

Análise “*in silico*” de genes de resistência ortólogos nos genomas de *Sorghum bicolor*, *Zea mays* e teosinto.

Ronaldo Omizolo de Souza⁽¹⁾; Isabella da Cruz Franco⁽¹⁾; Ramir Bavaresco Junior⁽¹⁾; Kamila Savala Pereira⁽¹⁾; Liliam Silvia Candido⁽²⁾; Rodrigo Matheus Pereira⁽²⁾.

⁽¹⁾ Graduando(a) em Biotecnologia; Universidade Federal da Grande Dourados; Dourados, MS; *ronaldo_omizolo@hotmail.com; ⁽²⁾ Professor(a) Adjunto; Universidade Federal da Grande Dourados, Dourados, MS;

RESUMO: Durante a domesticação da cultura do milho (*Zea mays* L.) pode ter ocorrido a transferência de genes de resistência (R) das espécies ancestrais e entre espécies aparentadas. Os genes R podem ser classificados em cinco classes, a CNL, TNL, RLP, RLK e outros. O objetivo deste trabalho foi identificar, *in silico*, genes R comuns entre a espécie *Zea mays* subsp. *mays* com *Zea mays* subsp. *parviglumis*, *Teosinto* e *Sorghum bicolor*. As sequências foram coletadas de bancos de dados biológicos e a análise foi realizada com o auxílio de ferramentas da bioinformática. No entanto, não foi possível encontrar o transcriptoma completo de todas as espécies nos bancos de dados biológicos consultados. Somente os transcriptomas do *Sorghum bicolor* e *Zea mays* subsp. *mays* foram obtidos integralmente. Para as outras duas espécies, foram utilizadas informações parciais encontradas no GenBank e no European Nucleotide Archive (ENA). Foi possível a identificação de 1085 genes R comuns entre o transcriptoma de *Zea mays* subsp. *mays* e as sequências utilizadas de cada uma das espécies. A comparação com o *Sorghum bicolor* retornou 930 genes R divididos em todas as classes. A comparação entre *Zea mays* subsp. *mays* e *Zea mays* subsp. *parviglumis*, possível ancestral do milho, totalizou 109 genes R comuns, enquanto que a análise com o *Teosinto* retornou 44 genes R. Também foram encontrados dois genes ortólogos entre todas as espécies. Assim, os resultados obtidos neste estudo poderão servir como base para futuros estudos sobre o compartilhamento de genes de resistência entre estas espécies.

Termos de indexação: Bioinformática, Gramíneas, Genômica, Milho, Sorgo, Teosinto.

INTRODUÇÃO

Há mais de 8000 anos o milho é cultivado em diversas partes do mundo e graças a sua adaptabilidade, representada por vários genótipos, pode ser cultivado desde o Equador até o limite das terras temperadas. Possui importância econômica, devido sua utilização para alimentação humana e animal (Barros & Calado, 2014).

O sorgo (*Sorghum bicolor*) é uma *Poacea* originária da África, utilizado principalmente na alimentação animal, na cadeia produtiva da suinocultura e avicultura. Também pode ser utilizado para a fabricação de bebidas alcoólicas. (Purcino, 2011).

O sorgo é um parente próximo do milho (*Zea mays*) e da cana-de-açúcar (*Saccharum spp*) (Springer et al., 1989). Entre estes, o milho é o melhor caracterizado geneticamente, possuindo um mapa genético bem desenvolvido (Beavis & Grant 1991; Burr et al., 1988). O sorgo e o milho apresentam o mesmo número de cromossomos (n=10). Entretanto o milho contém, no mínimo, três vezes mais DNA nuclear do que o sorgo (Michaelson et al., 1991).

O milho é uma planta cultivada de grande versatilidade. Os programas de melhoramento genético vêm desenvolvendo ao longo das décadas tipos diversificados de cultivares, entre outras razões, devido a existência de alta variabilidade genética do germoplasma da espécie. Com o advento da biotecnologia foram desenvolvidas tecnologias que permitem acesso a novas e variadas fontes de variabilidade genética, além das existentes naturalmente. Em especial, é possível citar o aprimoramento da tecnologia do DNA recombinante e da bioinformática (Coulter, 2014; Oliveira, 2010).

Nos últimos 10 anos, o mapeamento comparativo entre espécies tem fornecido evidências para uma conservação importante de marcadores e a ordem de genes entre genomas similares, o que resultou no surgimento das técnicas

de comparação de genomas. Estudos comparativos realizados com gramíneas demonstraram consistência em relação à colinearidade entre os genomas dentro da família (Feuillet & Keller, 2002). A maioria destas análises foi realizada com gramíneas de importância econômica, como arroz, trigo, cevada, milho, milheto, aveia e o sorgo (Gale & Devos, 1998; Bennetzen, 2000; Keller & Feuillet, 2000).

A utilização de recursos aplicados à seleção de sequências de DNA unidos as técnicas de bioinformática tornaram-se um pré-requisito para construção de modelos teóricos e experimentais. (Saeys et al., 2007).

Alguns estudos indicam que algumas variedades de teosinto, possível ancestral de *Zea mays*, são citologicamente indistinguíveis do milho e, portanto, são capazes de formar híbridos completamente férteis com o milho (Doebley, 2004).

Vários trabalhos têm discutido e pesquisado as diferenças entre estas espécies, alguns com análises moleculares identificaram uma variação de teosinto (*Zea mays* ssp. *parviglumis*) como o progenitor do milho (Doebley, 2004; Barros & Calado, 2014).

As plantas expressam genes de resistência (R) para reconhecer invasores e evitar a disseminação de agentes patogênicos. As plantas que não têm células imunológicas especializadas devem ser capazes de reconhecer de forma autônoma possíveis proteínas efetoras. Em resposta a proteínas efetoras liberadas por patógenos, as plantas necessitam manter um grande número de genes R que, direta ou indiretamente reconhecem os efetores e iniciam uma resposta efetiva de disparo de imunidade (ETI) (McDowell & Simon, 2008).

Os genes de resistência têm um importante papel para o reconhecimento de proteínas específicas, que são expressas por genes de avirulência (*Avr*) durante um possível ataque de micro-organismos ou pragas, e podem ser agrupados em cinco distintas classes funcionais, com base na presença de domínios específicos (Bent, 1996; Van Ooijen et al., 2007; Kim et al., 2012): a classe CNL 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); a classe TNL inclui aqueles que apresentam receptor com um domínio do tipo Toll-interleucina, um sítio de ligação de nucleotídeos e uma repetição rica em leucina (TIR- NB-LRR); a classe RLP, sigla de proteína do tipo receptora, grupo este, que contém um receptor com um domínio do tipo serina-treonina-quinase, e uma repetição extracelular rica em leucina (Ser/Tre -

LRR); a classe RLK contém aqueles com um domínio de cinase, e uma repetição extracelular rica em leucina extracelular (Kin-LRR); a classe “Outros” inclui todos os outros genes que têm sido descritos como conferindo resistência através de mecanismos moleculares diferentes, por exemplo Mlo e ASC-1. O grupo de genes R “Unknown” representa aqueles genes cuja função exata ou tipo ainda não foram estabelecidos (Sanseverino et al., 2012; Hammond-Kosack & Jones, 1997; Chisholm et al., 2006).

Devido a forte relação entre a função dos genes R e sua complexidade no genoma das plantas, uma comparação de conteúdos de genes R entre diferentes espécies, por meio de ferramentas da bioinformática, poderá fornecer informações úteis para aplicações práticas e importantes descobertas quanto à presença destes genes em diferentes espécies (Kim et al., 2012).

Levando este fato em consideração, esse trabalho foi realizado com o objetivo de verificar, *in silico*, a presença de genes de resistência comuns entre *Zea mays* subsp. *mays* em relação a *Zea mays* subsp. *Parviglumis*, *Teosinto* e *Sorghum bicolor*.

MATERIAL E MÉTODOS

Para a execução deste trabalho foram utilizadas as informações dos seguintes genomas: *Sorghum bicolor*, *Teosinto* e *Zea mays* subsp. *pavirglumis*. Sendo que os genomas dessas espécies foram comparadas com *Zea mays* subsp. *mays* na busca de genes de resistências comuns. As comparações realizadas foram feitas uma a uma, ou seja, os genes R identificados em *Zea mays* subsp. *mays* foram pesquisados entre os genes de resistência encontrados em cada uma das espécies comparadas.

As sequências de dados biológicos, no formato FASTA, foram obtidas de diversas fontes, sendo os transcriptomas do *Sorghum bicolor* e *Zea mays* subsp. *mays*, obtidos do banco de dados Phytozome V11.0 (<https://phytozome.jgi.doe.gov>); sequências de *Zea mays* subsp. *pavirglumis*, obtidas do European Nucleotide Archive (ENA) (<http://www.ebi.ac.uk/ena>) e sequências de *Teosinto*, obtidos do GenBank (<http://www.ncbi.nlm.nih.gov/nuccore>) o banco de dados de nucleotídeos do NCBI (National Center for Biotechnology Information).

As sequências de dados biológicos das espécies foram alinhadas contra sequências FASTAS do banco de dados biológicos PRG-Wiki (<http://prgdb.crg.eu>), um banco de dados de genes de resistência de plantas, utilizando o BLAST (Basic Local Alignment Search Tool), uma ferramenta de

alinhamento local de sequências. Para realizar o alinhamento entre as sequências similares foi utilizado o parâmetro de nota de corte para e-value de pelo menos $1e-5$, sendo o e-value a probabilidade do alinhamento ter ocorrido ao acaso. O alinhamento resultou em um conjunto de informações tabuladas.

Em seguida, criou-se um banco de dados local, utilizando o MySQL, uma ferramenta de gerenciamento de banco de dados gratuita. Foi adicionada uma tabela com informações anotadas sobre os genes R de plantas, obtida do PRG-Wiki para consecução de anotação dos dados alinhados.

Foram criadas novas tabelas no banco de dados com genes ortólogos entre *Zea mays* subsp. *mays* e as demais espécies utilizadas neste trabalho. Na sequência, informações provenientes da tabela Annotation (tabela que contém informações anotadas sobre os genes de resistência) foram adicionadas às tabelas de genes ortólogos.

RESULTADOS E DISCUSSÃO

A partir dos resultados obtidos entre os alinhamentos realizados de cada uma das espécies com as informações do BLAST de *Zea mays* subsp. *mays*, verificou-se a necessidade de filtrar as informações repetidas nos arquivos. Na **tabela 1** está o comparativo entre o número de sequências de entrada, o número de resultados obtidos demonstrando sequências de genes de resistência similares, bem como o resultado da comparação entre as demais espécies com *Zea mays* subsp. *mays*.

A partir disto, foi possível observar o número final de resultados obtidos após a filtragem de sequências repetidas que foi realizada. Durante a busca pelas informações biológicas do milho e demais espécies nos diferentes bancos de dados, não foi possível encontrar o transcriptoma completo de todas as espécies, informação especialmente importante para que as comparações fossem realizadas. Entretanto, foi possível utilizar os transcriptomas completos de *Sorghum bicolor* e *Zea mays* subsp. *mays*. Para as outras espécies utilizadas na comparação foram utilizadas todas as informações de sequências biológicas encontradas a partir dos bancos de dados biológicos GenBank (<http://www.ncbi.nlm.nih.gov/nuccore>) e do ENA (<http://www.ebi.ac.uk/ena>).

Tabela 1. Distribuição do número de sequências obtidas em bancos de dados biológicos para a realização das análises *in silico*.

Espécies e	Nº de	Nº após	Nº	Nº
------------	-------	---------	----	----

(Banco de dados)	sequências de entrada	BLAST vs Genes R	sequências as sem repetição	sequências filtradas vs <i>Z. mays</i> subsp. <i>mays</i>
<i>Z. mays</i> subsp. <i>mays</i> (Phytozome)	88760	10951	3032	-
<i>S. bicolor</i> (Phytozome)	47205	10086	2581	930
<i>Teosinto</i> (GenBank)	21395	1994	64	44
<i>Z. mays</i> subsp. <i>pavirglumis</i> (ENA)	23139	3226	145	109
TOTAL	180499	26257	5822	1085

Apesar destas limitações foram encontrados os resultados, apresentados na **tabela 1**. Conforme as análises foram sendo executadas foi possível filtrar as informações para a conferência de quais e quantos genes de resistência estão similarmente compartilhados entre *Zea mays* subsp. *mays* e as outras espécies. A **tabela 2** apresenta os resultados encontrados quando a comparação entre *Zea mays* subsp. *mays* e as outras espécies foi realizada. Cada um dos genes de resistência similares foi dividido nas diferentes classificações. O número total de genes R somando todas classificações encontradas na comparação também foi obtido.

Foram encontrados 1085 genes de resistência compartilhados entre o transcriptoma de *Zea mays* subsp. *mays* e as sequências utilizadas das outras espécies estudadas (**tabela 1**). Os resultados de comparação mais expressivos foram os encontrados com *Sorghum bicolor*, que mostrou ter genes R de todas as classes compartilhados com *Zea mays* subsp. *mays*, num total de 930 genes. A comparação entre *Zea mays* subsp. *mays* e *Zea mays* subsp. *pavirglumis*, que é considerado o ancestral do milho atual, resultou em 109 genes R compartilhados, pertencentes à todas as classes, excluindo as do tipo Mlo-like e T. Para o *Teosinto* também foram encontrados a presença de 44 genes R ortólogos, podendo ser resultado de uma possível ancestralidade. A comparação realizada entre as quatro espécies demonstrou a presença de 2 genes R em comum, ortólogos, porém ambos de classe ainda desconhecida.

Para todas as plantas estudadas, foi observado o prevaletamento para presença dos genes R das classes CNL, N, NL e RLP entre as sequências das plantas comparadas com *Zea mays* subsp. *mays*. Ambas as classes TNL e CNL visam especificamente proteínas efetoras patogênicas no interior da célula hospedeira, o que inicia a resposta imune desencadeada (ETI) (Hammond-Kosack e Jones, 1997). Já as classes RLP e RLK são receptores padrões de reconhecimento do patógeno/micro-organismo que desencadeia a imunidade para permitir o reconhecimento de uma

ampla gama de agentes patogênicos, como vírus e bactérias (Chisholm et al., 2006). Enquanto que na classe “Outros” estão os genes responsáveis por facilitar a comunicação entre os meios intracelular e extracelular, embora não contenham domínios completos do tipo RLP, NBS ou RLK. Por isso o termo “Outros” ou oth-R foi proposto por Sanseverino et al. (2012), e é utilizado para classificar estes tipos peculiares de análogos de genes de resistência.

CONCLUSÕES

Foi possível encontrar dois genes de resistência ortólogos a todas as plantas estudadas, porém estes genes ainda não possuem informações completas com relação à sua funcionalidade, além disso, algumas das espécies estudadas continham poucas sequências, ou não continham os transcriptomas disponíveis nos bancos de dados biológicos.

A anotação manual, realizada *in silico*, dos genes de resistência encontrados possibilitou a observação de genes de resistência ortólogos a todas as plantas. Isso poderá servir como guia para se iniciar um processo de validação em laboratório sobre as espécies avaliadas no trabalho.

REFERÊNCIAS

- BARROS, J. F. C.; CALADO, J. G. **A Cultura do Milho**. Universidade de Évora- Escola de Ciências e Tecnologia- Departamento de Fitotecnia, 2014.
- BEAVIS, W. D.; GRANT, D. A linkage map based on information from four F2 populations of maize (*Zea mays* L.). **Theoretical and Applied Genetics**, v. 82, n. 5, p. 636-644, 1991.
- BENNETZEN, J. L. Transposable element contributions to plant gene and genome evolution. **Plant molecular biology**, v. 42, n. 1, p. 251-269, 2000.
- BENT, Andrew F. Plant disease resistance genes: function meets structure. **The Plant Cell**, v. 8, n. 10, p. 1757, 1996.
- CHISHOLM, Stephen T. et al. Host-microbe interactions: shaping the evolution of the plant immune response. **Cell**, v. 124, n. 4, p. 803-814, 2006.
- COULTER, J. Advance corn hybrid selection with new trial results. **Minnesota Crop News**. University of Minnesota. Minnesota. 2014.
- DOEBLEY, J. The Genetics of Maize Evolution. University of Wisconsin. **Annual Reviews of Genetics**, 38:37-59, 2004.
- FEUILLET, C.; KELLER, B. Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. **Annals of botany**, v. 89, n. 1, p. 3-10, 2002.
- GALE, M. D.; DEVOS, K. M. Comparative genetics in the grasses. **Proceedings of the National Academy of Sciences**, v. 95, n. 5, p. 1971-1974, 1998.
- HAMMOND-KOSACK, Kim E.; JONES, Jonathan DG. Plant disease resistance genes. **Annual review of plant biology**, v. 48, n. 1, p. 575-607, 1997.
- KELLER, Beat; FEUILLET, Catherine. Colinearity and gene density in grass genomes. **Trends in plant science**, v. 5, n. 6, p. 246-251, 2000.
- KIM, Jungeun et al. A genome-wide comparison of NB-LRR type of resistance gene analogs (RGA) in the plant kingdom. **Molecules and cells**, v. 33, n. 4, p. 385-392, 2012.
- MCDOWELL, John M.; SIMON, Stacey A. Molecular diversity at the plant-pathogen interface. **Developmental & Comparative Immunology**, v. 32, n. 7, p. 736-744, 2008.
- MICHAELSON, Martin J. et al. Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. **American journal of botany**, p. 183-188, 1991.
- OLIVEIRA, E. C. Marcadores ISSR na formação de grupos heteróticos e na inferência evolutiva de milho pipoca. **Universidade Estadual do Norte Fluminense**. Campos dos Goytacazes. Rio de Janeiro – RJ. 2010.
- PURCINO, A.A.; Sorgo sacarino na Embrapa: Histórico, importância e usos. **AGROENERGIA Ano II**, Edição nº, p 6-7, 2011.
- SAEYS, Y.; INZA, I.; LARRAÑAGA, P. A review of feature selection techniques in bioinformatics. **Bioinformatics**, v. 23, n. 19, p. 2507-2517, 2007.
- SANSEVERINO, Walter; ERCOLANO, Maria Raffaella. In silico approach to predict candidate R proteins and to define their domain architecture. **BMC research notes**, v. 5, n. 1, p. 1, 2012.
- SPRINGER, P. S.; ZIMMER, E. A.; BENNETZEN, J. L. Genomic organization of the ribosomal DNA of sorghum and its close relatives. **Theoretical and applied genetics**, v. 77, n. 6, p. 844-850, 1989.
- VAN OOIJEN, G., VAN DEN BURG, H. A., CORNELISSEN, B. J., TAKKEN, F.L. Structure and function of resistance proteins in solanaceous plants. **Annual Review of Phytopathology**, 45, 43-72. 2007.

Tabela 2. Distribuição de classes de genes R encontradas nas espécies *S. bicolor*, *Zea mays* subsp. *parviglumis* e *Teosinto* comparadas a *Zea mays* subsp. *mays*.

Espécies comparadas com Zea mays	TNL	CN	CNL	Mlo-like	N	NL	RLK	RLK-GNK2	RLP	T	Outros	Unknown	Total
<i>S. bicolor</i>	18	10	38	5	48	59	2	15	111	5	3	616	930
<i>Teosinto</i>	1	-	2	-	4	2	8	-	-	-	-	27	44
<i>Z. mays</i> subsp. <i>pavirglumis</i>	3	1	10	-	5	3	2	1	15	-	3	66	109
Comum à todos	0	0	0	0	0	0	0	0	0	0	0	2	2
TOTAL													1085

