum* e espera-se que essas informações possam ser usadas em trabalhos futuros para o melhoramento da espécie, ou de outras espécies, principalmente em condições de estresse hídrico, como para superexpressar genes de resistência encontrados ou transferência de genes de resistência para outra planta. Sendo útil também para pastagem de bovinos, em condições de estresse hídrico. Cabe ressaltar que esse é o primeiro trabalho realizado envolvendo o transcriptoma desta planta.

7. REFERÊNCIAS

ABLAZOV, A.; TOMBULOGLU, H. Genome-wide identification of the mildew resistance locus O (MLO) gene family in novel cereal model species *Brachypodium distachyon*. **European journal of plant pathology**, v. 145, n. 2, p. 239-253, 2016.

AL CHAMI, Z.; AMER, N.; AL BITAR, L.; & CAVOSKI, I. Potential use of *Sorghum bicolor* and *Carthamus tinctorius* in phytoremediation of nickel, lead and zinc. **International journal of environmental science and technology**, v. 12, n. 12, p. 3957-3970, 2015.

ALMEIDA, W. R.; FERNANDES, F. C. Desenvolvimento de ferramentas de bioinformática para análise de dados de metagenômica e transcriptômica provenientes dos sequenciamentos de rna de anuros. **III Semana de Produção Científica**, p. 47, 2013.

ALMÉN, M. S.; NORDSTRÖM, K. J.; FREDRIKSSON, R.; & SCHIÖTH, H. B. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. **BMC biology**, v. 7, n. 1, p. 50, 2009.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, DJ. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 03-10, 1990.

ALVARENGA, S. M.; CAIXETA, E. T.; ZAMBOLIM, E. M.; ZAMBOLIM, L.; & SAKIYAMA, N. S. Categorização funcional de sequências expressas envolvidas na defesa do cafeeiro a doenças. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 6., 2009. Vitória, ES. **Anais Brasília, D.F: Embrapa - Café, 2011.**

ARAÚJO, N. D.; FARIAS, R. P.; PEREIRA, P. B.; FIGUEIRÊDO, F. M.; MORAIS, A. M. B.; SALDANHA, L. C.; & GABRIELI, J. E. A era da bioinformática: seu potencial e suas implicações para as ciências da saúde. **Estudos de biologia**, v. 30, n. 70/72, p. 143-148, 2017.

BAER, C. E.; IAVARONE, A. T.; ALBER, T.; & SASSETTI, C. M. Biochemical and spatial coincidence in the provisional Ser/Thr protein kinase interaction network of *Mycobacterium tuberculosis*. **Journal of Biological Chemistry**, v. 289, n. 30, p. 20422-20433, 2014.

BALARDIN, R. S.; SILVA, F. D. L.; DEBONA, D.; CORTE, G. D.; FAVERA, D. D.; & TORMEN, N. R. Tratamento de sementes com fungicidas e inseticidas como redutores dos efeitos do estresse hídrico em plantas de soja. **Ciência Rural**, v. 41, n. 7, p. 65-71, 2011.

BASANTE, C.; FERREIRA NETO, J. R. C.; PANDOLFI, V.; SILVA, K.; ROCHA, M. D. M.; NEPOMUCENO, A.; & KIDO, E. Transcriptoma supersage de feijão-caupi sob desidratação radicular. In **Embrapa Meio-Norte-Artigo em anais de congresso (ALICE)**. In: CONGRESSO NACIONAL DE FEIJÃO-CAUPI, 3., 2013, Recife. Feijão-Caupi como alternativa sustentável para os sistemas produtivos familiares e empresariais. Recife: IPA, 2013.

BENTHAM, A.; BURDETT, H.; ANDERSON, P. A.; WILLIAMS, S. J.; & KOBE, B. Animal NLRs provide structural insights into plant NLR function. **Annals of botany**, v. 119, n. 5, p. 827-702, 2016.

BIOINFORMATICS, B.. FastQC: a quality control tool for high throughput sequence data. **Cambridge, UK: Babraham Institute**, 2011.

BONFIM-SILVA, E. D. N. A.; SILVA, T. J. A.; CABRAL, C. E. A.; KROTH, B. E.; & REZENDE, D. Desenvolvimento inicial de gramíneas submetidas ao estresse hídrico. **Revista Caatinga**, v. 24, n. 2, p. 180-186, 2011.

BOTELHO, C. E.; MENDES, A. N. G.; CARVALHO, G. R.; BARTHOLO, G. F.; & CARVALHO, S. P. Seleção de progênies F4 de cafeeiros obtidas pelo cruzamento de Icatu com Catimor. **Ceres**, v. 57, n. 3, p. 274-281, 2015.

BRAGA, G. J.; MACIEL, G. A.; RAMOS, A. K. B.; CARVALHO, M. A.; FERNANDES, F. D.; JUNIOR; R. G.; & JANK, L. Performance of Nellore steers in *Panicum maximum* pastures. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 52., 2015, Belo Horizonte. **Zootecnia: otimizando recursos e potencialidades: anais**. Belo Horizonte: Sociedade Brasileira de Zootecnia, 2015.

BRITO, M. E.; SOARES, L. A.; FERNANDES, P. D.; LIMA, G. S.; SÁ, F. V. S.; & DE MELO, A. S. Comportamento fisiológico de combinações copa/porta-enxerto de citros sob estresse hídrico. **Revista Brasileira de Ciências Agrárias**, v. 7, suplemento, p. 857-865, 2012.

CAMPBELL, M. K.; & FARRELL, S. O. Bioquímica: Tradução da 8ª edição norte-americana. 2ª Edição. São Paulo – SP: **CENGAGE Learning**, 2016.

CATANHO, M. P.; MIRANDA, A. B.; DEGRAVE, W. M. S. Comparando genomas: bancos de dados e ferramentas computacionais para a análise comparativa de genomas procarióticos. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v. 1, n. 2, suplemento, 2007.

CHIARI, L.; BITENCOURT, G. A.; LAURA, V. A.; VALLE, C. B. D.; JANK, L.; & MORO, J. R. Aluminum tolerance on genotypes of signal grass. **Revista Brasileira de Zootecnia**, v. 40, n. 2, p. 245-250, 2011.

CHRISTOPOULOU, M.; MCHALE, L. K.; KOZIK, A.; WO, S. R.; WROBLEWSKI, T.; & MICHELMORE, R. W. Dissection of two complex clusters of resistance genes in lettuce (*Lactuca sativa*). **Molecular Plant-Microbe Interactions**, v. 28, n. 7, p. 751-765, 2015.

COUILLAULT, C.; PUJOL, N.; REBOUL, J.; SABATIER, L.; GUICHOU, J. F.; KOHARA, Y.; & EWBANK, J. J. TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. **Nature immunology**, v. 5, n. 5, p. 488, 2004.

DALIO, R. J.; MAGALHÃES, D. M.; RODRIGUES, C. M.; ARENA, G. D.; OLIVEIRA, T. S.; SOUZA-NETO, R. R.; & PACHECO, I. S. PAMPs, PRRs, effectors and R-genes associated with citrus–pathogen interactions. **Annals of botany**, v. 119, n. 5, p. 749-774, 2017.

DUARTE, E. A. A.; FILHO, P. A. M.; SANTOS, R. C. Características agronômicas e índice de colheita de diferentes genótipos de amendoim submetidos a estresse hídrico. **Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi**, v. 17, n. 8, p. 843-847, 2013.

ESPINDOLA, F. S.; CALÁBRIA, L. K.; REZENDE, A. A. A.; PEREIRA, B. B.; SANTANA, F. A.; AMARAL, I. M. R.; & SANTOS, L. P. Recursos de bioinformática aplicados às ciências ômicas como genômica, transcriptômica, proteômica, interatômica e metabolômica. **Bioscience Journal**, v. 26, n. 3, p. 463-477, 2010.

FALK, S. P.; WEISBLUM, B. Phosphorylation of the *Streptococcus pneumoniae* cell wall biosynthesis enzyme MurC by a eukaryotic-like Ser/Thr kinase. **FEMS microbiology letters**, v. 340, n. 1, p. 19-23, 2013.

FANG, X.; DONG, K.; WANG, X.; LIU, T.; HE, J.; REN, R.; & HUANG, M. A high density genetic map and QTL for agronomic and yield traits in Foxtail millet [*Setaria italica* (L.) P. Beauv.]. **BMC genomics**, v. 17, n. 1, p. 336, 2016.

FILIPPI, M. C. C.; D'ÁVILA, L. S.; GONÇALVES, F. J.; PONTE, E.; ARAÚJO, L. G.; SILVA, G. B.; & SILVA-LOBO, V. L. Incidência de brusone (*Magnaporthe oryzae*) na cultura do arroz. In **Embrapa Arroz e Feijão-Artigo em anais de congresso (ALICE)**. In: CONGRESSO BRASILEIRO DE ARROZ IRRIGADO, 9., 2015, Pelotas. Ciência e tecnologia para otimização da orizicultura: anais. Brasília, DF: Embrapa; Pelotas: Sosbai, 2015.

GAO, N.; WADHWANI, P.; MÜHLHÄUSER, P.; LIU, Q.; RIEMANN, M.; ULRICH, A. S.; & NICK, P. An antifungal protein from Ginkgo biloba binds actin and can trigger cell death. **Protoplasma**, v. 253, n. 4, p. 1159-1174, 2016.

GORDON, A.; HANNON, G. Fastx-toolkit: FASTQ. **A short-reads pre-processing tools. Unpublished http://hannonlab.cshl.edu/fastx_toolkit**, 2010.

GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; & CHEN, Z. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature biotechnology**, v. 29, n. 7, p. 644, 2011.

GUIMARÃES, P. M.; JOSÉ, A. C. F.; PROITE, K.; BERTIOLI, D. J.; & LEAL-BERTIOLI, S. C. Desenvolvimento de Marcadores Moleculares para Análogos a Genes de Resistência em *Arachis spp.* Silvestres. **Fitopatologia Brasileira**, v. 30, n. 6, p. 663-667, 2005.

HEESE, A.; HANN, D. R.; GIMENEZ-IBANEZ, S.; JONES, A. M.; HE, K.; LI, J.; & RATHJEN, J. P. The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. **Proceedings of the National Academy of Sciences**, v. 104, n. 29, p. 12217-12222, 2007.

HU, T. A glutathione s-transferase confers herbicide tolerance in rice. **Crop Breeding and Applied Biotechnology**, v. 14, n. 2, p. 76-81, 2014.

ITO, T. M.; RAMPIM, M. C.; POLIDO, P. B.; & SOUZA, S. G. H. Fatores de transcrição da família ap2/erf e resposta em plantas aos estresses abióticos. **Arquivos de Ciências Veterinárias e Zoologia da UNIPAR**, v. 15, n. 2, p. 207-214, 2013.

JANK, L.; SANTOS, M. F.; VALLE, C. B.; BARRIOS, S.; & RESENDE, R. Novas alternativas de cultivares de forrageiras e melhoramento para a sustentabilidade da pecuária. In **Embrapa Gado de Corte - Resumo em anais de congresso (ALICE)**. In: Simpósio De Adubação E Manejo De Pastagens, 4.:

Simpósio De Produção Animal A Pasto, 4., Dracena. Sustentabilidade do sistema produtivo. São Paulo: Cultura Acadêmica, 2017.

KELLEY, L. A.; MEZULIS, S.; YATES, C. M.; WASS, M. N.; & STERNBERG, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature protocols**, v. 10, n. 6, p. 845, 2015.

KREPS, J. A.; WU, Y.; CHANG, H. S.; ZHU, T.; WANG, X.; & HARPER, J. F. Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. **Plant physiology**, v. 130, n. 4, p. 2129-2141, 2002.

KUSHWAHA, S. K.; CHAUHAN, P.; HEDLUND, K.; & AHRÉN, D. NBSPred: a support vector machine-based high-throughput pipeline for plant resistance protein NBSLRR prediction. **Bioinformatics**, v. 32, n. 8, p. 1223-1225, 2015.

LANNA, A. C.; CARVALHO, M. D. F.; SILVEIRA, R. D. D.; HEINEMANN, A. B.; & BRONDANI, C. Protocolo de deficiência hídrica em arroz de terras altas para análise de transcriptoma. **Embrapa Arroz e Feijão - Comunicado Técnico (INFOTECA-E)**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 2013.

LI, P.; QUAN, X.; JIA, G.; XIAO, J.; CLOUTIER, S.; & YOU, F. M. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. **BMC genomics**, v. 17, n. 1, p. 852, 2016.

LI, Y. F.; WANG, Y.; TANG, Y.; KAKANI, V. G.; & MAHALINGAM, R. Transcriptome analysis of heat stress response in switchgrass (*Panicum virgatum* L.). **BMC plant biology**, v. 13, n. 1, p. 153, 2013.

LIU, M.; SHI, J.; LU, C. Identification of stress-responsive genes in *Ammopiptanthus mongolicus* using ESTs generated from cold-and drought-stressed seedlings. **BMC plant biology**, v. 13, n. 1, p. 88, 2013.

LIU, X.; TANG, S.; JIA, G.; SCHNABLE, J. C.; SU, H.; TANG, C.; & DIAO, X. The C-terminal motif of SiAGO1b is required for the regulation of growth, development and stress responses in foxtail millet (*Setaria italica* (L.) P. Beauv). **Journal of experimental botany**, v. 67, n. 11, p. 3237-3249, 2016.

MADDEN, T. The BLAST sequence analysis tool: 2. ed. **The NCBI Handbook [Internet]**. Bethesda (MD): National Center for Biotechnology Information (US), 2013.

MANOVICH, L. Banco de Dados. **Revista ECO-Pós**, v. 18, n. 1, p. 7-26, 2015.

MARONE, D.; RUSSO, M. A.; LAIDÒ, G.; LEONARDIS, A. M.; & MASTRANGELO, A. M. Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. **International journal of molecular sciences**, v. 14, n. 4, p. 7302-7326, 2013.

MASSRUHÁ, S. M. F. S.; LEITE, M. A. A.; JUNIOR, A. L.; ROMANI, L. A. S. Tecnologias da informação e comunicação e suas relações com a agricultura. **Embrapa Informática Agropecuária - Livro científico (ALICE)**, Brasília, DF: Embrapa, 2014.

MEIJER, A. H.; KRENS, S. G.; RODRIGUEZ, I. A. M.; HE, S.; BITTER, W.; SNAAR-JAGALSKA, B. E.; & SPAINK, H. P. Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish. **Molecular immunology**, v. 40, n. 11, p. 773-783, 2004.

MILLER, R. N. G.; ALVES, G. S. C.; VAN SLUYS, Marie-Anne. Plant immunity: unravelling the complexity of plant responses to biotic stresses. **Annals of botany**, v. 119, n. 5, p. 681-687, 2017.

MILLIGAN, S. B.; BODEAU, J.; YAGHOUBI, J.; KALOSHIAN, I.; ZABEL, P.; & WILLIAMSON, V. M. The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. **The Plant Cell**, v. 10, n. 8, p. 1307-1319, 1998.

MIYAKAWA, T.; MIYAZONO, K. I.; SAWANO, Y.; HATANO, K. I.; & TANOKURA, M. Crystal structure of ginkbilobin-2 with homology to the extracellular domain of plant cysteine-rich receptor-like kinases. **Proteins: Structure, Function, and Bioinformatics**, v. 77, n. 1, p. 247-251, 2009.

MORALES, R. G. F.; RESENDE, L. V.; BORDINI, I. C.; GALVÃO, A. G.; & REZENDE, F. C. Caracterização do tomateiro submetido ao déficit hídrico. **Scientia Agraria**, v. 16, n. 1, p. 09-17, 2015.

MOREIRA, L. M. Ciências genômicas: fundamentos e aplicações. **Ribeirão Preto: Sociedade Brasileira de Genética**, 2015.

NASCIMENTO, J. B.; BARRIGOSSI, J. A. F. O papel das enzimas antioxidantes na defesa das plantas contra insetos herbívoros e fitopatógenos. **Agrarian Academy**, v. 1, n. 1, p. 234-250, 2014.

NEJAT, N.; ROOKES, J.; MANTRI, N. L.; & CAHILL, D. M. Plant–pathogen interactions: toward development of next-generation disease-resistant plants. **Critical reviews in biotechnology**, v. 37, n. 2, p. 229-237, 2017.

NELSON, R.; WIESNER-HANKS, T.; WISSER, R.; & BALINT-KURTI, P. Navigating complexity to breed disease-resistant crops. **Nature Reviews Genetics**, v. 19, n. 1, p. 21, 2018.

NI, X.; YANG, J.; SUN, S.; & YANG, W. Identification and Analysis of Resistance-like Genes in the Tomato Genome. **Journal of Phytopathology**, v. 162, n. 3, p. 137-146, 2014.

NONATO, T.; JANK, L.; SANTOS, M.; CAMPOS, G.; SANTANA, S.; BARRIOS, S.; & SIMEÃO, R. Seleção de híbridos de *Panicum maximum* Jacq. candidatos a lançamento. In **Embrapa Gado de Corte-Resumo em anais de congresso (ALICE)**. In: WORKSHOP MELHORAMENTO VEGETAL: CONTRIBUIÇÕES, AVANÇOS E PERSPECTIVAS PARA O CERRADO BRASILEIRO, 2, Anais. Campo Grande, MS: SBMP, 2016.

O'NEILL, L. A. J.; BOWIE, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. **Nature Reviews Immunology**, v. 7, n. 5, p. 353, 2007.

ORGIL, U.; ARAKI, H.; TANGCHAIBURANA, S.; BERKEY, R.; & XIAO, S. Intraspecific genetic variations, fitness cost and benefit of RPW8, a disease resistance locus in *Arabidopsis thaliana*. **Genetics**, v. 176, n. 4, p. 2317-2333, 2007.

OSUNA-CRUZ, C. M.; PAYTUVI-GALLART, A.; DI DONATO, A.; SUNDESHA, V.; ANDOLFO, G.; AIESE CIGLIANO, R.; & ERCOLANO, M. R. PRGdb 3.0: a comprehensive platform for prediction and analysis of plant disease resistance genes. **Nucleic acids research**, v. 46, n. D1, p. D1197-D1201, 2017.

PAL, T.; JAISWAL, V.; CHAUHAN, R. S. DRPPP: A machine learning based tool for prediction of disease resistance proteins in plants. **Computers in biology and medicine**, v. 78, p. 42-48, 2016.

PANSTRUGA, R. Serpentine plant MLO proteins as entry portals for powdery mildew fungi. **Biochemical Society Transactions**, v. 33, n. 2, p. 389-392, 2005.

PEREIRA, A.; PAICIULLO, D. S. C.; GOMIDE, C. D. M.; & LEDO, F. D. S. Catálogo de forrageiras recomendadas pela Embrapa. **Embrapa Gado de Leite - Livro técnico (INFOTECA-E)**. Brasília, DF: Embrapa, 2016.

PERTEA, M.; PERTEA, G. M.; ANTONESCU, C. M.; CHANG, T. C.; MENDELL, J. T.; & SALZBERG, S. L. StringTie enables improved reconstruction

of a transcriptome from RNA-seq reads. **Nature biotechnology**, v. 33, n. 3, p. 290, 2015.

PETIT-HOUDENOT, Y.; FUDAL, I. Complex interactions between fungal avirulence genes and their corresponding plant resistance genes and consequences for disease resistance management. **Frontiers in plant science**, v. 8, n. 1072, sem paginação, 2017.

QUEIROZ, C. D. A.; FERNANDES, C. D.; VERZIGNASSI, J. R.; Valle, C. B. D.; JANK, L.; MALLMANN, G.; & BATISTA, M. V.. Reaction of accessions and cultivars of *Brachiaria* spp. and *Panicum maximum* to *Pratylenchus brachyurus*. **Summa Phytopathologica**, v. 40, n. 3, p. 226-230, 2014.

REDDY, P. S.; RAO, T. S. R. B.; SHARMA, K. K.; & VADEZ, V. Genome-wide identification and characterization of the aquaporin gene family in *Sorghum bicolor* (L.). **Plant Gene**, v. 1, p. 18-28, 2015.

REINSTÄDLER, A.; MÜLLER, J.; CZEMBOR, J. H.; PIFFANELLI, P.; & PANSTRUGA, R. Novel induced mlo mutant alleles in combination with site-directed mutagenesis reveal functionally important domains in the heptahelical barley Mlo protein. **BMC plant biology**, v. 10, n. 1, p. 31, 2010.

RODRIGUES, A. A. C.; NETO, E. B.; COELHO, R. S. B. Indução de resistência a *Fusarium oxysporum* f. sp. *tracheiphilum* em caupi: eficiência de indutores abióticos e atividade enzimática elicitada. **Fitopatologia Brasileira**, v. 31, n. 5, p. 492-499, 2006.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature protocols**, v. 5, n. 4, p. 725, 2010.

SALGADO, L. R.; LIMA, R.; SANTOS, B. F.; SHIRAKAWA, K. T.; ALMEIDA VILELA, M.; ALMEIDA, N. F.; & CHIARI, L. De novo RNA sequencing and analysis of the transcriptome of signalgrass (*Urochloa decumbens*) roots exposed to aluminum. **Plant Growth Regulation**, v. 83, n. 1, p. 157-170, 2017.

SANGI, S.; BIANCHETTI, R. E. CAPÍTULO XXI Da genômica à bioinformática. In: BOTÂNICA NO INVERNO 2017. **São Paulo: Instituto de Biociências da Universidade de São Paulo, Departamento de Botânica**, 2017, p. 275-283.

SANSEVERINO, W.; HERMOSO, A.; D'ALESSANDRO, R.; VLASOVA, A.; ANDOLFO, G.; FRUSCIANTE, L.; & ERCOLANO, M. R. PRGdb 2.0: towards

a community-based database model for the analysis of R-genes in plants. **Nucleic acids research**, v. 41, n. D1, p. D1167-D1171, 2012.

SANSEVERINO, W.; ROMA, G.; SIMONE, M.; FAINO, L.; MELITO, S.; STUPKA, E.; & ERCOLANO, M. R. PRGdb: a bioinformatics platform for plant resistance gene analysis. **Nucleic acids research**, v. 38, n. 1, p. D814-D821, 2009.

SANTOS, B. F.; VILELA, M. M.; LAURA, V. A.; NEPOMUCENO, A. L.; CHIARI, L. Análise de Transcriptoma de Raízes de plântulas de *urochloa decumbens* Cv. Basilisk (Syn. *Brachiaria decumbens* Cv. Basilisk) em Resposta ao Alumínio. EMBRAPA Gado de Corte. **Resumo apresentado no Workshop de Melhoramento Vegetal de Culturas do Cerrado – UFGD**, 2014.

SANTOS, F. R.; ORTEGA, J. M. Bioinformática aplicada à Genômica. **Melhoramento Genômico, Minas Gerais: UFV**, 2003.

SCHWEDE, T.; KOPP, J.; GUEX, N.; & PEITSCH, M. C. SWISS-MODEL: an automated protein homology-modeling server. **Nucleic acids research**, v. 31, n. 13, p. 3381-3385, 2003.

SHEN, Q.; ZHAO, J.; DU, C.; XIANG, Y.; CAO, J.; & QIN, X. Genome-scale identification of MLO domain-containing genes in soybean (*Glycine max* L. Merr.). **Genes & genetic systems**, v. 87, n. 2, p. 89-98, 2012.

SHIU, S. H.; KARLOWSKI, W. M.; PAN, R.; TZENG, Y. H.; MAYER, K. F.; & LI, W. H. Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. **The plant cell**, v. 16, n. 5, p. 1220-1234, 2004.

SHIU, S.; BLEECKER, A. B. Plant receptor-like kinase gene family: diversity, function, and signaling. **Science's STKE**, v. 2001, n. 113, p. re22-re22, 2001.

SILVA, B. V.; HORTA, B. A.; ALENCASTRO, R. B.; & PINTO, A. C. Proteínas quinases: características estruturais e inibidores químicos. **Química Nova**, v. 32, n. 2, p. 453-462, 2009.

SILVEIRA, J. A.; SILVA, S. L.; SILVA, E. N.; & VIÉGAS, R. A. Manejo da salinidade na agricultura: estudos básicos e aplicados. **Fortaleza, CE: Instituto Nacional de Ciências e Tecnologia em Salinidade**, p. 166-185, 2010.

SILVESTRINI, M.; MALUF, M. P.; GUERREIRO, O.; & GONÇALVES, W. Expressão de genes de resistência em resposta à infecção por nematoides.

Trabalho apresentado no Simpósio de Pesquisa dos Cafés do Brasil (4.: 2005: Londrina, PR). **Anais. Brasília, D.F.: Embrapa Café**, 2005.

SINGH, V. K.; SINGH, A. K.; CHAND, R.; & SINGH, B. D. Genome wide analysis of disease resistance mlo gene family in sorghum [*Sorghum bicolor* (L.) Moench]. **Journal of Plant Genomics**, v. 2, n. 1, p. 18-27, 2012.

SOUZA, B. C.; OLIVEIRA, R. S.; ARAÚJO, F. S.; LIMA, A. L. A.; & RODAL, M. J. N. Divergências funcionais e estratégias de resistência à seca entre espécies decíduas e sempre verdes tropicais. **Rodriguésia - Instituto de Pesquisas Jardim Botânico do Rio de Janeiro**, v. 66, n. 1, p. 021-032, 2015.

SOUZA, R. O.; CRUZ, F. I.; JUNIOR, R. B.; PEREIRA, K. S.; CANDIDO, L. S.; & Pereira, R. M. Análise “in silico” de genes de resistência ortólogos nos genomas de *Sorghum bicolor*, *Zea mays* e teosinto. In: XXXI CONGRESSO NACIONAL DE MILHO E SORGO. **Bento Gonçalves, RS: Associação Brasileira de Milho e Sorgo**, p.47-52, 2016.

SUZUKI, N.; RIVERO, R. M.; SHULAEV, V.; BLUMWALD, E.; & MITTLER, R. Abiotic and biotic stress combinations. **New Phytologist**, v. 203 n. 1, p. 32-43, 2014.

TOLEDO-SILVA, G.; CARDOSO-SILVA, C. B.; JANK, L.; & SOUZA, A. P. De novo transcriptome assembly for the tropical grass *Panicum maximum* Jacq. **PloS one**, v. 8, n. 7, p. e70781, 2013.

TRIVILIN, A. P. Superexpressão do gene codificante do peptídeo AtPep1 em *A. Thaliana* visando a obtenção de resistência à isolados de diferentes espécies do gênero *Pythium*. 108f. **Dissertação de mestrado - Universidade Federal do Rio Grande do Sul, Porto Alegre, RS**, 2008.

VERLI, H. Bioinformática: da biologia à flexibilidade molecular, 1. ed. **Porto Alegre: Universidade Federal do Rio Grande do Sul**, 2014.

VINECKY, F.; BRITO, K. M.; SILVA, F. R.; ANDRADE, A. C. Análise *in silico* de genes potencialmente envolvidos nas respostas aos estresses abióticos presentes na base de dados do genoma do café. Trabalho apresentado no Simpósio de Pesquisa dos Cafés do Brasil (4.: 2005: Londrina, PR). **Anais. Brasília, D.F.: Embrapa Café**, 2005.

VOET, D.; VOET, J. G. Bioquímica. 4ª edição. Porto Alegre – RS: **ARTMED EDITORA LTDA.**, 2013.

WANG, H.; NG, T. B. Ginkbilobin, a novel antifungal protein from Ginkgo biloba seeds with sequence similarity to embryo-abundant protein. **Biochemical and Biophysical Research Communications**, v. 279, n. 2, p. 407-411, 2000.

WANG, W.; DEVOTO, A.; TURNER, J. G.; & XIAO, S. Expression of the membrane-associated resistance protein RPW8 enhances basal defense against biotrophic pathogens. **Molecular plant-microbe interactions**, v. 20, n. 8, p. 966-976, 2007.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews genetics**, v. 10, n. 1, p. 57, 2009.

WIT, J.; GHOSH, A. Control of neural circuit formation by leucine-rich repeat proteins. **Trends in neurosciences**, v. 37, n. 10, p. 539-550, 2014.

XIAO, S.; BROWN, S.; PATRICK, E.; BREARLEY, C.; & TURNER, J. G. Enhanced transcription of the Arabidopsis disease resistance genes RPW8. 1 and RPW8. 2 via a salicylic acid-dependent amplification circuit is required for hypersensitive cell death. **The Plant Cell**, v. 15, n. 1, p. 33-45, 2003.

YAMAMOTO, M.; TAKEDA, K.; AKIRA, S. TIR domain-containing adaptors define the specificity of TLR signaling. **Molecular immunology**, v. 40, n. 12, p. 861-868, 2004.

YANG, S.; ZHANG, X.; YUE, J. X.; TIAN, D.; & CHEN, J. Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. **Molecular Genetics and Genomics**, v. 280, n. 3, p. 187-198, 2008.

YANG, X.; WANG, W.; COLEMAN, M.; ORGIL, U.; FENG, J.; MA, X., & XIAO, S. Arabidopsis 14-3-3 lambda is a positive regulator of RPW8-mediated disease resistance. **The Plant Journal**, v. 60, n. 3, p. 539-550, 2009.

YOU, Q.; ZHANG, L.; YI, X.; ZHANG, Z.; XU, W.; & SU, Z. SIFGD: *Setaria italica* functional genomics database. **Molecular plant**, v. 8, n. 6, p. 967-970, 2015.

YUE, H.; WANG, M.; LIU, S.; DU, X.; SONG, W.; & NIE, X. Transcriptome-wide identification and expression profiles of the WRKY transcription factor family in Broomcorn millet (*Panicum miliaceum* L.). **BMC genomics**, v. 17, n. 1, p. 343, 2016.

ZANELA, M. B.; DERETI, R. M. Boas Práticas Agropecuárias na Produção de Leite: da Pesquisa para o Produtor, 1. ed. Embrapa Clima Temperado - Documentos (INFOTECA-E). **Pelotas, RS: Embrapa Clima Temperado**, 2017.

ZHANG, Z.; FRADIN, E.; DE JONGE, R.; VAN ESSE, H. P.; SMIT, P.; LIU, C. M.; & THOMMA, B. P. Optimized agroinfiltration and virus-induced gene silencing to study Ve1-mediated *Verticillium* resistance in tobacco. **Molecular Plant-Microbe Interactions**, v. 26, n. 2, p. 182-190, 2013.

ZHONG, R.; YUAN, Y.; SPIEKERMAN, J. J.; GULEY, J. T.; EGBOSIUBA, J. C.; & YE, Z. H. Functional characterization of NAC and MYB transcription factors involved in regulation of biomass production in switchgrass (*Panicum virgatum*). **PloS one**, v. 10, n. 8, p. e0134611, 2015.