Vinicius Vilperte

TRANSCRIPTÔMICA E MIRNÔMICA COMPARATIVA DE VARIEDADES TRANSGÊNICAS DE MILHO

Dissertação submetida ao Programa de Pós-Graduação em Recursos Genéticos Vegetais da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Ciências
Orientador: Prof. Dr. Rubens Onofre Nodari

Florianópolis
2016

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Vilperte, Vinicius
Transcriptômica e mirnômica comparativa de variedades transgênicas de milho / Vinicius Vilperte ; orientador, Rubens Onofre Nodari - Florianópolis, SC, 2016.
158 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Agrárias. Programa de Pós Graduação em Recursos Genéticos Vegetais.

Inclui referências

1. Recursos Genéticos Vegetais. 2. Organismos Geneticamente Modificados. 3. Análise de Risco. 4. Transcriptoma. 5. miRNA. I. Nodari, Rubens Onofre. II. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Recursos Genéticos Vegetais. III. Título.

AGRADECIMENTOS

Agradeço primeiramente à minha família, em especial aos meus pais Enio e Maria Ires pela excelente educação, conselhos, apoio, contribuição financeira, e por tudo que me proporcionaram de forma incondicional para que eu pudesse concretizar mais essa etapa da minha vida. Gostaria também de agradecer aos meus irmãos Júlio e Pablo.

À Universidade Federal de Santa Catarina pelo ensino gratuito e de qualidade.

Ao Programa de Pós-Graduação em Recursos Genéticos Vegetais e à todos os seus professores e servidores.

Ao CNPq e ao GenØk pelas bolsas de pesquisa e financiamento nos últimos anos.

Ao Prof. Dr. Rubens Onofre Nodari pelos 7 anos de orientação e pela confiança no meu trabalho.

Aos presentes e ex- colegas do LFDGV: Daniela, Lilian, Gustavo, Denise, Rafael, Vanessa, Clarissa, Daniel, Montagna, Raissa et al. pela descontração, pelos finais de semanas compartilhados no lab, por todo o apoio e pelas cervejas pós-expediente.

À Sarah Agapito pelos anos de colaboração, "co-orientação" e ajuda na elaboração e desenvolvimento deste trabalho

Aos amigos de algum jeito sempre presentes: Karin, Marcel, Athos, Nivaldo, Raphaella, Ágata, Mariana, Leonardo, Carolina et al. Obrigado pela amizade e pela torcida.

À todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

“The important thing is not to stop questioning; curiosity has its own reason for existing.”

(Albert Einstein, 1955)

RESUMO

Organismo Geneticamente Modificado (OGM) é definido como um organismo cujo material genético – DNA/RNA - tenha sido modificado por qualquer técnica de engenharia genética e, portanto, pode apresentar riscos advindos da sua manipulação. Para minimizar estes riscos, órgãos reguladores internacionais exigem que o OGM desenvolvido seja submetido à uma análise de risco para a identificação e prevenção de efeitos adversos que podem levar a mudanças não intencionais pela presença de transgene(s) tanto em aspectos moleculares, quanto aspectos ecológicos e de saúde humana e animal. A principal abordagem em análises de risco envolve o conceito de equivalência substancial, onde as características do organismo modificado são comparadas com a sua linha parental (considerado seguro para usos de alimentação humana e animal), envolvendo estudos de performance agronômica, fenotípica e alimentação animal, ou ainda de composição química e nutricional. O problema é que análises baseadas na equivalência substancial não levam em consideração as possíveis mudanças não intencionais geradas pela própria modificação genética, a qual poderia causar potenciais alterações em proteínas (ou seja, possíveis produtos tóxicos e alergênicos). Portanto, o objetivo deste trabalho foi testar duas novas abordagens moleculares para caracterizar os potenciais efeitos sinérgicos e antagonistas dos OGMs em nível molecular. A primeira abordagem visou avaliar comparativamente o perfil de expressão gênica (transcriptoma) de uma variedade de milho GM estaqueada (contendo duas proteínas inseticidas CRY e duas proteínas EPSPS que conferem tolerância ao herbicida glifosato), seus parentais GM simples e a linha isogênica não modificada, todas sob o mesmo background genético. Está combinação de genótipos permite isolar as potenciais alterações no transcriptoma que derivam da combinação dos dois transgenes na variedade estaqueada. A segunda abordagem visou avaliar comparativamente o perfil de expressão de micro RNAs (miRNAs) na mesma combinação de genótipos GMs de milho. miRNAs possuem a capacidade de regular a expressão gênica de diversos genes endógenos e, portanto, desempenham papéis fundamentais em diversos processos biológicos em um organismo. Os resultados das análises demonstraram que, para o perfil transcriptômico, diversos processos biológicos apresentaram diferenças em regulação, principalmente àqueles envolvidos em vias de redox, modificações pós-transducionais e regulação da transcrição gênica. Já para a análise dos perfis de expressão de miRNAs, alguns apresentaram uma regulação diferencial, e estes

demonstraram ser responsáveis pela regulação de fatores de transcrição endógenos, principalmente àqueles envolvidos em desenvolvimento foliar, mecanismos de resistência à estresse, transdução de sinais e processamento e tradução de RNA. Portanto, conclui-se que as abordagens moleculares utilizadas podem ser aplicadas como ferramentas úteis para aumentar a abrangência e confiabilidade em avaliações de riscos de OGMs. Por fim, recomendamos que outras investigações, visando compreender mais detalhadamente a relevância das mudanças encontradas, sejam realizadas, além de sugerir que as agências regulatórias de biossegurança de OGMs considerem que este tipo de estudo em futuras avaliações de risco para a liberação comercial de novos OGMs.

Palavras-chave: Organismos Geneticamente Modificados, análise de risco, transcriptoma, miRNA, expressão gênica.

ABSTRACT

Genetically Modified Organism (GMO) is defined as an organism in which the genetic material – DNA/RNA – has been modified by any technique of genetic engineering and, therefore, could present risks resulting from its manipulation. In order to minimize these risks, international regulatory bodies demand that the newly developed GMO is submitted to a risk assessment in order to identify and prevent adverse effects of transgene(s) that could lead to unintended changes in the GMOs, both at molecular level and ecological and human and animal health aspects. However, the main approach used in risk assessment involves the concept of substantial equivalence, where the characteristics of the modified organism are compared to its parental line (considered safe for animal and human food and feed uses), involving studies of agronomic, phenotypic and animal feeding performances, as well as chemical and nutritional compositions. The problem is that analyses based on substantial equivalence do not take in consideration the possible non-intended changes generated by the genetic modification itself, which could cause potential alterations in proteins (i.e. possible toxic and allergenic products). Therefore, the goal of this study was to test two new molecular approaches to characterize the potential synergic and antagonistic effects of GMOs at the molecular level. The first approach aimed to comparatively evaluate the gene expression profile (transcriptomics) of a stacked GM maize variety (containing two insecticidal CRY proteins and two EPSPS proteins which confer tolerance to the glyphosate herbicide), its single GM parental and the near-isogenic non-modified line, all of them under the same genetic background. This combination of genotypes allows isolating the potential alterations in the transcriptomics that derive from the combination of the two transgenes in the stacked variety. The second approach aimed to comparatively evaluate the expression profile of micro RNAs (miRNAs) in the same combination of GM maize genotypes. miRNAs have the ability of regulating gene expression of several endogenous genes and, therefore, play major role in a range of biological processes in an organism. The results of the analyses demonstrated that, for the transcriptomic profile, several biological processes showed differences in regulation, particularly those involved in redox pathways, post-translational modifications and regulations of transcription. On the other side, the analysis of the miRNAs expression profiles showed a differential regulation, responsible for targeting endogenous transcription factors, particularly those involved in leaf

development, mechanisms of stress resistance, signal transduction and RNA processing and translation. Therefore, it is concluded that the used molecular approaches could be applied as useful tools to increase broadness and confidence in risk assessments of GMOs. At last, we recommend that other investigations, aiming to address in more detail the relevance of such changes, should be conducted, besides suggesting that GMO safety regulatory bodies take into consideration these types of studies and require that it becomes mandatory in the risk assessment for the releasing of new GMOs.

Key words: Genetically Modified Organisms, risk assessment, transcriptomics, miRNA, gene expression.

LISTA DE FIGURAS

REVISÃO BIBLIOGRÁFICA

CAPÍTULO I

CAPÍTULO II

LISTA DE TABELAS

CAPÍTULO I

CAPÍTULO II

LISTA DE ABREVIATURAS E SIGLAS

µL	Microlitro
µM	Micromolar
A	Adenina
bp	Pares de base
C	Citosina
CTNBio	Comissão Técnica Nacional de Biossegurança
DNA	Ácido desoxirribonucleico
dsDNA	DNA de fita dupla
EFSA	European Food Safety Authority
g	Força centrífuga relativa
G	Guanina
GM	Geneticamente Modificado
kDA	Kilo Dalton
mg	Miligrama
mL	Mililitro
mM	Milimolar
miRNA	Micro RNA
mRNA	RNA mensageiro
ng	Nanograma
nt	Nucleotídeo
NGS	Sequenciamento de Nova Geração
NRQ	Quantificação Relativa Normalizada
OGM	Organismo Geneticamente Modificado
PCR	Reação de Polimerase em Cadeia
pmol	Picomol
PTGS	Silenciamento genético pós-transcricional
RNA	Ácido ribonucleico
RNAi	RNA interferência
RNA-seq	Sequenciamento de RNA
rpm	Rotações por minuto
RT-qPCR	PCR Quantitativa Real-Time
sRNA	Pequeno RNA
T	Timina
TF	Fator de transcrição
TGS	Silenciamento genético transcricional
U	Uracila
V	Volts

SUMÁRIO

1. INTRODUÇÃO E JUSTIFICATIVA

O termo engenharia genética é definido, segundo a Lei nº 11.105, de 24 de março de 2005, como sendo a atividade de produção e manipulação de moléculas de DNA/RNA recombinante (BRASIL, 2005). Inicialmente conhecida como tecnologia do DNA recombinante, esta tecnologia permite a transferência de genes entre espécies pertencentes a diferentes reinos e que não possuem probabilidades de cruzamento natural. Sua origem ocorreu em 1973, resultante de um experimento dos geneticistas Herbert Boyer e Stanley Cohen, que descobriram um método de clonagem em que um gene de RNA ribossomal de sapo poderia ser transferido e expresso em bactérias (COHEN et al. 1973). Atualmente, a expressão transgenia é usada para descrever organismos com (trans)genes ou genes exógenos.

A descoberta da tecnologia do DNA recombinante foi considerada um marco importante na história da genética. Porém, os geneticistas tomaram consciência do perigo iminente da criação de elementos de DNA infecciosos, provindos da manipulação de genes recombinantes e de suas imprevisíveis propriedades biológicas (BERG et al. 1974). Em 1975, um grupo de geneticistas redigiu e assinou a *Declaração de Asilomar*, a qual chamou atenção para os riscos potenciais do uso da tecnologia do DNA recombinante. Foi instaurada então uma moratória na engenharia genética até a elaboração de regras rígidas para o seu uso seguro (BERG et al. 1975).

Os Organismos Geneticamente Modificados (OGMs) podem ser definidos, segundo a Lei nº 11.105, de 24 de março de 2005, como um organismo cujo material genético – DNA/RNA tenha sido modificado por qualquer técnica de engenharia genética (BRASIL, 2005) e, portanto, podem apresentar riscos advindos da sua manipulação. Pensando nisso, a Convenção sobre Diversidade Biológica (CBD, do inglês *Convention on Biological Diversity*) adotou no ano de 2000 o Protocolo de Cartagena, o qual tem por objetivo "*assegurar níveis adequados de segurança no campo de transferência, manipulação e uso de organismos vivos modificados resultantes da engenharia genética que possam ter efeitos adversos na conservação e no uso sustentável da diversidade biológica, levando em conta riscos para a saúde humana, e com enfoque específico em movimentos transfronteiriços*" (CBD, 2000).

Na inexistência de estatísticas oficiais, segundo o relatório sobre a Situação Global das Culturas Biotecnológicas/GM Comercializadas em 2014, realizado pelo Serviço Internacional para Aquisição de Aplicações em Agrobiotecnologia (ISAAA, sigla em inglês), 181.5

milhões de hectares foram cultivados com culturas geneticamente modificadas (GM), um crescimento anual de 3,6% em relação a 2013. Além disso, 2014 foi o primeiro ano em que países em desenvolvimento cultivaram mais hectares com culturas GM do que países industrializados. A área total de culturas GM aumentou de 1,7 milhões de hectares (1996) para 181,5 milhões (2013), fazendo das culturas GM a mais rápida tecnologia adotada nos últimos anos. Os Estados Unidos aparece em primeiro lugar em área plantada (73,1 milhões de hectares), seguido pelo Brasil (42,2 milhões) e Argentina (24,3 milhões) (JAMES, 2014).

O Brasil é considerado, nos últimos cinco anos, como um dos países que mais aumentou a sua produção de OGMs globalmente. Em 2014, o Brasil foi responsável pelo cultivo de 23% do total dos 181 milhões de hectares cultivados no mundo. Pelo segundo ano consecutivo, o principal fator que impulsionou esse crescimento foi o plantio no país da primeira soja estaqueada (com resistência a insetos e tolerância a herbicidas) em 5,2 milhões de hectares (um crescimento aproximado de 136% em relação a 2013) (JAMES, 2014). Ainda segundo o relatório da ISAAA, 28% dos 181,1 milhões de hectares plantados com variedades GM foram ocupados por eventos estaqueados (i.e. possuem duas ou mais características introduzidas por cruzamento convencional) mostrando o contínuo aumento da utilização destes eventos em relação ao ano anterior (27% dos 175 milhões de hectares). É esperado que essa ampla utilização de eventos estaqueados continue crescendo cada vez mais (JAMES, 2014). O Brasil figura como um dos países que mais auxilia no crescimento da utilização de eventos estaqueados: dos 49 eventos de plantas GM liberados comercialmente no país, 25 deles são eventos estaqueados (CTNBIO, 2016).

A análise de risco de novos OGMs, segundo o Anexo III do Protocolo de Cartagena sobre Biossegurança, tem como principais objetivos identificar e avaliar os efeitos adversos potenciais dos organismos vivos modificados na conservação e no uso sustentável da diversidade biológica no provável meio receptor, levando também em conta os riscos para a saúde humana e animal e ao meio ambiente. Esse tipo de análise deve ser realizada caso-a-caso, levando em consideração as características específicas de cada novo OGM (CBD, 2000; HEINEMANN et al. 2011). Entretanto, para a CTNBio (Resolução Normativa nº 5 de 2008, Art. 4º), quando se trata de eventos estaqueados, as análises de risco se baseiam no princípio de que, se ambos os parentais do evento estaqueado (i.e. eventos simples) foram considerados seguros, o evento estaqueado também será (CTNBIO,

2008). Assim, a *priori*, para o órgão regulamentador brasileiro, não existem interações entre os dois transgenes. Porém, este tipo de abordagem pode levantar algumas questões relevantes no âmbito da biossegurança, tais como: Quais as consequências desse novo evento? Existe a possibilidade de ocorrência de efeitos pleiotrópicos devido a interações entre os dois transgenes? Existe a possibilidade de efeitos adversos devido à inserção do novo transgene?

Visando responder estas e outras perguntas, técnicas chamadas de *profiling* já são consideradas rotineiras em estudos de análise de riscos, sendo aplicadas em estudos de caracterizações moleculares e composicionais (AHTEG, 2010). No âmbito da segurança dos OGMs, técnicas que permitem a análise em ampla escala de milhões de genes, proteínas ou metabólitos oferecem subsídios para uma avaliação mais profunda dos possíveis efeitos adversos oriundos da transgenia (DAVIES et al. 2010). As técnicas *profiling* podem ser classificadas em três tipos de abordagem: *targeted profiling*, *semi-targeted profiling* e *untargeted profiling* (HEINEMANN et al. 2011). As abordagens *targeted* visam avaliar parâmetros pré-determinados, tendo como consequência a captura de informações bastante específicas sobre possíveis efeitos adversos. Dentre as técnicas utilizadas neste tipo de abordagem encontram-se aquelas baseadas em afinidade ou hibridização, como por exemplo, *Southern blot*, *Western blot* e *Northern blot*. As abordagens *semi-targeted*, assim como as *targeted*, também visam analisar moléculas específicas, mas utilizam técnicas e metodologias capazes de detectar uma classe maior de moléculas de interesse. Normalmente, técnicas capazes de detectar variantes de mRNA e proteínas são aplicadas neste tipo de abordagem. Já as abordagens *untargeted* visam avaliar uma classe inteira de moléculas de interesse, como por exemplo, mudanças nas quantidades de proteínas e RNA (HEINEMANN et al. 2013).

Sendo assim, este trabalho tem por objetivo utilizar as diferentes abordagens de *profiling* para investigar e agregar conhecimento científico sobre possíveis alterações nos produtos da expressão gênica e expressão de micro RNAs (miRNAs) de eventos transgênicos simples e estaqueado, assim como dos seus comparadores adequados. Os resultados deverão contribuir para a diminuição das incertezas, geração de subsídios para os órgãos de biossegurança e identificação de possíveis riscos decorrentes destas transformações genéticas.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Analisar os perfis de expressão de mRNA e miRNA de híbridos transgênicos de milho contendo genes de resistência à herbicida e genes de tolerância à insetos.

2.2. OBJETIVOS ESPECÍFICOS

a) Realizar sequenciamento de nova geração de mRNAs e miRNAs em híbridos de milho contendo eventos transgênicos simples e estaqueados, assim como a sua isolinha convencional e uma variedade crioula;

b) Realizar análise comparativa de mRNA entre híbridos de milho contendo eventos transgênicos simples e estaqueados, assim como na sua isolinha convencional e uma variedade crioula;

c) Realizar análise comparativa de miRNA entre híbridos de milho contendo eventos transgênicos simples e estaqueados, assim como na sua isolinha convencional e uma variedade crioula;

3. REVISÃO BIBLIOGRÁFICA

3.1. *NEXT-GENERATION SEQUENCING* E SUAS APLICAÇÕES EM PLANTAS

As primeiras gerações de sequenciamento de DNA foram desenvolvida por Frederick Sanger (SANGER et al. 1977) e por Allan Maxam e Walter Gilbert em 1977 (MAXAM; GILBERT, 1977). O sequenciamento Sanger baseia-se no método de terminação de cadeia, enquanto o sequenciamento Maxam-Gilbert se baseia em um método de modificação química do DNA Devido à menor complexidade técnica e à capacidade de ampliação do método, o sequenciamento Sanger foi o que prevaleceu entre os métodos de sequenciamento de primeira geração (SCHADT et al. 2010).

Todos os campos de conhecimento envolvendo sequenciamento de DNA, em especial a genética, têm sofrido uma imensa revolução nos últimos anos devido ao desenvolvimento e evolução das tecnologias de sequenciamento de nova geração (NGS, do inglês *Next-Generation Sequencing*) (VAN DIJK et al. 2014; QUAIL et al. 2012), também chamadas de sequenciamento de segunda geração. As principais características que tornaram estas tecnologias mais atraentes, em oposição às tecnologias de sequenciamento de primeira geração são: i) preparo de bibliotecas de NGS livres de sistemas celulares de clonagem em bactérias; ii) produção de milhões de reações de sequenciamento em paralelo; e iii) resultado do sequenciamento detectado sem a necessidade de eletroforese (VAN DIJK et al. 2014).

A primeira tecnologia NGS a ser lançada, em 2005, foi o método de pirosequenciamento da plataforma Roche 454 (VAN DIJK et al. 2014). Desde então, diversos métodos NGS vêm sendo criados e aprimorados anualmente. Estes métodos podem ser divididos em duas principais categorias de acordo com a metodologia de sequenciamento: sequenciamento por síntese e sequenciamento por ligação. No sequenciamento por síntese, a determinação da composição de bases se baseia na detecção de quimiluminescência, fluorescência ou mudanças de pH criada pela incorporação de nucleotídeos durante a síntese da cadeia complementar de DNA. Em resumo, o DNA é fragmentado em tamanho apropriado, ligado à adaptadores e amplificado para aumentar o sinal fluorescente (EGAN et al. 2012). Além da plataforma Roche 454, as outras tecnologias utilizadas são as das plataformas Illumina (Illumina®) e Ion Torrent (Thermo Fischer Scientific). Já na síntese por ligação, a determinação da sequência ocorre através da sensibilidade de

mismatch da DNA ligase (LANDEGREN et al. 1988). Este método utiliza sondas de oligonucleotídeos de tamanhos variados acoplados com *tags* fluorescentes, de acordo com o nucleotídeo a ser determinado (EGAN et al. 2012). A principal tecnologia que utiliza esta metodologia é a da plataforma SOLiD (*Sequencing by Oligonucleotide Ligation and Detection*) (Applied Biosystems/Thermo Fischer Scientific).

Tecnologias mais avançadas estão sendo desenvolvidas, as chamadas Sequenciamento de Terceira Geração (do inglês *Third-Generation Sequencing*). De acordo com Egan et al. (2012), estas metodologias são capazes de produzir sinais detectáveis da incorporação de nucleotídeos por quimiluminescência durante o sequenciamento de DNA de uma única molécula, eliminando assim a necessidade de amplificação de cópias de DNA. As tecnologias de terceira geração já disponíveis são Helicos Genetic Analysis Platform (Helicos BioSciences Corporation) e PacBio (Pacific Biosciences®). Existe também outra tecnologia em desenvolvimento, a Oxford Nanopore (Oxford Nanopore™ Technologies), que além das características mencionadas acima, ainda se baseia na detecção de bases sem o uso de florescências e obtenção de sequências longas (BUERMANS; DEN DUNNEN, 2014).

Com o advento do NGS aliado a facilidade e velocidade das ferramentas bioinformáticas, diversos estudos considerados de alta complexidade tornaram-se viáveis. Uma das principais aplicações do NGS é o sequenciamento e caracterização completa do genoma de espécies de interesse, também chamado de *Whole-Genome Shotgun* (WGS). Diversas espécies de plantas já possuem o seu genoma, ou pelo menos a maior parte dele, sequenciado e anotado, como por exemplo, *Arabdopsis thaliana* (THE ARABIDOPSIS GENOME INIATIVE, 2010), *Triticum aestivum* (BRENCHLEY et al. 2012) e *Zea mays* (SCHNABLE et al. 2009).

Outra aplicação que vêm sendo largamente utilizada é o ChIP-Seq (*Chromatin Immunoprecipitation Sequencing*), o qual consiste em uma estratégia que combina a imunoprecipitação de cromatina (ChIP) com o NGS para identificar sítios de ligação de proteínas associadas ao DNA, como por exemplo fatores de transcrição (TFs) e histonas (KAUFMANN et al. 2009; RICARDI et al. 2014). Esta metodologia vem sendo aplicada com mais frequência em plantas modelos, como *Arabdopsis thaliana* (KAUFMANN et al. 2009; YANT et al. 2010; MOYROUD et al. 2011), mas também existem estudos com plantas não modelos, como *Solanum lycopersicum* (RICARDI et al. 2014), *Eucalyptus grandis* (HUSSEY et al. 2015) e *Phaseolus vulgaris* (AYYAPPAN et al. 2015).

Estudos na área de epigenética e metilação de DNA utilizaram por muito tempo métodos de sequenciamento bissulfito baseados em eletroforese capilar, devido à sua capacidade de resolução, mas com aplicação restrita à regiões pequenas (KRUEGER et al. 2012). Com o advento das tecnologias NGS, métodos baseados na conversão por bissulfito acoplado com a tecnologia NGS surgem como métodos promissores, uma vez que são capazes de gerar perfis completos do genoma metilado em uma resolução de bases únicas (CHENG; ZHU, 2013). A maioria dos estudos e conhecimento acerca de perfis de metilação em plantas são com *A. thaliana* (COKUS et al. 2008; LISTER et al. 2008; HAGMANN et al. 2015), mas outras espécies também vêm sendo alvos de estudos epigenéticos (ZHONG et al. 2013; LI et al. 2014; RAMBANI et al. 2015).

Recentemente, estudos nos campos da metagenômica e metatranscriptômica, definidos como a análise genética direta dos genomas e transcriptomas presentes em uma amostra ambiental sem a necessidade prévia de cultivos clonais (OULAS et al. 2015), se tornaram viáveis e promissores devido às técnicas de NGS. Estudos nestas áreas visam investigar e responder questões sobre quais organismos estão presentes em comunidades ambientais, quais as suas funções e como estes organismos interagem e mantêm um balanço ecológico no ambiente (OULAS et al. 2015; MELCHER et al. 2014). A principal aplicação destes estudos em plantas se dá no campo da interação planta-micro-organismo, tanto em estudos com interações na rizosfera (BROWN et al. 2012; KNIEF, 2014; LIAO et al. 2014), quanto interações com micro-organismos endógenos (SESSITSCH et al. 2012).

Finalmente, na temática de estudo dos OGMs, a utilização de tecnologias NGS ainda é limitada. A maioria dos estudos conduzidos visam a caracterização do transgene inserido ao nível de rearranjos, *indels* e cópias inseridas (KOVALIC et al. 2012; YANG et al. 2013; FRITSCH et al. 2015), mas poucos visam analisar possíveis efeitos adversos no perfil de expressão gênica (ou seja, no transcriptoma). Entre àqueles que estudaram o perfil transcriptômico, a grande maioria utiliza *microarrays* (GREGERSEN et al. 2005; CHENG et al. 2008; BATISTA et al. 2008), e poucos deles utilizam tecnologias NGS, como por exemplo o RNA-Seq (KAWAKATSU et al. 2013).

3.2. TRANSCRIPTÔMICA

O RNA, considerado um ácido nucleico de fita única, só foi separado do mundo do DNA em 1958 quando Francis Crick formulou a hipótese do dogma central da biologia, onde ele expôs a ideia de que a informação genética é transcrita do DNA para o RNA e posteriormente traduzida de RNA para proteína (CRICK, 1958; DONG; CHEN, 2013). Ainda em 1958, Francis Crick formulou a hipótese da existência de um "adaptador", o qual seria específico para cada aminoácido e seria capaz de se ligar, através de pareamento de bases, na molécula de RNA mensageiro (mRNA). Este "adaptador" foi identificado mais tarde como sendo o RNA transportador (tRNA) (HOAGLAND et al. 1958). Apesar de F. Crick ter formulado a hipótese do pareamento do tRNA com o mRNA, a ideia da existência do mRNA e a sua associação com o ribossomo no processo de transcrição só foi formulado alguns anos depois por Jacob e Monod (JACOB; MONOD, 1961; DONG; CHEN, 2013).

Já em 1977, alguns pesquisadores demonstraram a existência de sequências dos genes que codificam para proteínas (éxons) e outras que não codificam (íntrons), e que, durante o *splicing*, os íntrons são retirados e degradados e os éxons são montados em diferentes mRNAs, fenômeno chamado de *splicing* alternativo (BERGET et al. 1977; CHOW et al. 1977). Nos anos 90, alguns cientistas observaram a existência de um fenômeno onde moléculas de RNA eram capazes de inibir a expressão gênica (NAPOLI et al. 1990; ECKER et al. 1986), o que mais tarde foi chamado de RNA interferência (RNAi), no qual moléculas de fita dupla de RNA (dsRNA, do inglês *double-stranded RNA*) são capazes de reconhecer sequências específicas de mRNA e levar à sua degradação (FIRE et al. 1998).

O transcriptoma, definido como o conjunto completo de transcritos de DNA (i.e. RNAs) em uma célula e suas quantidades, para um estágio específico de desenvolvimento ou condição fisiológica, representa uma importante relação entre o fenótipo e as informações codificadas pelo DNA (WANG et al. 2009; VALDÉS et al. 2013). Estudos transcriptômicos se tornaram um campo de estudos promissor na era pós-genômica (LOCKHART; WINZELER, 2000), devido à razões, como a capacidade de refletir a dinâmica espaço-temporal da expressão gênica, suporte à estudos de proteômica, e possibilidade de estudos de estrutura e função de RNA não-codantes (DONG; CHEN, 2013), detalhadas no item 3.3. Os principais objetivos de estudos transcriptômicos são: i) identificar todos os tipos de transcritos,

incluindo mRNAs, RNAs não-codantes (ncRNAs) e pequenos RNAs (sRNAs); ii) determinar a estrutura dos genes, em termos de locais de início, extremidades 5' e 3', padrões de *splicing* e outras modificações pós-transcricionais; e iii) quantificar mudanças nos níveis de expressão de cada transcrito (WANG et al. 2009).

Ao longo dos anos, diversas tecnologias capazes de analisar a expressão de múltiplos genes vêm sendo aprimoradas para a aplicação em estudos transcriptômicos (WARD et al. 2012), incluindo abordagens baseadas em hibridização e sequenciamento (WANG et al. 2009). Das tecnologias de hibridização, os *microarrays* de cDNA (SCHENA et al. 1995) são os mais utilizados e se baseiam na hibridização de cDNA do organismo de interesse, o qual foi marcado com fluorescência, em sondas de DNA em um chip, seguido da detecção da fluorescência relativa (WARD et al. 2012).

Mais especificamente, *microarrays* (ou método de chip) são *pools* de sondas de oligonucleotídeos curtos, representando diversos genes, ligadas a um substrato (geralmente lâminas de vidro) em locais pré-definidos (VALDÉS et al. 2013). Bibliotecas de cDNA marcadas com fluorescência são incubadas e hibridizadas com as sondas, permitindo então a quantificação da abundância relativa de milhões de transcritos de duas ou mais amostras (SCHENA et al. 1995; KARAKACH et al. 2010). *Microarrays* de alta densidade permitem a análise de expressão gênica a baixos custos, além de poderem ser criados de acordo com o objetivo das análises, como por exemplo, a detecção de diferentes variantes de um transcrito (CLARK et al. 2002).

Existem também os *microarrays* chamados de *tiling genome arrays*, os quais não exigem um conhecimento prévio do transcriptoma da espécie de interesse e utilizam um conjunto de sondas que se sobrepõem para a detecção da expressão gênica do genoma completo (STOLC et al. 2005). Entretanto, esta categoria de *microarrays* apresenta um alto custo para estudos de genomas grandes, além de apresentar um alto *background* de hibridização devido à impossibilidade de distinguir transcritos com alta homologia entre si (DRAGHICI et al. 2006).

Em contraste aos métodos de *microarrays*, as abordagens de sequenciamento são capazes de determinar diretamente a sequência de cDNA (WANG et al. 2009). Entre os métodos de sequenciamento utilizados para estudos de expressão gênica diferencial, um dos primeiros desenvolvidos, em escala de transcriptoma completo, foi o sequenciamento pelo método Sanger de bibliotecas de cDNA ou *expressed sequence tags* (ESTs), os quais foram utilizadas por muito

tempo (BOGUSKI et al. 1994), com trabalhos recentes ainda utilizando esta tecnologia (BLAIR et al. 2011). Entretanto, essa metodologia apresenta resultados insatisfatórios, preços altos e os dados geralmente não são quantitativos (WANG et al. 2009).

Outros métodos desenvolvidos na mesma época foram o *Differential Display Reverse Transcriptase Polymerase Chain Reaction* (DDRT-PCR) (LIANG; PARDEE, 1992) e *Serial Analysis of Gene Expression* (SAGE) (VELCULESCU et al. 1995). O DDRT-PCR utiliza um primer seletivo para a amplificação de *subpools* de mRNA que podem ser visualizados em géis de sequenciamento. Apesar dos vários estudos utilizando esse método, ele demanda muito esforço e alguns dos resultados não são passíveis de reprodução (MALHOTRA et al. 1998; WARD et al. 2012). Já o SAGE se baseia no sequenciamento de fragmentos de cDNA unidos através do métodos Sanger, e a subsequente separação dos resultados e mapeamento contra um genoma ou ESTs (ADAMS et al. 1991).

As tecnologias baseadas em sequenciamento Sanger mais recentes são *Cap Analysis Gene Expression* (CAGE) (SHIRAKI et al. 2003) e *Massively Parallel Signature Sequencing* (MPSS) (BRENNER et al. 2000). A tecnologia CAGE baseia-se na preparação e sequenciamento de *tags* de concatâmeros de DNA derivados dos 20 nucleotídeos inicias do 5' *end* de mRNAs, possibilitando a análise de sítios de início de transcrição e uso de promotores (SHIRAKI et al. 2003). Já a tecnologia MPSS produz dados similares aos do SAGE, produzindo sequências curtas que podem ser alinhadas com genomas de referência ou ESTs (BRENNER et al. 2000).

Apesar de estas tecnologias apresentarem alta capacidade e precisão em gerar dados de níveis de expressão gênica, todas elas se baseiam no sequenciamento Sanger, que é considerado caro e laborioso. Em algumas das tecnologias, por exemplo, o MPSS, a preparação das bibliotecas exige uma fase de clonagem em *beads* bastante complexa, fazendo com que a tecnologia não tenha sido amplamente disponível e utilizada (WARD et al. 2012). Além disso, apenas uma parte dos transcritos é analisada e geralmente não é possível diferencias variantes de transcritos (WANG et al. 2009; ZHAO et al. 2014).

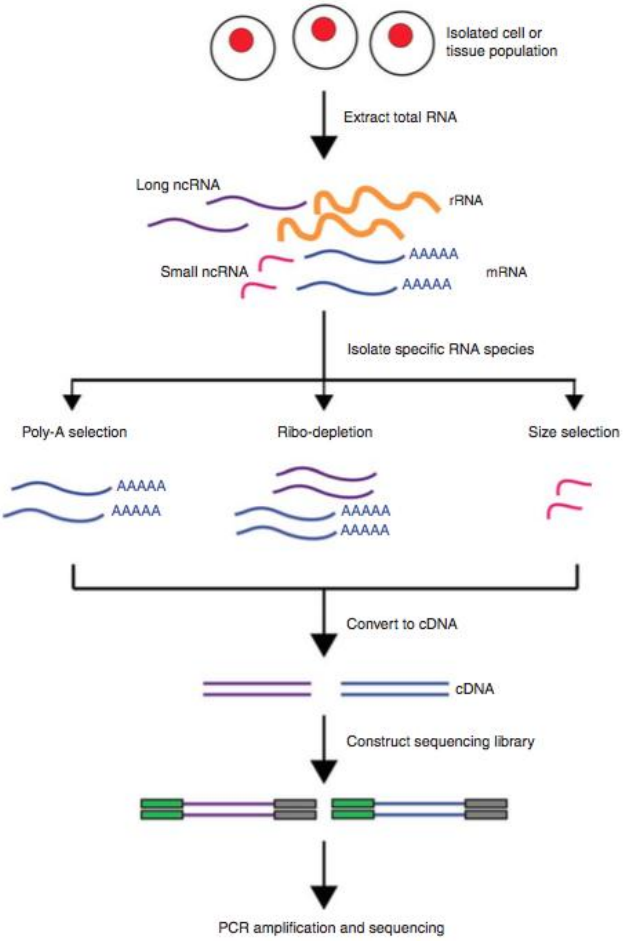
Durante muitos anos, o uso de *microarrays* têm sido a principal escolha em estudos de expressão gênica em larga escala devido à otimização e padronização dos instrumentos e protocolos (MALONE; OLIVER, 2011; ZHAO et al. 2014). Atualmente, os *microarrays* ainda são altamente utilizados e podem ser prontamente utilizados graças ao seu baixo custo e facilidade de uso (ZHAO et al. 2014). Entretanto, esta

tecnologia apresenta problemas em alcançar uma caracterização compreensiva e precisa do transcriptoma devido a algumas limitações (VALDÉS et al. 2013), como por exemplo a precisão na quantificação da expressão, principalmente de transcritos em baixa abundância, devido ao background de hibridização (ZHAO et al. 2014). Além disso, existe uma grande diferença nas propriedades de hibridização das sondas utilizadas e estas são capazes de detectar somente os transcritos para as quais foram desenhadas, não detectando possíveis variantes (ZHAO et al. 2014).

Como alternativa às limitações dos *microarrays*, os métodos de RNA-Seq, definidos como sequenciamento direto de transcritos através de tecnologias de NGS (ZHAO et al. 2014), estão disponíveis para estudos de transcriptoma e são independentes de qualquer sequência anotada para o organismo de interesse, contando com diversos avanços das tecnologias de *microarrays* e várias químicas de sequenciamento (VALDÉS et al. 2013). Além disso, o RNA-Seq apresenta uma maior capacidade de detectar transcritos de baixa abundância, diferenciar isoformas e possibilitar a identificação de variantes genéticas, além de possibilitar a detecção de uma maior quantidade de genes diferencialmente expressos com *fold-change* mais altos (ZHAO et al. 2014).

Para a realização de um experimento de RNA-Seq, os seguintes passos são normalmente aplicados: 1) extração e isolamento do RNA de interesse; 2) conversão do RNA para DNA complementar (cDNA); 3) preparo da biblioteca de cDNA; e 4) amplificação e sequenciamento da biblioteca em uma plataforma NGS (Figura 1). Após o sequenciamento, os *reads* obtidos são alinhados com um genoma de referência ou montados por meio da estratégia *de novo* (sem um genoma de referência) (WANG et al. 2009). Além disso, de acordo com o objetivo do experimento, diversos detalhes devem ser considerados antes da realização do RNA-Seq, como por exemplo, o uso de replicatas biológicas e técnicas, profundidade de sequenciamento, cobertura desejada do transcriptoma e a escolha da plataforma de NGS adequada (KURKUBA; MONTGOMERY, 2015).

Figura 1. Workflow de um experimento de RNA-Seq. Primeiramente, o RNA total é isolado da amostra de interesse, seguido de um isolamento do tipo de RNA adequado para análise. Em seguida, o RNA isolado é convertido em DNA complementar (cDNA), o qual é utilizado para a construção de bibliotecas enriquecidas e posteriormente amplificadas via PCR.



Fonte: Figura adaptada de Kurkuba e Montgomery (2015).

3.3. *NON-CODING* RNAS E MECANISMOS DE SILENCIAMENTO DE GENES

ncRNA (*non-coding RNA*) é a denominação utilizada para pequenas moléculas de RNA que não codificam proteínas funcionais (CHAPMAN; CARRINGTON, 2007). De acordo com as suas funções, eles podem ser divididos em *house-keeping* e ncRNAs regulatórios. Normalmente, ncRNAs considerados *house-keeping* exercem funções estruturais e catalíticas (tRNAs e rRNAs envolvidos no processo de tradução), *small nuclear* RNAs (snRNAs – envolvidos no processo de *splicing* de mRNA), *small nucleolar* RNAs (snoRNAs – envolvidos no processo de *splicing* de rRNA), entre outros (EDDY et al. 2001). Outras formas de RNA também são conhecidas, como o circRNA (*circular* RNA), mas as suas funções reguladoras e vias metabólicas ainda não estão totalmente elucidadas (KOSIK, 2013).

Os ncRNAs regulatórios estão envolvidos em diversos processos biológicos. Estes ncRNAs podem ser divididos, de acordo com o seu comprimento, em pequenos e longos ncRNAs (lncRNAs) DONG; CHEN, 2013). Os pequenos ncRNAs possuem comprimento menor que 200 nucleotídeos e incluem microRNAs (miRNAs), *short interfering* RNA (siRNA), *trans-acting* RNA (tasiRNA) e piwi-interacting RNA (piRNA), enquanto os lncRNAs possuem comprimento maior que 200 nt, variando até 100 kb (MERCER et al. 2009). Os miRNAs fazem parte de processos biológicos cruciais, como resposta à estresse biológico, desenvolvimento e comportamento celular (KROL et al. 2010). Já siRNAs e piRNAs, produzidos por vias diferentes às dos miRNAs, estão envolvidos no silenciamento gênico de transposons e sequências repetitivas (LIPPMAN; MARTIENSSEN, 2004; PENG; LIN, 2013).

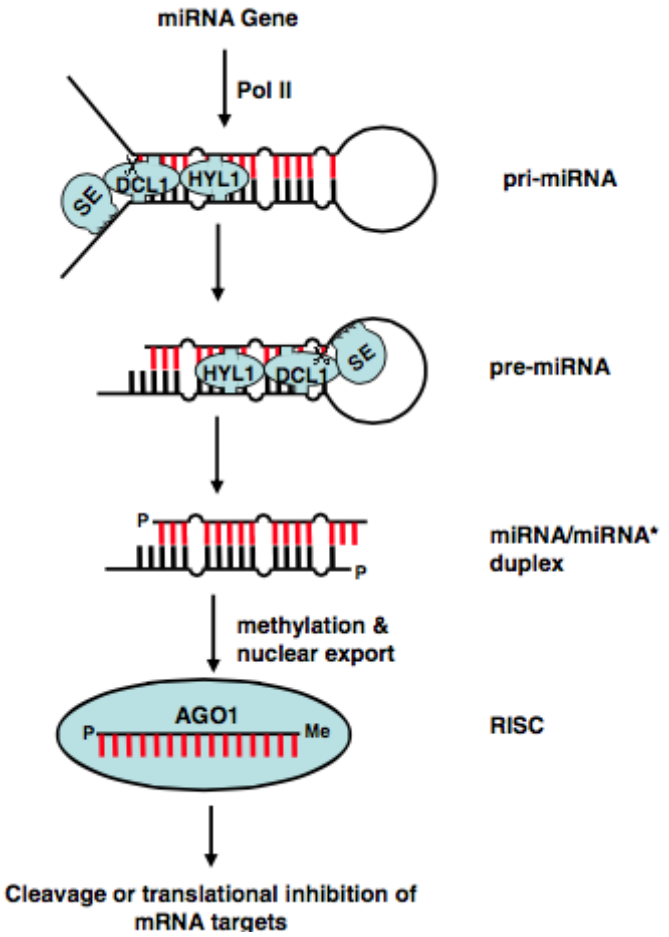
Os lncRNAs são normalmente transcritos da RNA polimerase II, com ausência de *open reading frames*, poliadenilados e com localização abundante no núcleo (MERCER et al. 2009). Os lncRNAs podem ser classificados em 5 categorias de acordo com a sua localização: senso, antisenso, bidirecionais, intrônicos e intergênicos. Além de atuarem em diversos processos biológicos, como por exemplo, resposta a estresse, localização, *splicing* alternativo, remodelação de cromatina, lncRNAs também podem afetar processos celulares, como ciclo celular, sobrevivência, migração e metabolismo (ZHANG et al. 2013).

miRNAs constituem uma grande família de elementos reguladores de expressão gênica, com tamanho variando de 18 a 26 nt, e

que controlam diversos processos celulares em organismos eucarióticos (KROL et al. 2010). Em animais, a maioria dos miRNAs são processados de longos transcritos em forma de *hairpin* através de consecutivas ações de enzimas membros da família da RNA III, DROSHA e DICER, enquanto em plantas somente a enzima DICER é responsável pelo processamento de miRNAs (CARTHEW; SONTHEIMER, 2009). A maioria das plantas possuem mais de 100 genes de miRNAs (chamados de MIR) (NOZAWA et al. 2012), os quais estão localizados quase que exclusivamente em regiões intergênicas do genoma (REINHART et al. 2002).

Em plantas, a biogênese dos miRNAs ocorre exclusivamente no núcleo, através dos seguintes processos: 1) transcrição pela RNA polymerase II de genes de miRNAs, dando origem ao miRNA primário (pri-miRNA) que se reestrutura em forma de *hairpin*; 2) processamento do pri-miRNA por uma enzima RNAse III, a Dicer-like 1 (DCL1 – contendo dois domínios de dupla fita de RNA-binding) (JONES-RHOADES et al. 2006) e seus cofatores associados de RNA-binding Hyponastic Leaves 1 (HYL1 - contendo dois domínios de dupla fita de RNA-binding) (HAN et al. 2003) e Serrate (SE - uma *zinc finger* do tipo C2H2-type) (YANG et al. 2006), para a formação do miRNA; 3) metilação do miRNA/miRNA* duplex, seguido da exportação para o citoplasma e incorporação na proteína Agonaute 1 (AGO1) para formar um dos componentes principais do complexo de silenciamento induzido por RNA (RISC, do inglês *RNA-induced silencing complex*) (ZHU, 2008; ROGERS; CHEN, 2013) (Figura 2).

Figura 2. Ilustração representativa da biogênese de miRNAs em plantas.



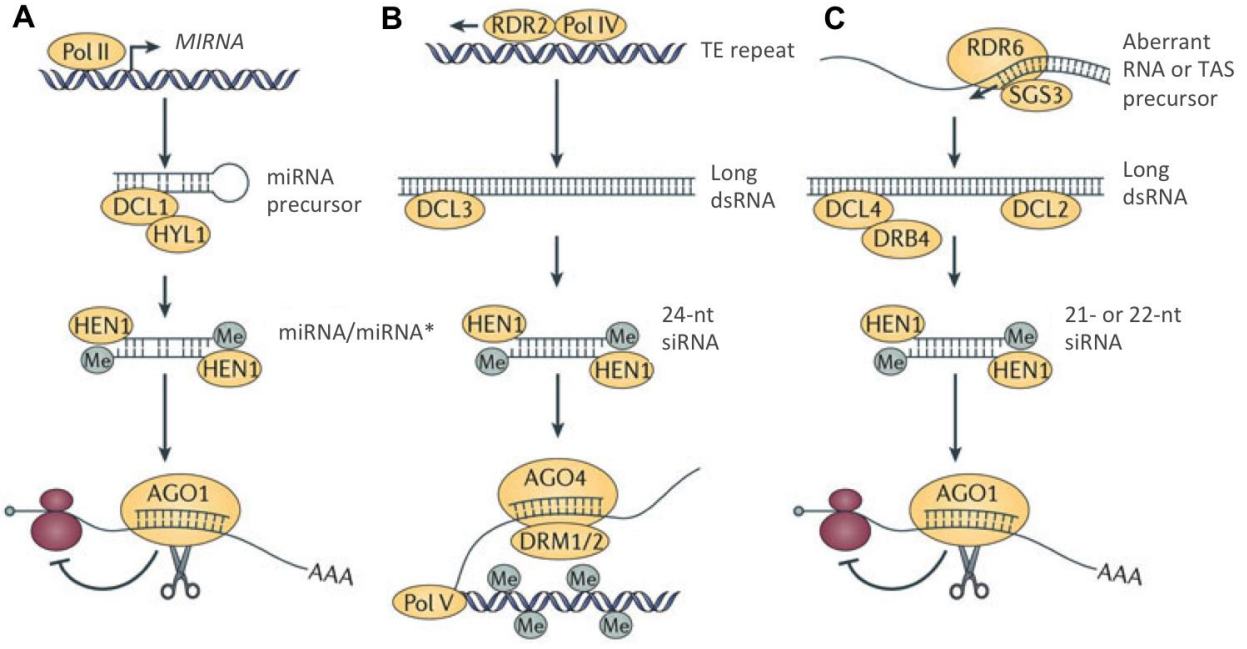
Fonte: Figura adaptada de Zhu (2008)

Já os siRNAs são gerados através de RNAs de dupla fita que se originam de diferentes origens, como RNAs transcritos de regiões invertidas, pares de transcritos em cis-antisenso, pela ação de *RNA-dependent RNA polymerases* (RDRs) que convertem RNA de fita simples em dsRNA, entre outros. O dsRNA é clivado em siRNAs curtos (21 – 24 nt) pela ação de proteínas DCLs, as quais definem o tamanho do siRNA de acordo com a sua atividade catalítica. Similar aos

miRNAs, os siRNAs também são incorporados em complexos RISC, podendo interferir na regulação gênica a níveis pós-transcricionais ou transcricionais (por meio de vias *RNA-directed DNA Methylation*-RdDM) (KHRAIWESH et al. 2012).

A figura 3 apresenta um exemplo da biogênese e modo de silenciamento de miRNAs e siRNAs em plantas. A figura 3A apresenta uma ilustração da biogênese de miRNAs e seu posterior modo de silenciamento pós-transcricional. A figura 3B apresenta a biogênese de siRNAs provenientes de transcritos de regiões repetitivas ou elementos transponíveis e o seu subsequente modo de silenciamento transcricional. Já a figura 3C apresenta a biogênese de siRNAs derivados de outros ncRNAs, mRNAs ou RNAs endógenos em antisenso, com o seu posterior modo de silenciamento pós-transcricional (PUMPLIN; VOINNET, 2013).

Figura 3. Biogênese de miRNAs e siRNAs em plantas. A) Biogênese de miRNAs e seu posterior modo de silenciamento pós-transcricional. B) Biogênese de siRNAs provenientes de transcritos de regiões repetitivas ou elementos transponíveis e o seu subsequente modo de silenciamento transcricional. C) Biogênese de siRNAs derivados de outros ncRNAs, mRNAs ou RNAs endógenos em antisenso, com o seu posterior modo de silenciamento pós-transcricional.



Fonte: Figura adaptada de Pumplin e Voinnet (2013).

Devido à capacidade dos sRNAs de reprimir a expressão gênica, a regulação mediada por RNA é normalmente chamada de silenciamento por RNA, silenciamento gênico ou RNA interferência (RNAi) (VAUCHERET, 2006). RNAi é um mecanismo de regulação gênica que ocorre naturalmente contra a presença de dsRNAs que interferem na tradução de transcritos de mRNA, levando a uma supressão da expressão gênica (FIRE et al., 1998). No citoplasma, sRNAs podem induzir o silenciamento gênico pós-transcricional (PTGS, do inglês *Post-Transcriptional Gene Silencing*) por meio da degradação ou repressão traducional de RNA complementares. No núcleo, sRNAs induzem o silenciamento gênico transcricional (TGS, do inglês *Transcriptional Gene Silencing*) por meio da indução de modificações epigenéticas em regiões homólogas aos siRNAs, como metilação do DNA e histonas (MATZKE; MOSHER, 2014). O TGS ocorre através do bloqueio da transcrição, comumente mediado por uma rota relacionada com a rota de RNAi, a rota de *RNA-dependent DNA Methylation* (RdDM) (FAGARD; VAUCHERET, 2000). Neste mecanismo, um miRNA aberrante, produzido via rota de RNAi, se pareia com uma sequência homóloga no genoma e fornece um substrato atrativo para enzimas *cytosine methyltransferases* (CMTases), as quais catalisam a transferência de um grupo metil *AdoMet* para o anel de citosina e, portanto, são responsáveis pela metilação do DNA (BENDER, 2004).

3.4. ESTUDOS DE BIOSSEGURANÇA DE EVENTOS ESTAQUEADOS

Eventos estaqueados, ou piramidados, podem ser denominados como a combinação de duas ou mais características através de cruzamento tradicional (TAVERNIERS et al. 2008). No Brasil, eventos estaqueados são considerados novos OGMs, entretanto, uma análise de risco completa não é necessária se os seus parentais simples já estão aprovados, conforme determina a Resolução Normativa n° 5 de 2008 que prevê uma análise de risco simplificada nestes casos, a qual exige menos estudos de segurança do que para os parentais (CTNBIO, 2008). Já na União Europeia, os eventos estaqueados também são considerados novos OGMs, sendo exigidos estudos, quanto à sua segurança, similares aos exigidos para os eventos simples (DE SCHRIJVER et al. 2007), mas com alguns aspectos não sendo considerados relevantes para os estaqueados (KOK et al. 2014). Os órgãos reguladores internacionais demandam que, para a análise de risco de eventos estaqueados, ambos os parentais GMs, os quais já foram previamente analisados, devem ser

utilizados como comparadores. Na falta das linhas isogênicas simples, o uso da linha isogênica convencional deve ser usada como o comparador primário (CBD, 2012).

Estudos acerca da caracterização molecular destes eventos estaqueados são escassos, e a comparação dos níveis de expressão dos seus transgenes e mesmo de seus genes endógenos com os parentais simples é quase inexistente. Agapito-Tenfen et al. (2014) estudaram a estabilidade dos mesmo híbridos utilizados nas análises desta dissertação, através de análises proteômicas e de expressão de mRNA. Transcritos transgênicos do evento estaqueado apresentaram uma redução de expressão média de cerca de 34% quando comparados com os parentais. A análise proteômica mostrou a existência de 22 proteínas diferencialmente moduladas entre as variedades, com destaque para duas vias metabólicas: energia/carboidrato e metabolismo de desintoxicação. Os autores sugerem que a piramidização de dois insertos transgênicos no genoma de uma variedade GM de milho pode levar a alterações na expressão de genes endógenos. Além disso, a identificação de proteínas relacionadas com vias metabólicas energéticas e a sua diferente modulação entre a variedade estaqueada e simples pode estar relacionada com uma maior demanda de produção de proteínas transgênicas nas células das plantas do evento estaqueado.

Poucos estudos sobre avaliação de riscos ambientais foram conduzidos com eventos estaqueados. Schuppener et al. (2012) alimentaram larvas de *Aglais urticae*, uma espécie de borboleta comum na Europa, com diferentes doses de pólen de milho do evento MON89034 x MON88017 (proteínas CRY1A.105 e CRY2AB2 x proteína CRY3BB1 e resistência à glifosato) para avaliar possíveis efeitos adversos. Larvas alimentadas com 200 grãos de pólen/cm^2 tiveram uma redução na atividade de alimentação. Diferenças significativas no tempo de desenvolvimento foram observadas em densidades de pólen de 300 grãos/cm^2 e em taxas de sobrevivência em densidade de 400 grãos/cm^2. Todavia, os autores concluíram que o risco do evento MON89034 x MON88017 para populações de *A. urticae* é desprezível. Entretanto, comparadores adequados (i.e. parentais contendo os eventos simples e a variedade convencional) não foram utilizados nas alimentações. Höss et al. (2015) avaliaram o efeito do cultivo do evento MON89034 x MON88017 na comunidade de nematoides do solo e nenhum efeito adverso foi encontrado na estrutura dos mesmo. Entretanto, os parentais contendo os eventos simples (neste caso, MON89034 e MON88017) também não foram incluídos nas análises.

Steijven et al. (2015) acessaram o efeito do pólen do evento MON89034 x MON88017 em larvas de *Apis melífera*. Os autores não encontraram efeito nas diferentes doses ministradas em termos de sobrevivência e atraso de desenvolvimento das larvas. Entretanto, efeitos dose-dependentes foram encontrados para o peso pré-pupa, sugerindo a existência de efeitos pleiotrópicos. Um estudo anterior utilizando o mesmo evento foi realizado, no qual parâmetros similares foram analisados, além de efeitos na composição de bactérias intestinais (HENDRIKSMA et al., 2013). Segundo os autores, efeitos adversos não foram observados em nenhum dos parâmetros avaliados. Novamente, é importante ressaltar que para ambos os estudos, os comparadores adequados não foram utilizados.

Estudos ambientais com eventos estaqueados também foram realizados por grupos de pesquisas ligados à empresas desenvolvedoras de OGMs. Thaylor et al. (2007a, 2007b) conduziram experimentos de alimentação com galinhas utilizando dois eventos estaqueados distintos, MON89034 x MON88017 (proteínas CRY1A.105 e CRY2AB2 x proteína CRY3BB1 e resistência à glifosato) e MON89034 x NK603 (proteínas CRY1A.105 e CRY2AB2 x resistência à glifosato). Em ambos os estudos, os autores expuseram as galinhas a oito dietas alimentares (teste, controle e seis dietas de referência com milho convencional), e nenhuma diferença estatística foi encontrada para os parâmetros analisados (peso corporal, consumo de ração, rendimento de carcaça, etc.). Entretanto, os autores também não utilizaram os comparadores adequados para avaliar os reais efeitos das rações.

4. REFERÊNCIAS BIBLIOGRÁFICAS

AD HOC TECHNICAL EXPERT GROUP (AHTEG): United Nations Environment Programme, CBD: Guidance document on risk assessment of living modified organisms, 2010.

ADAMS, M.D.; KELLEY, J.M.; GOCAYNE, J.D., DUBNICK, M.; POLYMEROPOULOS, M.H.; XIAO, H. et al. Complementary DNA sequencing: Expressed sequence tags and Human Genome Project. Science, v.252, 1651–1656, 1991.

AGAPITO-TENFEN, S.Z.; VILPERTE, V.; BENEVENUTO, R.F.; ROVER, C.M.; TRAAVIK, T.I.; NODARI, R.O. Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. BMC Plant Biology, v.14, 346, 2014.

AYYAPPAN, V.; KALAVACHARLA, V.: THIMMAPURAM, J.; BHIDE, K.P.; SRIPATHI, V.R.; SMOLINSKI, T.G. et al. Genome-Wide Profiling of Histone Modifications (H3K9me2 and H4K12ac) and Gene Expression in Rust (Uromyces appendiculatus) Inoculated Common Bean (Phaseolus vulgaris L.). PLoS One, v.10(7), e0132176, 2015.

BATISTA, R.; SAIBO, N.; LOURENCO, T.; OLIVEIRA, M.M Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. PNAS, v.105(9), 3640–3645, 2008.

BENDER, J. DNA methylation and epigenetics. Annual Review of Plant Biology, v.55, 41–68, 2004.

BERG, P. et al. Potential biohazards of recombinant DNA molecules. Science, v.185(26), 1974.

BERGET, S.M.; MOORE, C.; SHARP, P.A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. PNAS, v.74, 3171–3175, 1977.

BLAIR, M.W.; FERNANDEZ, A.C.; ISHITANI, M.; MORETA, D.; SEKI, M.; AYLING, S. et al. Construction and EST sequencing of full-length, drought stress cDNA libraries for common beans (Phaseolus

vulgaris L.). BMC Plant Biol, v.11, 171, 2011.

BOGUSKI, M.S.; TOLSTOSHEV, C.M; BASSETT, D.E.J. Gene discovery in dbEST. Science, v.265, 1993–1994, 1994.

BRASIL. Lei nº 11.105, de 24 de março de 2005. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/lei/l11105.htm. Acessado em: 20 de Janeiro de 2016.

BRENCHLEY, R.; SPANNAGL, M.; PFEIFER, M.; BARKER, G.L.A.; D'AMORE, R.; ALLEN, A.M. et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature, v.491, 705-710, 2012.

BRENNER, S., JOHNSON, M.; BRIDGHAM, J.; GOLDA, G.;. LLOYD, D.H.; JOHNSON, D. et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature Biotechnology, v.18, 630–634, 2000.

BROWN, S.D.; UTTURKAR, S.M.; KLINGEMAN, D.M.; JOHNSON, C.M.; MARTIN, S.L.; LAND, M.L. et al. Twenty-one genome sequences from Pseudomonas species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of Populus deltoides. J Bacteriol, v.194(21), 5991-3, 2012.

BUERMANS, H.P.J.; DEN DUNNEN, J.T. Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta, v.1842, 1932–1941, 2014.

CARTHEW, R. W.; SONTHEIMER, E. J. Origins and mechanisms of miRNAs and siRNAs. Cell, v.136, 642–655, 2009.

CONVENTION ON BIOLOGICAL DIVERSITY. Guidance On Risk Assessment Of Living Modified Organisms. Hyderabad, India, 1-5 October, 2012.

CHAPMAN, E. J.; CARRINGTON, J. C. Specialization and evolution of endogenous small RNA pathways. Nature Reviews Genetics, v.8, 884-896, 2007.

CHENG, K.C.; BEAULIEU, J.; IQUIRA, E.; BELZILE, F.J.; FORTIN,

M.G.; STRÖMVIK, M.V. Effect of transgenes on global gene expression in soybean is within the natural range of variation of conventional cultivars. J Agric Food Chem, v.56(9), 3057-67, 2008.

CHENG, L.; ZHU, Y. A classification approach for DNA methylation profiling with bisulfite next-generation sequencing data. Bioinformatics, v.30(2), 172-9, 2014.
CHOW, L.T.; GELINAS, R.E.; BROKER, T.R.; ROBERTS, R.J. An amazing Sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell, v.12, 1–8, 1977.

CLARK, T.A.; SUGNET, C.W.; ARES, M.J. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science, v.296(5569), 907-10, 2002.

COHEN, S. N.; CHANG, A. C. Y.; BOYER, H. W.; HELLING, R. B. Construction of biologically functional bacterial plasmid in vitro. PNAS, v.70, 3240-3244, 1973.

COKUS, S.J., FENG, S., ZHANG, X., CHEN, Z., MERRIMAN, B., HAUDENSCHILD, C.D. et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature, v.452, 215–219, 2008.

CONVENTION ON BIOLOGICAL DIVERSITY. Cartagena Protocol on Biosafety to the Convention on Biological Diversity: text and annexes. Montreal: Secretariat of the Convention on Biological Diversity, 30p., 2000.

CRICK, F.H. On protein synthesis. Symp Soc Exp Biol, v.12, 138–163, 1958.

CTNBIO. Resolução Normativa Nº 5, de 12 de março de 2008. Disponível em: http://www.ctnbio.gov.br/index.php/content/view/11444.html. Acessado em: 28 de Outubro de 2014.

CTNBIO. Resumo Geral de Plantas Geneticamente modificadas aprovadas para Comercialização. 2016 Disponível em: http://www.ctnbio.gov.br/upd_blob/0002/2086.pdf. Acessado em 16 de janeiro de 2016.

DONG, Z.; CHEN, Y. Transcriptomics: advances and approaches. Sci China Life Sci, v.56(10), 960-7, 2013.

DRAGHICI, S.; KHATRI, P.; EKLUND, A.C.; SZALLASI, Z. Reliability and reproducibility issues in DNA microarray measurements. Trends Genet, v.22, 101–109, 2006.

ECKER, J.R.; DAVIS, R.W. Inhibition of gene expression in plant cells by expression of antisense RNA. PNAS, v.83, 5372–5376, 1986.

EDDY, S.R. Non-coding RNA genes and the modern RNA world. Nature Reviews, v.2, 919-929, 2001.

EFSA Panel on Genetically Modified Organisms (GMO). Guidance on selection of comparators for the risk assessment of genetically modified plants and derived food and feed. EFSA Journal, 9(5), 2149, 2011.

EGAN, A.N.; SCHLUETER, J.; SPOONER, D.M. Applications of next-generation sequencing in plant biology. Am J Bot. v.99(2), 175-85, 2012.

FAGARD, M.; VAUCHERET, H. (Trans)gene silencing in plants: How Many Mechanisms? Annu. Rev. Plant Physiol. Plant Mol. Biol., 2000.

FIRE, A.; XU, S.; MONTGOMERY, M. K.; KOSTAS, S. A.; DRIVER, S. E.; MELLO, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature, v.391, 806–811, 1998.

FRITSCH, L.; FISCHER, R.; WAMBACH, C.; DUDEK, M.; SCHILLBERG, S.; SCHRÖPER F. Next-generation sequencing is a robust strategy for the high-throughput detection of zygosity in transgenic maize. Transgenic Res, 24(4), 615-23, 2015.

GREGERSEN, P.L.; BRINCH-PEDERSEN, H.; HOLM, P.B. A microarray- based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. Transgenic Res, v.14, 887–905, 2005.

HAGMANN, J.; BECKER, C.; MÜLLER, J.; STEGLE, O.; MEYER, R.C.; WANG G. et al. Century-scale Methylome Stability in a Recently

Diverged Arabidopsis thaliana Lineage. PLoS Genet, v.11(1), p.e1004920, 2015.

HAN, M.; GOUD, S.; SONG, L.; FEDOROFF, N. The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. PNAS, v.101(4), 1093–1098, 2003.

HEINEMANN, J. A.; AGAPITO-TENFEN, S. Z.; CARMAN, J. A. A comparative evaluation of the regulation of GM crops or products containing dsRNA and suggested improvements to risk assessments. Environment International, v.55, 43–55, 2013.

HEINEMANN, J.A.; KURENBACH, B.; QUIST, D. Molecular profiling - a tool for addressing emerging gaps in the comparative risk assessment of GMOs. Environ Int, 37, 1285–1293, 2011.

HOAGLAND, M.B.; STEPHENSON, M.L.; SCOTT, J.F.; HECHT, L.I.; ZAMECNIK, P.C. A soluble ribonucleic acid intermediate in protein synthesis. J Biol Chem, v.231, 241–257, 1958.

HÖSS, S.; REIFF, N.; OTTERMANNS, R.; PAGEL-WIEDER, S.; DOHRMANN, A.B.; TEBBEE, C.C. et al. Risk assessment of the cultivation of a stacked Bt-maize variety (MON89034 × MON88017) for nematode communities. Soil Biology and Biochemistry, v.91, 109–118, 2015.

HUSSEY, S.G.; MIZRACHI, E.; GROOVER, A.; BERGER, D.K.; MYBURG, A.A. Genome-wide mapping of histone H3 lysine 4 trimethylation in Eucalyptus grandis developing xylem. BMC Plant Biol, v.15, 117, 2015.

JACOB, F.; MONOD. J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol, v.3, 318–356, 1961.

JONES-RHOADES, M.W.; BARTEL, D.P.; BARTEL, B. MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology, v.57, 19-53, 2006.

KARAKACH, T.K.; FLIGHT, R.M.; DOUGLAS, S.E.; WENTZELL, P.D. An introduction to DNA microarrays for gene expression analysis, OMICS, v.104, 28–52, 2010.

KAUFMANN, K.; MUIÑO, J.M.; JAUREGUI, R.; AIROLDI, C.A.; SMACZNIAK, C.; KRAJEWSKI, P. et al. Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biol, v.7(4), e1000090, 2009.

KAWAKATSU, T.; KAWAHARA, Y.; ITOH, T.; TAKAIWA, F. A whole-genome analysis of a transgenic rice seed-based edible vaccine against cedar pollen allergy. DNA Res, 20(6), 623-31, 2013.

KHRAIWESH, B.; ARIF, M.A.; SEUMEL, G.I.; OSSOWSKI, S.; WEIGEL, D.; RESKI, R. et al. Transcriptional control of gene expression by microRNA. Cell, v.140, 111–122, 2010.

KNIEF, C. Analysis of plant microbe interactions in the era of next generation sequencing Technologies. Front Plant Sci, v.5, 216, 2014.

KOSIK, K. S. Circles reshape the RNA world. Nature, v.495, 322-324, 2013.

KOVALIC, D.; GARNAAT, C.; GUO, L.; YAN, Y.; GROAT, J.; SILVANOVICH, A. et al. The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology. The Plant Genome, v.5(3), 149-163, 2012.

KROL, J.; LOEDIGE, I., FILIPOWICZ, W. The widespread regulation of microRNA biogenesis, function and decay. Nature Rev. Genet, v.11, 597–610, 2010.

KRUEGER, F.; KRECK, B.; FRANKE, A.; ANDREWS, S.R. DNA methylome analysis using short bisulfite sequencing data. Nat Methods, 9(2), 145-51, 2012.

KURKUBA, K.R.; MONTGOMERY, S.B. RNA Sequencing and Analysis. Cold Spring Harbor Protocols, 2015.

LANDEGREN, U.; KAISER, R.; SANDERS, J.; HOOD, L. A ligase-mediated gene detection technique. Science, v.241, 1077–1080, 1988.

LI, Q.; EICHTEN, S.R.; HERMANSON, P.J.; ZAUNBRECHER, V.M.; SONG, J.; WENDT, J. et al. Genetic perturbation of the maize methylome. Plant Cell, 26(12), 4602-16, 2014.

LIANG, P.; PARDEE, A.B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science, v.257, 967–971, 1992.

LIAO, H.L.; CHEN, Y.; BRUNS, T.D.; PEAY, K.G.; TAYLOR, J.W. et al. Metatranscriptomic analysis of ectomycorrhizal roots reveals genes associated with Piloderma-Pinus symbiosis: improved methodologies for assessing gene expression in situ. Environ Microbiol, 16(12), 3730-42, 2014.

LIPPMAN, Z.; MARTIENSSEN, R. The role of RNA interference in heterochromatic silencing. Nature, v.431, 364–370, 2004.

LISTER, R.; O'MALLEY, R.C.; TONTI-FILIPPINI, J.; GREGORY, B.D.; BERRY, C.C.; MILLAR, A.H. et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell, v.133, 523–536, 2008.

LOCKHART, D.J.; WINZELER, E.A. Genomics, gene expression and DNA arrays. Nature, v.405, 827–836, 2000.

MALHOTRA, K.; FOLTZ, L.; MAHONEY, W.C.; SCHUELER, P.A. Interaction and effect of annealing temperature on primers used in differential display RT-PCR. Nucleic Acids Research, v.26, 854–856, 1998.

MALONE, J.H.; OLIVER, B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biology, v9, 34, 2011.

MATZKE, M.A.; MOSHER, R.A. RNA- directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet, v.15, 394 408, 2014.

MAXAM, A.M; GILBERT, W. A new method for sequencing DNA. PNAS, v.74, 560–564, 1977.

MELCHER, U.; VERMA, R.; SCHNEIDER, W.L. Metagenomic search

strategies for interactions among plants and multiple microbes. Front Plant Sci, v.5, 268, 2014.

MERCER, T.R.; DINGER, M.E.; MATTICK, J.S. Long non-coding RNAs: insights into functions. Nat Rev Genet, 10(3), 155–159, 2009.

MOYROUD, E.; MINGUET, E.G.; OTT, F.; YANT, L.; POSÉ, D.; MONNIAUX, M. et al. Prediction of regulatory interactions from genome sequences using a biophysical model for the Arabidopsis LEAFY transcription factor. Plant Cell, 23(4), 1293-306, 2011.

NAPOLI, C.; LEMIEUX, C.; JORGENSEN, R. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. Plant Cell, v.2, 279–289, 1990.

NOZAWA, M.; MIURA, S.; NEI, M. Origins and evolution of microRNA genes in plant species. Genome Biol Evol, v.4, 230–239, 2012.

OULAS, A.; PAVLOUDI, C.; POLYMENAKOU, P.; PAVLOPOULOS, G.A.; PAPANIKOLAOU, N.; KOTOULAS, G. et al. Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. Bioinform Biol Insights, v.9, 75–88, 2015.

PENG, J.C.; LIN, H. Beyond transposons: The epigenetic and somatic functions of the Piwi-piRNA mechanism. Curr Opin Cell Biol, v.67(25), 190–194, 2013.

PUMPLIN, N.; VOINNET, O. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. Nat Rev Microbiol, v.11(11), 745-60, 2013.

QUAIL, M.A.; SMITH, M.; COUPLAND, P.; OTTO, T.D.; HARRIS, S.R.; CONNOR, T.R. et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics, v.13, 341, 2014.

RAMBANI, A.; RICE, J.H.; LIU, J.; LANE, T.; RANJAN, P.; MAZAREI, M. et al. The Methylome of Soybean Roots during the

Compatible Interaction with the Soybean Cyst Nematode. Plant Physiol, v.168(4), 1364-77, 2015.

REINHART, B.J.; WEINSTEIN, E.G.; RHOADES, M.W.; BARTEL, B.; BARTEL, D.P. MicroRNAs in plants. Genes Dev, v.16, 1616–1626, 2002.

RICARDI, M.M.; GONZÁLEZ, R.M.; ZHONG, S.; DOMÍNGUEZ, P.G.; DUFFY, T.; TURJANSKI, P.G. et al. Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought stress-responsive transcription fator. BMC Plant Biology, v.14, 29, 2014.

ROGERS, K.; CHEN, X. Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. The Plant Cell, v.25, 2383–2399, 2013.

SANGER, F.; NICKLEN, S.; COULSON, A.R. DNA sequencing with chain-terminating inhibitors. PNAS, v.74, 5463–5467, 1977.

SCHADT, E.E.; TURNER S.; KASARSKIS, A. A window into third-generation sequencing. Hum Mol Genet, v.19, R227–R240, 2010.

SCHENA, M.; SHALON, D.; DAVIS, R.W.; BROWN, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, v.270, 467–470, 1995.

SCHNABLE, P.S.; WARE, D.; FULTON, R.S.; STEIN, J.C.; WEI, F.; PASTERNAK, S. et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science, v.326, 1112-1115, 2009.

SCHUPPENER, M.; MÜHLHAUSE, J.; MÜLLER, A. K.; RAUSCHEN, S. Environmental risk assessment for the small tortoiseshell Aglais urticae and a stacked Bt-maize with combined resistances against Lepidoptera and Chrysomelidae in central European agrarian landscapes. Mol Ecol, v.21, 4646-4662, 2012.

SESSITSCH, A.; HARDOIM, P.; DÖRING, J.; WEILHARTER, A.; KRAUSE, A.; WOYKE, T. et al. Functional Characteristics of an Endophyte Community Colonizing Rice Roots as Revealed by Metagenomic Analysis. MPMI, v.25(1), 28–36, 2012.

SHIRAKI, T.; KONDO, S.; KATAYAMA, S.; WAKI, K.; KASUKAWA, T.; KAWAJI, H. et al. Cap analysis gene expression. for high-throughput analysis of transcriptional starting point and identification of promoter usage. PNAS, v.100, 15776–15781, 2003.

STEIJVEN, K.; STEFFAN-DEWENTER, I.; HÄRTEL, S. Testing dose-dependent effects of stacked Bt maize pollen on in vitro-reared honey bee larvae. Apidologie , 1-11, 2015.

STOLC, V.; SAMANTA, M.P.; TONGPRASIT, W.; SETHI, H.; LIANG, S.; NELSON, D.C. et al. Identification of transcribed sequences in Arabidopsis thaliana by using high- resolution genome tiling arrays. PNAS, v.102, 4453–4458, 2005.

TAVERNIERS, I.; PAPAZOVA, N.; BERTHEAU, Y.; DE LOOSE, M.; HOLST-JENSEN, A. Gene stacking in transgenic plants: towards compliance between definitions, terminology, and detection within the EU regulatory framework. Environ Biosafety Res, v.7, 197-218, 2008.

TAYLOR, M.; HARTNELL, G.; NEMETH, M.; LUCAS, D.; DAVIS, S. Comparison of broiler performance when fed diets containing grain from second- generation insect-protected and glifosato-tolerant, conventional control or commercial reference corn. Poultry Science, v.86, 1972-1979, 2007a.

TAYLOR, M.; LUCAS, D.; NEMETH, M.; DAVIS, S.; HARTNELL, G. Comparison of Broiler Performance and Carcass Parameters When Fed Diets Containing Combined Trait Insect-Protected and Glifosato-Tolerant Corn (MON 89034 × NK603), Control, or Conventional Reference Corn. Poultry Science, v.86, 1988–1994, 2007b.

THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, v.408, 796-815, 2000.

VALDÉS, A.; IBÁÑEZ, C.; SIMÓ, C.; GARCÍA-CAÑAS, V. Recent transcriptomics advances and emerging applications in food science. Trends in Analytical Chemistry, v.52, 142–154, 2013.

VAN DIJK, E.L.; AUGER, H.; JASZCZYSZYN, Y.; THERMES, C. Ten years of next-generation sequencing technology. Trends Genet,

v.30(9), 418-26, 2014.

VAUCHERET, H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes & Development, v.20, 759–771, 2006.

VELCULESCU, V.; ZHANG, L.; VOGELSTEIN, B. Serial analysis of gene expression. Science, v.270, 484–487, 1995.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet, v.10, 57–63, 2009.

WARD, J.A.; PONNALA, L.; WEBER, C.A. Strategies for transcriptome analysis in nonmodel plants. Am J Bot, 99(2), 267-76, 2012.

YANG, L.; LIU, Z.; LU, F.; DONG, A.; HUANG, H. SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. Plant J, 47(6), 841-50, 2006.

YANG, L.; WANG, C.; HOLST-JENSEN, A.; MORISSET, D.; LIN, Y.; ZHANG, D. Characterization of GM events by insert knowledge adapted re-sequencing approaches. Nature Scientific Reports, 3, Article number: 2839, 2013.

YANT, L.; MATHIEU, J.; DINH, T.T.; OTT, F.; LANZ, C.; WOLLMANN, H. et al. Orchestration of the Floral Transition and Floral Development in Arabidopsis by the Bifunctional Transcription Factor APETALA2[W][OA]. Plant Cell, 22(7), 2156–2170, 2010.

ZHANG, H.; CHEN, Z.; WANG, X.; HUANG, Z.; HE, Z.; CHEN, Y. Long non-coding RNA: a new player in cancer. J Hematol Oncol, v.6, 37, 2013.

ZHAO, S.; FUNG-LEUNG, W.; BITTNER, A.; NGO, K.; LIU, X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLoS ONE, v.9(1), p.e78644, 2014.

ZHONG, S.; FEI, Z.; CHEN, Y.; ZHENG, Y.; HUANG, M.; VREBALOV, J. et al. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening.

Nature Biotechnology, v.31(2), 154-161, 2013.

ZHU, J. Reconstituting plant miRNA biogenesis. PNAS, v.105(29), 9851-2, 2008.

5. CAPÍTULO I

COMPARATIVE TRANSCRIPTOME OF BT- AND EPSPS-EXPRESSING TRANSGENIC MAIZE

ABSTRACT

Background: The risk assessment of genetically modified plants is a requirement, in which the new crop is assessed for their health and ecological safety, prior to their release in the environment. However, data on possible synergic and antagonistic effects of the introduction of transgenes in the overall gene expression are not required in risk assessment of transgenic crops, due the lack of specific guidelines. Profiling techniques allow the simultaneous measurement and comparison of thousands of cell components without prior knowledge of their identity and, therefore, can be considered useful tools to assess unintended effects arising from genetic modification. Here we report the first results of a transcriptome profiling analysis of a stacked commercial maize hybrid containing insecticidal (CRY1A.105 and CRY2AB2 proteins), and herbicide tolerant (EPSPS protein) traits, in comparison to the single GM and near-isogenic hybrids with the same genetic background.

Results: Our results demonstrate that several biological processes, in both single GM *vs* near-isogenic non-GM comparisons and stacked GM *vs* single and near-isogenic hybrids, are differentially regulated. Single GM hybrids presented major differences in redox (oxidoreductase activity and monooxygenase), protein glycosylation, and phosphorylation and transferase activity pathways. In addition, the stacked GM variety showed to be differentially regulated for pathways related to regulation of transcription, such as DNA binding, protein serine/threonine kinase activity, protein kinase activity, and ATP binding. Finally, our study shows the relevance of investigating the adequate comparators, such as landrace varieties, to possibly disclose differences in GM lines that might fit within the variation observed in non-modified materials for risk assessment.

Conclusion: Our findings indicate the genome changes may influence the overall gene expression and hence several crucial biological processes in the stacked GM maize. Moreover, molecular profiling

could be applied as a useful tool in ways to increase broadness and confidence in risk assessments of single and stacked GM crops. In face of these findings, we recommend that further investigations should be conducted, in order to address the biological relevance of gene expression changes, and that untargeted profiling studies should be taken in account in risk assessment analysis by regulatory authorities.

Key-words: Transcriptomics, RNA-Seq, genetically modified organisms, stacked GMO, risk assessment, molecular profiling.

1. BACKGROUND

The use of agricultural GMOs has been growing steadily over the last decade, although official data are unavailable. In 2014, 181.5 million hectares were grown with GM crops, an annual growth rate of 3.6% from 2013. From that, about 28% (51 million hectares) were grown with GM events containing two or more traits (JAMES, 2014) combined by traditional breeding, referred as "stacked" or "pyramided" events (TAVERNIERS et al. 2008). Regulatory practice within the European Union (EU) consider stacked events as new GM organisms, and additional information on the stability of transgene insertions, expression levels and potential antagonistic or synergistic interactions should be provided prior to marketing (DE SCHRIJVER et al. 2007; EFSA, 2007; AHTEG, 2010).

There is a lack of scientific literature regarding molecular characterization and/or data on synergic and antagonistic effects of stacked GM when compared to parental single GM and near-isogenic non-GM lines (AGAPITO-TENFEN et al. 2014). Studies regarding the possible ecological effects of stacked GM events have been published, but the adequate comparators (single GM and near-isogenic non-GM lines) are often not included in the analysis (HÖSS et al. 2015; SCHUPPENER et al. 2012; HARDISTY et al. 2013). To the best of our knowledge, our previous study was the first one to address the possible changes in endogenous protein expression of stacked events using the adequate comparators, which enabled us to isolate effects arising from the transgene stacking (AGAPITO-TENFEN et al. 2014). Recent discussions about potential risks of stacked events, as well as the opinion of the European Food Safety Authority (EFSA) on this issue, have highlighted the lack of consensus with regard to whether stacked GMOs should be subject to specific assessments (SPÖK et al. 2007), and, therefore, they cannot be considered generally recognized as safe without specific supporting evidence (DE SCHRIJVER et al. 2007).

There is an indication of an increasing evolution in the tools used to assess the risks and unintended effects of GMOs; in particular, there is a growing focus on the development of high-throughput, non-targeted and broad scale approaches (DAVIES, 2010). Recent developments in Next Generation Sequencing (NGS) technologies allows access to more detailed information and refined tools, which leads to more accurate detection of unintended effects of GM plants, with transcriptomics being considered the most complete coverage of potential unintended effects (MEYERS et al., 2004). Transcriptomics,

defined as the complete set of DNA transcripts (i.e. RNAs) in one cell and their quantity, for a specific developmental stage or physiological condition (WANG et al. 2009; VALDÉS et al. 2013), is included among them. Transcriptomics studies have become a promising field of study in the post-genomic era (LOCKHART; WINZELER, 2000), due its ability to quantify changes in transcript expression levels and to support to proteomic studies (WANG et al. 2009; DONG; CHEN, 2013) and, therefore, are suitable profiling method for addressing possible unintended effects arising from GM plants. Transcriptomics techniques have been used to assess transcriptome changes in GM plants (MONTERO et al. 2011; LAMBIRTH et al. 2015; KAWAKATSU et al. 2013; GREGERSEN et al. 2005), but few of them aimed to investigate commercial GM plants (COLL et al. 2009; COLL et al. 2008).

Therefore, the aim of this study was to evaluate changes in transcript expression of single and stacked GM events under highly controlled conditions and to provide insight into the formulation of specific guidelines for the risk assessment of stacked events. We hypothesized that the combination of two transgenes could differentially modulate endogenous transcript expression, which might have an effect on the plant metabolism and physiology. To test these hypotheses, we have used a GM stacked maize genotype containing *cry1A.105/cry2Ab2* and *epsps* cassettes expressing both insect resistance and herbicide tolerance as unlinked traits, as well as genotypes of each single transgene alone, being all maize hybrids in the same genetic background. The seed set of stacked and single GM maize events, as well as the conventional near-isogenic counterpart developed in the same genetic background and a landrace variety, enables the isolation of potential effects derived from stacking two transgenes.

2. METHODOLOGY

2.1. Plant material and growth chamber conditions

Five maize varieties were used in this study (Table 1). Two of them are non-GM maize seeds, the hybrid AG8025 (named here as 'conventional') from Sementes Agroceres and the open pollinated variety Pixurum 5 (named here as 'landrace'). These are named in this study as 'Conv' and 'Land', respectively. Pixurum 5 has been developed and maintained by small farmers in South Brazil for several years (CANCI; BRASSIANI, 2004). The other three varieties are GM and have the same genetic background as the conventional variety since they are produced from the same endogamic parental lines. These are: AG8025RR2 (unique identifier MON-ØØ6Ø3-6 from Monsanto Company, glyphosate herbicide tolerance, Sementes Agroceres); AG8025PRO (unique identifier MON-89Ø34-3 from Monsanto Company, resistance to lepidopteran species, Sementes Agroceres) and AG8025PRO2 (unique identifier MON-89Ø34-3 × MON-ØØ6Ø3-6 from Monsanto Company, stacked event resistant to lepidopteran species and glyphosate-based herbicides, Sementes Agroceres). These are named in this study as RR, Bt and RRxBt, respectively. The AG8025 variety is the hybrid progeny of the single cross between maternal endogamous line "A" with the paternal endogamous line "B". Thus, the used hybrid variety seeds have high genetic similarity (all seeds should be AB genotype in the absence of self pollination). All these five commercial varieties were produced by the aforementioned company/farmers and are commonly found in the market in Brazil.

The cultivation of MON-ØØ6Ø3-6, MON-89Ø34-3, and MON-89Ø34-3 × MON-ØØ6Ø3-6 has been approved in Brazil in 2008, 2009 and 2010 respectively (CTNBIO, 2008; CTNBIO, 2009; CTNBIO, 2010). The stacked hybrid MON-89Ø34-3 × MON-ØØ6Ø3-6 expresses two insecticidal proteins (Cry1A.105 and Cry2Ab2 proteins derived from *Bacillus thuringiensis*, under the regulation of P-35S and FMV promoters, respectively), which are active against certain lepidopteran insect species, and two identical EPSPS proteins (under the regulation of P-Ract1 and P-35S promoters) providing tolerance to the herbicide glyphosate. The novel traits of each parent line have been combined through traditional plant breeding to produce this new hybrid. The experimental approach currently applied for the comparative assessment requires the use of conventional counterpart and the single-event counterparts, all with genetic background as close as possible to the GM

plant, as control (EFSA, 2007; CBD, 2013; CODEX ALIMENTARIUS, 2003).

After the confirmation by PCR of the transgenic events in both single and stacked GM seeds and the absence in the 'Conv' and 'Land' ones (data not shown), the seeds from all the five varieties were grown side by side in growth chambers (EletrolabTM model 202/3) set to 16 h light period and 25 °C (± 2 °C). Seedlings were germinated and grown in Plantmax HT substrate (Buschle & Lepper S.A.) and watered daily. No pesticide or fertilizer was applied. Around 50 plants were grown in climate chambers out of which 30 plants were randomly sampled per maize variety (genotype). The collected samples were separated in three groups of ten plants. The ten plants of each group were pooled and were considered one biological replicate. Maize leaves were collected at V4 stage (20 days after seedling). Leaf pieces were cut out, weighed and placed in 3.8 ml cryogenic tubes before immersion in liquid nitrogen. The samples were kept at −80 °C until mRNA isolation.

Table 1. Transgenic and non-transgenic commercial maize varieties used in this study.

Commercial name	GM event	Transgenes	Sample size	Labeled in this study
AG8025RR2	MON-ØØ6Ø3-6	epsps/epsps	30	RR
AG8025PRO	MON-89Ø34-3	cry1A.105/cry2Ab2	30	Bt
AG8025PRO2	MON-89Ø34-3 x MON-ØØ6Ø3-6	cry1A.105/cry2Ab2 x epsps/epsps	30	RRxBt
AG8025	None	None	30	Conv
Pixurum 5	None	None	30	Land

Note: The unique identifier used is from the Biosafety Clearing House (BCH)

2.2. mRNA isolation and deep sequencing

Total RNA was isolated from approximately 100 mg of frozen leaf tissue using RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. In brief, samples were homogenized with guanidine-isothiocyanate lyses buffer and further purified using silica-membrane. During purification, in-column DNA digestion was performed using RNAse-free DNAse I supplied by Qiagen to eliminate any remaining DNA. The isolated RNA was quantified using NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, USA) and resolved in MOPS 1% denaturing gel.

RNA samples (1μg) were sent to FASTERIS SA (Geneva, Switzerland) for library construction (15 cDNA libraries) and sequencing. The libraries sequencing was conducted using the HiSeq SBS Kit v4 (Illumina®) in an Illumina HiSeq 2500, with number of cycles of 2x125+7 (*paired-end*) in one lane of the HiSeq Flow Cell v4 (Illumina®). Basecalling was performed using the pipelines HiSeq Control Software 2.2.38, RTA 1.18.61.0, CASAVA-1.8.2.

2.3. Pre-processing and read mapping

All low quality reads with FASTq values below 13 were removed, and 5' and 3' adapter, as well as index sequences, were trimmed using the Genome Analyzer Pipeline (Illumina) at Fasteris SA. Moreover, quality control with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was performed and no bases with Phred quality score below 30 were found; thus no additional trimming was necessary. However, RNA-Seq libraries are normally produced using random hexamer primers, which introduces a bias on the sequence composition, usually on the beginning of the reads. Therefore, the first 15 bases of the reads were trimmed in order to avoid the introduction of any further biases on the analyses.

The filtered reads were mapped against the *Zea mays* genome (B73, RefGen_v3, release 25), deposited in the Ensembl plants database (KERSEY et al. 2014), using the TopHat2 v2.1.0 tool (KIM et al. 2013). TopHat2 combines the ability to identify new splicing locations with the direct mapping with known transcripts, producing a precise alignment, even for highly repetitive genomes or with the presence of pseudo genes (KIM et al. 2013).

The library sequencing of one of the Bt replicates showed a high value of duplicated reads, which generated mapping errors and

precluding its use in the analysis of differential expression. The Illumina sequencing consists of three main steps: i) library preparation; ii) cluster amplification, sequencing by synthesis and image analysis; and iii) post-processing of data. There is a possibility of bias in all steps, for instance, the suppression of sequencing of reads with high GC content in high density clusters in the Illumina flow-cell, protocols of amplification and the enzymes used for PCRs, and also the replacement of sequencing kits, protocols and instruments. In addition, biases can vary among different laboratories, and even among lanes of the same flow-cell (AIRD et al. 2011). Therefore, the aforementioned replicate was excluded from the analyses in order to avoid any bias in the expression analysis; thus, the sequencing data for the Bt variety are from two biological replicates.

2.4. Differential expression of mRNAs

For the differential expression analysis, the following comparisons were conducted: Conv *vs* RR, Conv *vs* Bt, Conv *vs* RRxBt, Conv *vs* Land, RR *vs* Bt, RR *vs* RRxBt, RR *vs* Land, Bt *vs* RRxBt e RRxBt *vs* Land. The analysis were conducted using the Tuxedo package (TRAPNELL et al. 2012), which is able to assemble transcripts, estimate their abundance, and test the expression and differential regulation of RNA-Seq libraries. Moreover, it estimates the relative abundance of transcripts based on the amount of reads that support each other, taking in consideration possible biases in protocols of library preparation.

Tuxedo has different tools integrated in its package, which were used in the present study in different steps in order to conduct the differential expression analysis. After the read mapping against the *Zea mays* genome (B73, RefGen_v3, release 25) using TopHat2, Cufflinks v.2.2.0 tool was used to assemble the transcripts for each of the 14 libraries. Following, Cuffmerge tool was used to merge the previously assembled transcripts in a single annotation file. The last used tool was Cuffdiff, which is able to detect genes and transcripts that are differentially expressed between samples, as well as to detect differential splicing and promoter use. For plotting the results, the CummeRbund v.2.7.2 tool was used, together with the language and statistical environment R (R CORE TEAM, 2015).

2.5. Pathway enrichment analysis

The differentially expressed transcripts for each comparison were submitted to enrichment analysis using the online tool agriGO v1.2 (DU et al. 2010), using the tool Single Enrichment Analysis (SEA), with the following parameters: 1) Selected species: *Zea mays* ssp V5a; 2) Statistical test method: Hypergeometric; 3) Multi-test adjustment method: Hochberg (FDR); 4) Significance level of 0.05; 5) Minimum number of 5 mapping entries; and 6) Gene ontology type: Plant GO Slim. Following, the online tool REVIGO (SUPEK et al. 2011) was used to remove the redundant Gene Ontology (GO) terms. Only significant GO terms (False Discovery Rate (FDR) values < 0.05) were used, with the following parameters: 1) Allowed similarity: small (0.5); 2) Database with GO term sizes: *Zea mays*; e 3) Semantic similarity measure: SimRel.

3. RESULTS AND DISCUSSION

3.1. Illumina sequencing and alignment to reference genome

Our transcriptome sequence data included gene expression profiling of five different maize varieties. The RNA-Seq method applied is able to generate absolute information of gene expression, instead of relative expressions as the microarrays (ZHENG et al. 2013). We sequenced 14 cDNA libraries, with three biological replicates for all the varieties (Conv, RR, RRxBt and Land) except for Bt (only two biological replicates). The Illumina sequencing generated 207,490,337 paired reads, each of which was 125 bp in length (which were reduced to 110 bp after trimming), encompassing 64.03 Gb of sequence data (Table 2). The majority of the reads aligned to nuclear regions of the *Zea mays* genome (B73, RefGen_v3, release 25) (73.63%), followed by chloroplastidial regions (1.00%) and mitochondrial regions (0.61%) (Table 2).

From all identified transcripts, 55,022 belong to nuclear regions, 30 to chloroplastidial and 33 to mitochondrial regions. The largest transcript identified showed 14,668 bp and the smallest one showed 122 bp. The highest FPKM value (Fragments Per Kilobases per Million) identified was 15778.8, belonging to a nuclear transcript (Table 3). From all identified transcripts, 39,569 were present in all varieties, 1,071 were exclusively present in the Conv samples, 1,225 in the RR samples, 891 in the Bt samples and 1,115 in the RRxBt samples. The remaining transcripts were present in combinations of two or three varieties and are shown in Figure 1. The Land variety was not included in this analysis due its large difference in transcript abundance and expression from the other varieties. Therefore, this difference will be discussed further in this paper.

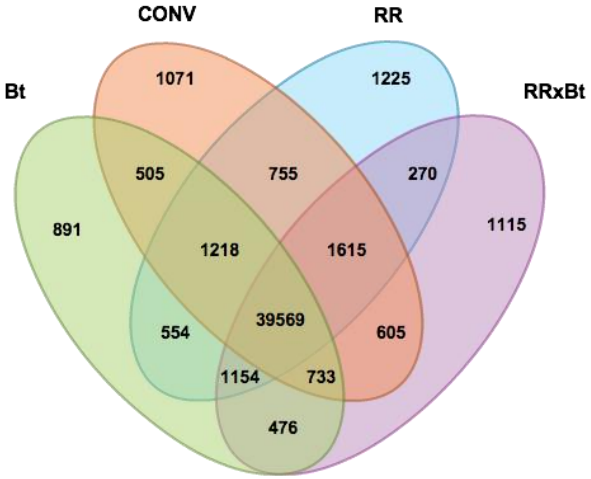
Table 2. Summary statistics for the cDNA libraries based on RNA-Seq data.

Sample	Total paired reads (2 x 110 bp)	Nuclear		Chloroplast		Mitochondria	
		Aligned pairs	Alignment rate (%)	Aligned pairs	Alignment rate (%)	Aligned pairs	Alignment rate (%)
RR 1	13,590,508	9,812,836	72.20	267,102	1.97	160,169	1.18
RR 2	15,125,600	11,126,349	73.56	164,533	1.09	95,670	0.63
RR 3	15,195,992	11,391,460	74.96	116,187	0.76	68,611	0.45
Bt 1	17,563,707	13,283,609	75.63	140,847	0.80	85,337	0.49
Bt 2	15,353,139	11,163,531	72.71	187,909	1.22	119,075	0.78
RRxBt 1	15,570,997	11,288,949	72.50	177,662	1.14	110,439	0.71
RR x Bt 2	14,235,425	9,910,421	69.62	243,239	1.71	149,319	1.05
RR x Bt 3	15,483,875	10,996,232	71.02	225,663	1.46	122,595	0.79
Conv 1	13,963,298	10,750,253	76.99	70,330	0.50	38,161	0.27
Conv 2	12,496,072	9,475,933	75.83	85,537	0.68	43,475	0.35
Conv 3	12,600,659	9,726,236	77.19	66,758	0.53	34,576	0.27
Land 1	14,213,813	10,123,354	71.22	131,163	0.92	137,877	0.97
Land 2	15,993,075	11,735,548	73.38	95,939	0.60	49,999	0.31
Land 3	16,104,177	11,907,200	73.94	88,796	0.55	44,547	0.28
Average	14,820,738	10,906,565	73.63	147,262	1.00	89,989	0.61

Table 3. Alignment statistics of the cDNA libraries with *Zea mays* reference genome.

Alignment characteristics	Type of genome		
	Nuclear	**Chloroplastidial**	**Mitochondrial**
Number of identified transcripts	55022	30	33
Length of largest transcript (bp)	14668	6595	3766
Length of smallest transcript (bp)	122	246	318
Max FPKM value - Conv	15709.0	48.8	2.2
Max FPKM value - RR	12032.2	24.2	1.9
Max FPKM value - Bt	13934.4	33.0	1.2
Max FPKM value - RRxBt	14809.5	29.8	4.1
Max FPKM value - Land	15778.8	52.7	4.3

Figure 1. Venn diagram of identified transcripts in each of the studied varieties.



Source: Author

3.2. Differential expression of single GM varieties

3.2.1. Quality assessment and expression profile

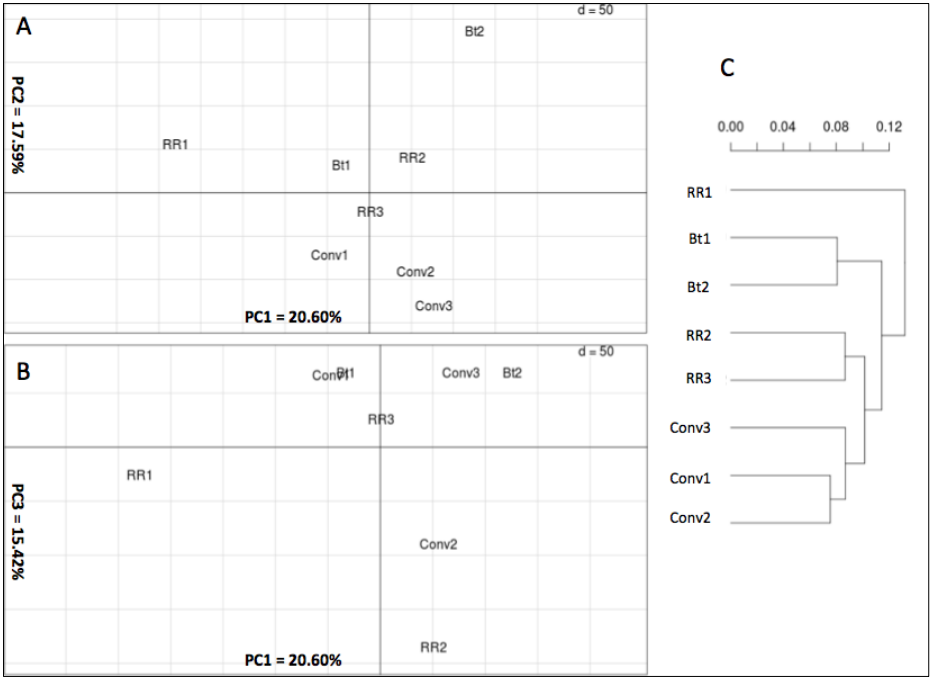
In order to analyze the difference in transcript expression between single GM varieties and their conventional isogenic line, the following comparisons were performed: Conv *vs* Bt and Conv *vs* RR. We have performed a dimensionality reduction strategy, the PCA analysis, and a cluster analyze, summarized in a dendrogram, which includes Conv, RR and Bt samples, in order to explore the relationship between these varieties. Results from both analyses are represented 2-dimensionally using their PC1, PC2 and PC3 scores (in two separated plots) (Figure 2A and B). PCA showed a separation in the first plot (PC1 x PC2) of the non-GM samples from the single GM samples, which explained the second most variability in the dataset (17.59% - PC2). The other plot (PC1 x PC3) didn't show a clear separation of any group of samples. The dendrogram in Figure 2C clearly shows a separation of the three varieties, with their biological replicates clustering together (with the exception of the RR1 biological replicate). The RR1 sample showed a different pattern of transcripts expression, which could be either by errors in the technique (AIRD et al. 2011) or by a biological variance in the samples. Since each replicate consists of a 10-plant pool, it is unlikely that this group of plants would show a high differentiation from the other pools.

Results of the multivariate analysis showed that the single GM varieties grouped separately from their conventional near-isogenic line. Barros et al. (2010) conducted a proteomics study using the same RR transgenic event utilized in the present study and a different Bt event (MON810 event - MON-ØØ81Ø-6) in the same genetic background. The authors found that RR maize samples were grouped separately from Bt and conventional samples grown at field conditions. Similar patterns were also observed in the same study for microarray and gas chromatographic/mass spectrometric metabolite profile analysis. Coll et al. (2010) showed that the majority of the detected quantitative variation in their transcriptomics study was related to the environment or the plant genetic background; however, transgenic and their conventional near-isogenic lines are frequently observed in separated groups by PCA.

Agapito-Tenfen et al. (2013) have also conducted a proteomic analysis of single GM maize and its near-isogenic line, but using another Bt event (MON-ØØ81Ø-6) grown under two different agroecosystems. Similarly, the authors found that the environment was the major source

of influence to the maize proteome; however, the different genotypes (Bt and near-isogenic conventional line) accounted for the second major source of variability, grouped in different clusters in the PCA.

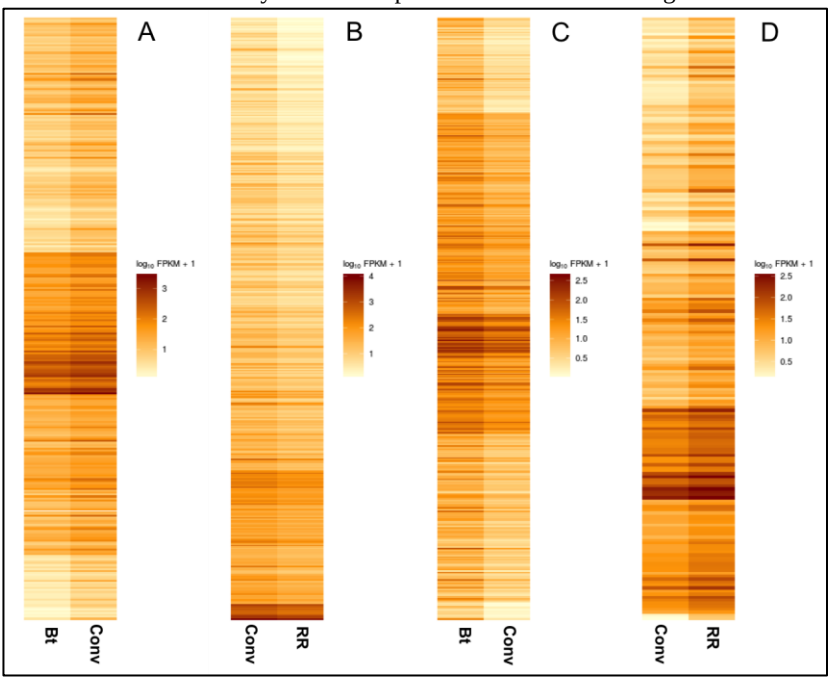
Figure 2. Principal Component Analysis and dendrogram of single GM (RR and Bt) and conventional varieties. A) PCA plot of PC1 x PC2. B) PCA plot of PC1 x PC3. C) Dendrogram of single and conventional varieties. The dendrogram was construct with the differentially expressed transcript, using Jensen-Shanon divergence (JSD) (GROSSE et al. 2002).



Source: Author

The differential expression analysis between Conv and Bt showed that 475 transcripts were up regulated in the Bt samples, while 315 were down regulated (*p-adjusted* value < 0.05). In addition, 10 differentially expressed transcripts were found exclusively in the Bt samples and four in the Conv samples. For the Conv *vs* RR comparison, 200 transcripts were up regulated in the RR samples, while 434 were down regulated (*p-adjusted* value < 0.05). Moreover, seven differentially expressed transcripts were found exclusively in the RR samples and four in the Conv samples. The profiles of up and down regulated transcripts for both Conv *vs* Bt and Conv *vs* RR comparisons are shown in Figure 3.

Figure 3. Heatmap of up and down regulated transcripts for the comparisons of single GM varieties and their near-isogenic conventional variety. A and B) Profile of up regulated genes for the Conv variety in comparison with the single varieties. C and D) Profile of down regulated genes for the Conv variety in comparison with the single varieties.



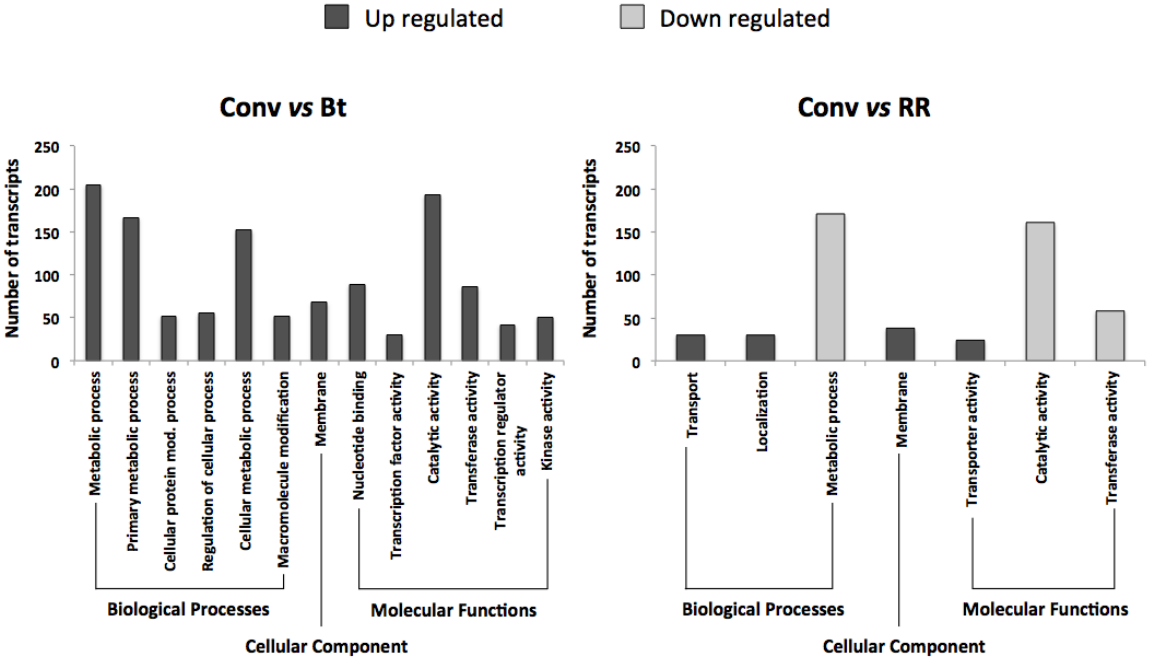
3.2.2. Disturbance in Redox, Post-Translational Modifications (PTM) and other biological processes

An enrichment analysis was performed in order to rank associations between the set of differentially regulated transcripts representing metabolic pathways with a respective statistical probability. All significantly enriched GO terms (FDR < 0.05) for Conv *vs* Bt and Conv *vs* RR comparisons are shown in Figure 4. The main point of enrichment analysis is that if a biological process is differentially modulated in a dataset, the co-functioning genes should have a higher (enriched) potential to be selected as a relevant group for the study. This approach shifts the analysis from an individual gene-oriented to a relevant gene group-based analysis (HUANG et al. 2008).

The enrichment analysis for the Conv *vs* Bt comparison showed that the differentially regulated transcripts were all up regulated in the transgenic samples when compared to the conventional samples. The majority of differentially regulated transcripts for the Conv *vs* RR comparison were down regulated in the transgenic samples when compared to the conventional ones. Moreover, most of the transcripts were assigned to GO terms of Biological Processes, followed by Molecular Functions. To further discuss these findings, we are going to focus on the differentially regulated biological processes for each of the comparisons.

The most altered biological processes in the Conv *vs* Bt comparison were metabolic process (GO:0008152), cellular metabolic process (GO:0044237) and primary metabolic process (GO:0044238), all of them being up regulated in the transgenic samples. The five transcripts with the highest fold-change assigned to these GO terms are shown in Table 4. We have used the logarithm base 2 of the fold-change (Log2FC) as a measure of expression changes. Log2FC is the most widely used alternative transformation of the ratio, once it has the advantage of producing a continuous spectrum of values and treating up- and down-regulated genes in a similar fashion. Therefore, a gene up-regulated by a factor of 2 has a Log2FC of 1, a gene down-regulated by a factor of 2 has a Log2FC of −1, and a gene expressed at a constant level (with a FC of 1) has a Log2FC equal to zero (QUACKENBUSH, 2002).

Figure 4. Enrichment analysis of the differentially regulated transcripts for the comparisons of single transgenic varieties with their near isogenic line. Differentially regulated pathways (FDR < 0.05) found in the agriGO online tool were submitted to removal of redundant GO terms using REVIGO online tool.



Source: Author

Table 4. Description and related GO terms of the five transcripts with the highest fold-change assigned to biological processes pathways in the Conv *vs* Bt enrichment analysis.

Ensembl Plants ID	Description	Related GO Terms	Log2FC
GRMZM2G019515_T01	Putative - Disulfide oxidoreductase/ monooxygenase/ oxidoreductase (UniProt:B6SKY5)	GO:0050661 NADP or NADPH binding GO:0050660 FAD binding GO:0004499 flavin-containing monooxygenase activity GO:0043087 regulation of GTPase activity GO:0055114 oxidation reduction GO:0016491 oxidoreductase activity	4.56
GRMZM2G050915_T01	Uncharacterized protein (UniProt:A0A096QMG9)	GO:0055114 oxidation reduction GO:0016491 oxidoreductase activity	4.49
GRMZM2G101405_T01	Uncharacterized protein (UniProt:A0A096RL33)	GO:0003700 transcription factor activity GO:0043565 sequence-specific DNA binding GO:0045449 regulation of transcription	3.04
GRMZM2G176630_T01	Uncharacterized protein (UniProt:A0A096T0H9)	GO:0006486 protein amino acid glycosylation GO:0016020 membrane GO:0005529 sugar binding GO:0050825 ice binding GO:0042309 homoiothermy GO:0050826 response to freezing GO:0008378 galactosyltransferase activity	2.99
GRMZM2G124042_T01	Uncharacterized protein (UniProt:A0A096S0Z8)	GO:0008415 acyltransferase activity	2.72

In the comparison of Bt samples with its conventional near isogenic variety, two out of the five most up regulated transcripts are involved in redox pathways (GO:0055114 oxidation reduction and GO:0016491 oxidoreductase activity). The most up regulated transcript is a Disulfide oxidoreductase enzyme, involved in Disulfide transfer pathways. Those pathways take part in the endoplasmic reticulum (ER) and chloroplast and play crucial roles in the development of protein storage organelles and the biogenesis of chloroplast, respectively (ONDA, 2013). Disulfide oxidoreductase enzymes play major roles in introducing disulfide bonds into polypeptides, a key step in oxidative protein folding, which covalently link the side chains of pairs of Cys residues, impart thermodynamic and mechanical stability to proteins, and control protein folding and activity (FASS, 2012). Zolla et al. (2008), in their proteomic study comparing another Bt and near isogenic non-Bt maize varieties, found a differentially regulated protein (protein disulfide isomerase) involved in the same pathways.

Moreover, Disulfide oxidoreductase enzyme transcript is also assigned to a monooxygenase function, particularly involved in the flavin-containing monooxygenase pathway (GO:0004499). Flavin-containing monooxygenases (FMOs) are involved in the process of nonnutritional foreign compounds metabolism known as xenobiotics. Their main function is to add molecular oxygen to lipophilic compounds, making them soluble to ensure rapid excretion (ESWARAMOORTHY et al. 2006). Researches in plants have showed specific functions for plant FMOs in auxin biosynthesis (ZHAO et al. 2001) and also a role in pathogen defense (BARTSCH et al. 2006). Agapito-Tenfen et al. (2013), in their proteomics analysis of Bt and near isogenic non-GM counterpart, have found an over expression in GM plants of 2-cysteine peroxiredoxin BAS1 (2-CP) proteins, which are highly sensitive to inactivation by reactive oxygen species, whereas 2-CP detoxifies H_2O_2 under normal conditions as well as under oxidative stress. The same results were also found in their later paper using the same varieties of this present study (AGAPITO-TENFEN et al. 2014).

The analysis also showed that the GRMZM2G176630_T01 transcript was assigned to a differentially modulated biological process of protein amino acid glycosylation (GO:0006486). Glycosylation is one of the most abundant posttranslational modifications (PTM) of proteins, and it plays a major role in protein folding, interaction, stability, mobility and signal transduction (ROTH et al. 2012). Agapito-Tenfen et al. (2013) found in their proteomic study that several proteins were identified more than one time in different spots. These proteins are

considered to represent different protein isoforms resulting from posttranslational modifications, which introduce changes of molecular weight (MW) and/or isoelectric point (pI).

Although 2-D electrophoresis coupled with MS/MS peptide identification is capable of detecting protein isoforms due to changes in MW and pI, it does not detect which PTM has occurred in the isoform, unless blotting techniques for specific modifications are also applied (GRAVEL et al. 1994). The analysis of protein glycans is complicated by their vast variety and the large number of potential glycosylation combinations: even a single protein can undergo a number of *N*- and *O*-glycosylations. Thus, detailed characterization of glycans often requires the use of several methods, with the most reliable analytical tools currently available being chromatography and mass spectrometry (ROTH et al. 2012) that is out of the scope of the present work.

In the Conv *vs* RR comparison, the most altered biological process was the metabolic process (GO:0008152), but unlike in the Bt samples, this pathway was down regulated in the transgenic (RR) samples. The five transcripts with the highest fold-change assigned to this GO term are shown in Table 5. Two out of the five most down regulated transcripts in the RR samples are involved in transferase activity of hexosyl groups (GO:0016758), one of them identified as a Cis-zeatin *O*-glucosyltransferase. Zeatin is a cytokinin, an essential plant hormone, promoting cell division and differentiation in tissue culture, besides regulating several other events in whole plants such as bud formation, leaf expansion, delay of senescence and seed germination (MARTIN et al. 2001). O-glucosyltransferases are responsible for converting zeatin in its *O*-glucoside form, which, although temporarily inactivates the hormone, it also protects zeatin from cytokinin oxidases/ dehydrogenases that can degrade zeatin but not its *O*-glucoside form (ARMSTRONG, 1994).

Another down regulated transcript identified in the RR samples, GRMZM2G442404_T01, was assigned to three protein phosphorylation processes (GO:0008287 protein serine/threonine phosphatase complex, GO:0004722 protein serine/threonine phosphatase activity, GO:0006470 protein amino acid dephosphorylation). Protein phosphorylation is one of the mostly characterized modifications involved in regulation of transcription (KARIN, 1994), and it is known to control functioning of transcription factors (SUBOTA et al. 2007). Phosphorylation can modify, through structural changes, DNA-binding activity and modulating functions of transcription factors (WHITMARSH; DAVIS, 2000). Interestingly, another transcript in the RR samples showed to be

down regulated (GRMZM2G301908_T01, identified as a Fasciclin-like arabinogalactan protein 10), and it was assigned to significantly modulated DNA-binding and regulation of transcription pathways, which can be affected by protein phosphorylation changes.

Our results showed that several metabolic pathways were differentially regulated in the single GM events in comparison to the near-isogenic conventional line, with most of them being up regulated in the Bt samples and down regulated in the RR samples. These pathways were responsible for several key plant biological processes, such as protein folding, protein glycosylation and phosphorylation, xenobiotics and signal transduction. Overall, the analyses show that the transcriptomic profiles of single GM events (expressing either insecticidal CRY or herbicide tolerant EPSPS proteins) are different from their near-isogenic conventional line.

Table 5. Description and related GO terms of the five transcripts with the highest fold-change assigned to biological processes pathways in the Conv *vs* RR enrichment analysis.

Ensembl Plants ID	Description	Related GO Terms	Log2FC
GRMZM2G334336_T01	Uncharacterized protein (UniProt:A0A096T9Z9)	GO:0016758 transferase activity, transf. hexosyl groups	-3.26
GRMZM2G010044_T03	Uncharacterized protein (UniProt:K7U8M1)	GO:0008652 cellular amino acid biosynthetic process GO:0016836 hydrolyase activity GO:0051539 4 iron, 4 sulfur cluster binding	-2.87
GRMZM2G301908_T01	Fasciclin-like arabinogalactan protein 10 (UniProt:K7U4B2)	GO:0005634 nucleus GO:0050826 response to freezing GO:0016068 type I hypersensitivity GO:0006355 regulation of transc., DNA-dependent GO:0050825 ice binding GO:0042309 homoiothermy GO:0003677 DNA binding GO:0004879 ligand-dependent nuclear receptor act.	-2.64
GRMZM2G442404_T01	Uncharacterized protein (UniProt:B6TEB8)	GO:0003824 catalytic activity GO:0008287 protein serine/threonine phosph. complex GO:0004722 protein serine/threonine phosph. act. GO:0006470 protein amino acid dephosphorylation	-2.49
GRMZM2G120016_T01	Cis-zeatin *O*-glucosyltransferase (UniProt:A0A096RYM5)	GO:0016758 transferase activity, transf. hexosyl groups	-2.31

3.3. Differential expression of staked GM varieties

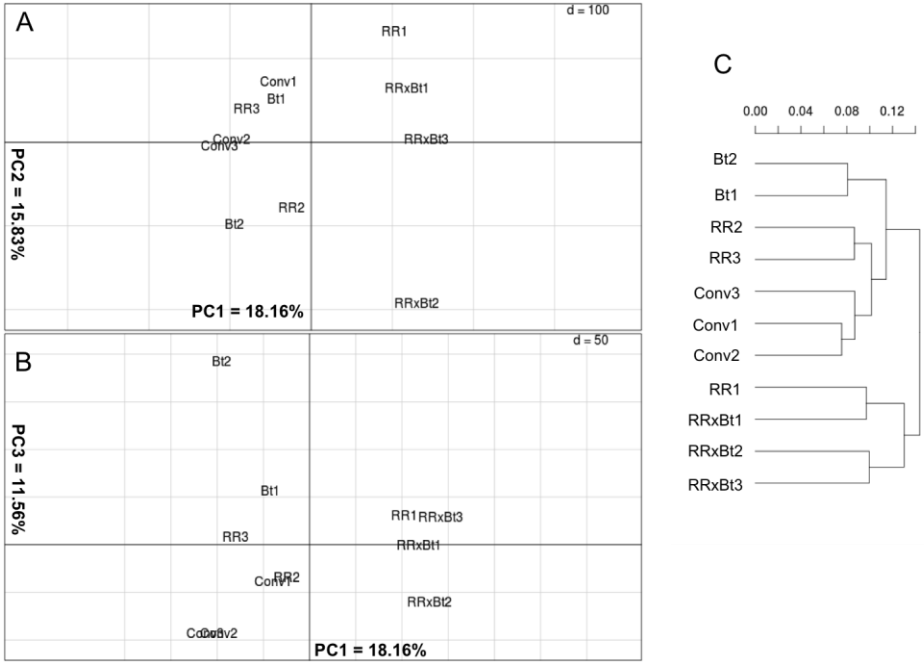
3.3.1. Quality assessment and expression profile

In order to assess the possible differences in transcript expression in the stacked GM variety compared to its single GM isogenic lines and conventional near isogenic line, the following comparisons were conducted: Conv *vs* RRxBt, RR *vs* RRxBt and Bt *vs* RRxBt. The PCA plots and dendrogram for the stacked samples and its near isogenic single GM and conventional varieties is shown in Figure 5. A clear separation of the RRxBt samples in PC1 was revealed by the first plot in PCA analysis (PC1 x PC2), accounting for the majority of the variation in the dataset (18.16%). In the second plot (PC1 x PC3), the RRxBt samples were again separate plotted, accounting for the same amount of variation. Lastly, the dendrogram in Figure 5C also clearly shows a separation of all varieties, with the stacked GM one being more distant from its near isogenic single GM and conventional varieties.

Similar results using the same GM varieties were found when analyzing their protein profile. Stacked GM plants were grouped separately from the other samples (RR, Bt, Conventional and Landrace). The first plot of the PCA analysis (PC1 x PC2) showed a grouping of stacked GM and Bt samples from the other, which accounted for 28.1% of the variation in the proteomics dataset. Although Bt and RRxBt samples were grouped in the same plot, their clusters were distant from each other (AGAPITO-TENFEN et al. 2014). In the present work, the Bt samples were not grouped together in the same plot of the RRxBt samples.

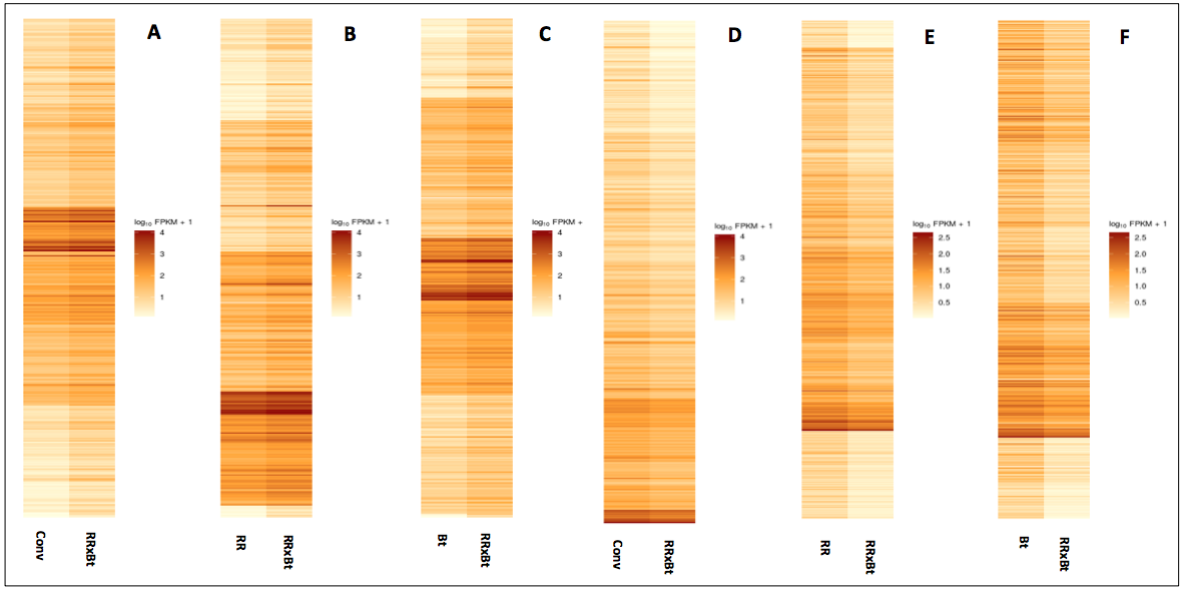
The differential analysis of Conv *vs* RRxBt showed that 375 transcripts were up regulated in the transgenic samples, while 1548 were down regulated (*p-adjusted* value < 0.05). In addition, six differentially expressed transcripts were found exclusively in the RRxBt samples and 16 in the non-GM samples. For the RR *vs* RRxBt comparison, 369 transcripts were up regulated in the transgenic samples and 706 were down regulated. Moreover, one transcript was exclusive from the RRxBt samples and seven from the RR samples. Finally, for the Bt *vs* RRxBt comparison, 389 transcripts were differentially up regulated in the stacked GM samples, while 1293 transcripts were down regulated. Two transcripts were exclusively found in the RRxBt samples, and 14 in the Bt samples. The profiles of up and down regulated transcripts for all comparisons are shown in Figure 6.

Figure 5. Principal Component Analysis and dendrogram of stacked GM and its near isogenic single GM and conventional varieties. A) PCA plot of PC1 x PC2. B) PCA plot of PC1 x PC3. C) Dendrogram of stacked GM, single GM and conventional varieties. The dendrogram was construct with the differentially expressed transcript, using Jensen-Shanon divergence (JSD) (GROSSE et al. 2002).



Source: Author

Figure 6. Heatmap of up and down regulated transcripts for the comparisons of stacked GM variety and their near-isogenic single GM and conventional varieties. A, B and C) Profile of up regulated genes for the RRxBt variety in comparison with the single GM and near-isogenic non-GM varieties. D, E and F) Profile of down regulated genes for the RRxBt variety in comparison with the single GM and near-isogenic non-GM varieties.



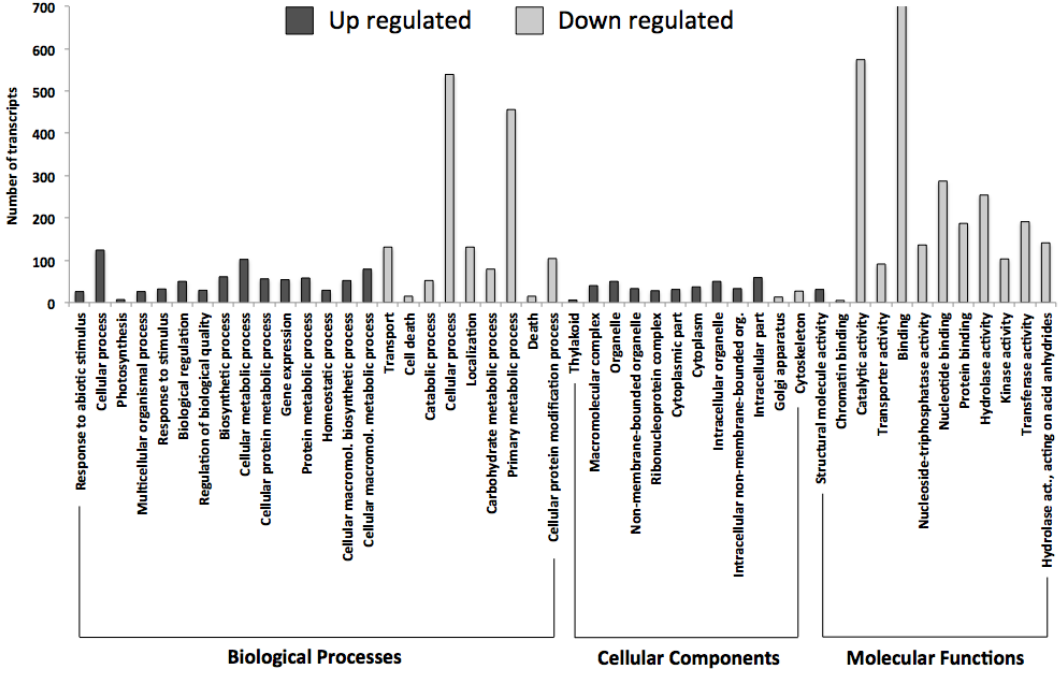
Source: Author

3.3.2. Disturbance in Transcription Factor (TF) pathways in stacked GM variety

We have also performed an enrichment analysis in order to rank associations between the set of differentially regulated transcripts representing metabolic pathways with a respective statistical probability. All significantly enriched GO terms (FDR < 0.05) for Conv *vs* RRxBt, RR *vs* RRxBt and Bt *vs* RRxBt comparisons are shown in Figure 7, 8 and 9, respectively. The enrichment analysis for the Conv *vs* RRxBt and Bt *vs* RRxBt comparisons showed similar patterns, where the differentially regulated transcripts were assigned to GO terms of Molecular Functions, followed by Biological Processes and Cellular Components, with the majority being down regulated in the stacked GM samples. The RR *vs* RRxBt comparison showed that the transcripts were mostly assigned to Biological Processes, followed by Molecular Functions and Cellular Components. Unlike the Conv *vs* RRxBt and Bt *vs* RRxBt comparisons, similar numbers of transcripts were up and down regulated in the stacked GM samples. Similarly to the comparisons of single GM *vs* near-isogenic line, the differentially modulated biological processes will be further discussed for each comparison.

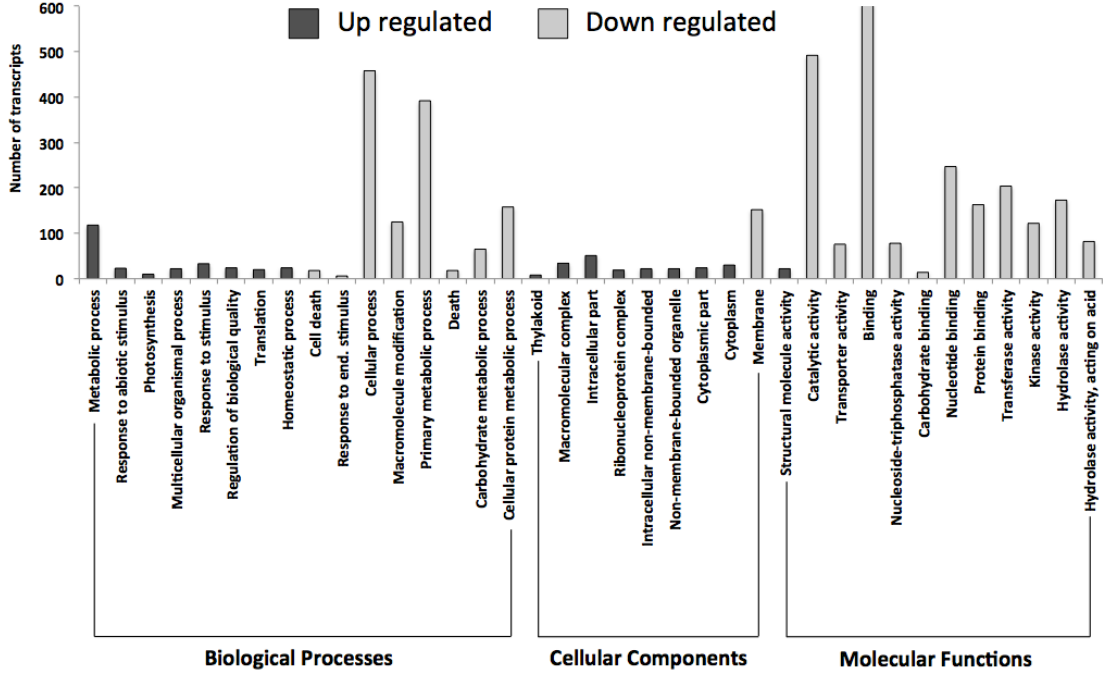
The most altered biological processes in the Conv *vs* RRxBt and Bt *vs* RRxBt comparisons were cellular process (GO:0009987) and primary metabolic process (GO:0044238), both of them being down regulated in RRxBt samples. The five transcripts with the highest fold-change assigned to these GO terms are shown in Table 6 and Table 7, respectively. For the RR *vs* RRxBt comparison, the most altered biological process was cellular process (GO:0009987) as well, with a similar down regulation in the RRxBt samples. The five transcripts with the highest fold-change assigned to this GO term are shown in Table 8.

Figure 7. Enrichment analysis of the differentially regulated transcripts for Conv *vs* RRxBt comparisons. Differentially regulated pathways (FDR < 0.05) found in the agriGO online tool were submitted to removal of redundant GO terms using REVIGO online tool.



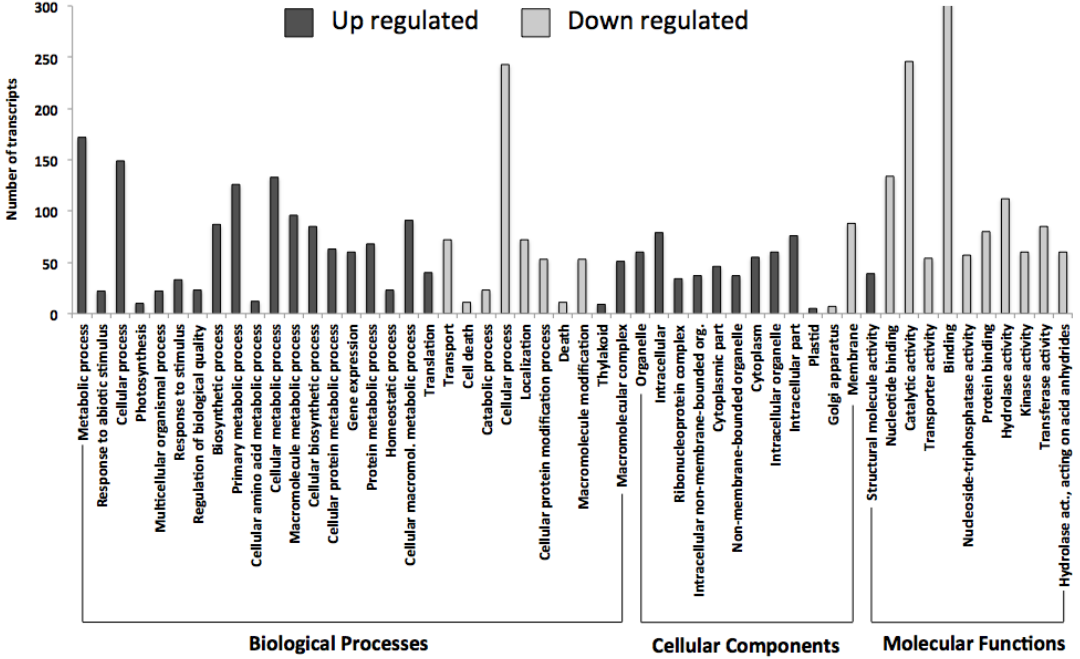
Source: Author

Figure 8. Enrichment analysis of the differentially regulated transcripts for Bt *vs* RRxBt comparisons. Differentially regulated pathways (FDR < 0.05) found in the agriGO online tool were submitted to removal of redundant GO terms using REVIGO online tool.



Source: Author

Figure 9. Enrichment analysis of the differentially regulated transcripts in RR *vs* RRxBt comparisons. Differentially regulated pathways (FDR < 0.05) found in the agriGO online tool were submitted to removal of redundant GO terms using REVIGO online tool.



Source: Author

In the Conv *vs* RRxBt comparison, two transcripts were identified as being Putative AP2/EREBP transcription factor superfamily proteins (GRMZM2G438202_T01 and GRMZM2G020150_T01). These transcripts, together with a third one (GRMZM2G319167_T01), were assigned to pathways related to regulation of transcription (GO:0003700, GO:0006355, GO:0003677). Transcription factors (TFs) play important roles in maintaining expression of protein genes in the genomes, with the AP2/EREBP (APETALA2/ethylene-responsive element-binding protein) being a large family of TF genes (SHARONI et al. 2011). Several members of the AP2/EREBP superfamily play important roles in a range of biological processes, such as developmental and growth processes (APETALA2 - OKAMURO et al., 1997), regulation of primary and secondary metabolism, regulating jasmonate-responsive genes (ORCA3 - VAN DER FITS; MEMELINK, 2001) and disease resistance (ZHOU et al. 1997). In addition, AP2/EREBP members also play a role in abiotic stress, more specific in drought tolerance (KIZIS et al. 2001).

Although it is known that abscisic acid (ABA) is a major physiological signal that induces drought responses in plants (MUNDY; CHUA, 1988), other studies have shown an ABA-independent pathway capable of triggering drought response through members of AP2/EREBP family of transc ription factors (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 1994). Over-expression of AP2/EREBP domain-containing transcription factor gene in transgenic *Arabidopsis thaliana* showed enhanced salt and freezing stresses (TANG et al. 2011), affected growth and development (ZHOU et al. 2013), and enhanced disease resistance and salt tolerance in tobacco (GUO et al. 2004).

The Bt *vs* RRxBt comparison also showed a down-regulated transcript in the RRxBt samples, involved in regulation of transcription, the WRKY DNA-binding domain superfamily protein (AC209050.3_FGT003). The transcription of *WRKY* genes is strongly and rapidly up regulated in response to wounding, pathogen infection or abiotic stresses in numerous plant species (EULGEM et al. 2000), showing the role of WRKY TFs as both positive and negative regulators of gene expression (reviewed in EULGEM; SOMSSICH, 2007). Similarly to the AP2/EREBP TFs, WRKYs have also shown to play major roles in drought stress mechanisms in plants (review in TRIPATHI et al. 2014).

Additionally, Asai et al. (2002) reported a role of WRKY factors associated with defense-induced mitogen-activated protein

kinase (MAPK) signaling cascades. *A. thaliana* WRKY factors have been identified as components of a MAPK pathway that confers resistance to both bacterial and fungal pathogens. A MAPK cascade is composed of a MAP kinase (MAPK), a MAP kinase kinase (MAP2K), and a MAP kinase kinase kinase (MAP3K). The mode of action of MAPK cascades normally involves a stimulus of plasma membrane receptors, which then activate MAP3Ks or, in some cases, MAP4Ks. Subsequent phosphorylations ensue as MAP3Ks activate downstream MAP2Ks, leading to an activation of MAPKs. Lastly, MAPKs then target various effector proteins in the cytoplasm or nucleus, which include other kinases, enzymes, or transcription factors (RODRIGUEZ et al. 2010; KRYSAN et al. 2002; KHOKHLATCHEV et al. 1998). Interestingly, another down-regulated transcript identified in RRxBt samples is a Putative MAPKKK family protein kinase (GRMZM2G305066_T01). MAPK cascades are shown to be involved in signaling a range of biotic and abiotic stresses, such as wounding and pathogen infection, temperature stress or drought, as well plant hormones, such as ethylene and auxin (WRZACZEK; HIRT, 2001).

Similar to the comparisons with Conv and Bt samples, in the RR *vs* RRxBt, three down regulated transcripts in the RRxBt samples have being assigned to regulation of transcription pathways (GO:0004674 protein serine/threonine kinase activity; GO:0005524 ATP binding; GO:0005515 protein binding; GO:0004672 protein kinase activity). One of these transcripts was identified as a Putative leucine-rich repeat receptor-like protein kinase family protein (GRMZM5G867798_T01). Ligand receptor-like kinases (RLKs) signaling pathways are crucial regulators of cell specification in plants (DE SMET et al. 2009). The majority of plant RLKs belong to the family of serine/threonine kinases, and most have extracellular Leucine-rich repeats (LRRs; BECRAFT, 1998). RLKs have been reported to regulate several plant processes, such as disease resistance (AFZAL et al. 2008), brassinosteroid signaling (ZHU et al. 2013; FRIEDRICHSEN et al. 2000) and cell growth (HÉMATY; HÖFTE, 2008). Zolla et al. (2008) have identified three proteins related to kinase activity in their proteomic study of Bt and non-Bt near-isogenic line, two of them being down-regulated in the GM samples (adenosine kinase and cytosolic 3-phospoglycerate kinase) and one completely repressed in the GM samples (fructokinase-1).

Two of the main mechanisms of pathogen defense in plants are direct related to innate immune systems (JONES; DANGL, 2006). The first one involves the action of trans membrane receptors, so called

pattern recognition receptors (PRRs), which detect conserved pathogen associated molecular patterns (PAMPs) and initiate plant defense responses. This mechanism is called PAMP-triggered immunity (PTI) (RODRIGUEZ et al. 2010). A well-characterized plant PRR, the *A. thaliana* flg22 receptor FLS2, is a highly conserved leucine-rich repeat receptor kinase (LRR-RK), and studies have shown its relation with MAPK cascades in mechanism of plant defense (CHINCHILLA et al. 2007). The other mechanism involves the plant ability of detecting microbial effector proteins via immune receptors, called resistance (*R*) proteins. *R* proteins are able to trigger forms of localized host cell death, called the hypersensitive response (HR). This mechanism is called effector-triggered immunity (ETI) (RODRIGUEZ et al. 2010). Studies in tobacco and tomato have shown evidence of the involvement of MAPK cascades in ETI and *R* gene signaling (ROMEIS et al. 1999; DEL POZO et al. 2004).

The differential expression analyses of the stacked GM event in comparison with the adequate comparators showed significant changes in the regulation of several metabolic pathways, with the majority of them being down regulated in the stacked samples. Most of the identified transcripts in these pathways were involved in transcription factor activity, taking part in major biological processes, such as mechanisms of abiotic stress, developmental and growth processes. In addition, it is important to emphasize that several transcripts involved in mechanisms of disease resistance were identified as being down-regulated in stacked GM samples, for instance: AP2/EREBP transcription factor, WRKY DNA-binding transcription factor, MAPKKK family protein kinase and Leucine-Rich repeat receptor-like kinase (LRR-RK). In conclusion, our results showed that stacked GM events (expressing both insecticidal CRY and herbicide tolerant EPSPS proteins) are different from their single GM and near-isogenic non-GM varieties, in respect of their transcriptomic profile.

Table 6. Description and related GO terms of the five transcripts with the highest fold-change assigned to biological processes pathways in the Conv *vs* RRxBt enrichment analysis.

Ensembl Plants ID	Description	Related GO Terms	Log2FC
GRMZM2G438202_T01	Putative AP2/EREBP transcription factor superfamily protein (UniProt:K7TSB0)	GO:0003700 transcription factor activity GO:0006355 regulation of transcription, DNA-dependent GO:0003677 DNA binding	-4.85
GRMZM2G319167_T01	Uncharacterized protein (UniProt:A0A096T7C7)	GO:0006355 regulation of transcription, DNA-dependent GO:0005634 nucleus GO:0005488 binding GO:0004879 ligand-dependent nuclear receptor activity GO:0003677 DNA binding	-3.63
GRMZM2G563728_T01	Uncharacterized protein (UniProt:A0A096U2M3)	oGO:0003779 actin binding GO:0015629 actin cytoskeleton GO:0030036 actin cytoskeleton organization	-3.43
GRMZM5G862488_T01	Uncharacterized protein (UniProt:A0A096UDU8)	GO:0019358 nicotinate nucleotide salvage GO:0019363 pyridine nuclt. biosynthetic process GO:0004516 nicotinate phosphoribosyltransf. act.	-3.42
GRMZM2G020150_T01	Putative AP2/EREBP transcription factor superfamily protein (UniProt:A0A096Q2T2)	GO:0003700 transcription factor activity GO:0006355 regulation of transcription, DNA-dependent GO:0003677 DNA binding	-3.36

Table 7. Description and related GO terms of the five transcripts with the highest fold-change assigned to biological processes pathways in the Bt *vs* RRxBt enrichment analysis.

Ensembl Plants ID	Description	Related GO Terms	Log2FC
GRMZM2G082580_T01	Uncharacterized protein (UniProt:A0A096R8X3)	GO:0016020 membrane GO:0016760 cellulose synthase activity GO:0008270 zinc ion binding GO:0005515 protein binding GO:0030244 cellulose biosynthetic process	-3.93
GRMZM2G066428_T01	Uncharacterized protein (UniProt:A0A096QY71)	GO:0005975 carbohydrate metabolic process GO:0016868 intramolecular transferase activity, phosphotransferases	-3.49
GRMZM2G305066_T01	Putative MAPKKK family protein kinase (UniProt:K7VKT3)	GO:0004674 protein serine/threonine kinase act. GO:0006468 protein amino acid phosphorylation GO:0005524 ATP binding GO:0004672 protein kinase activity	-3.23
GRMZM2G032602_T01	Disease resistance gene analog PIC17 (UniProt:A0A096QAR4)	GO:0006915 apoptosis GO:0017111 nucleoside-triphosphatase activity GO:0000166 nucleotide binding GO:0006952 defense response GO:0005515 protein binding GO:0005524 ATP binding	-3.14
AC209050.3_FGT003	WRKY DNA-binding domain superfamily prot. (UniProt:K7UYX9)	GO:0003700 transcription factor activity GO:0043565 sequence-specific DNA binding GO:0045449 regulation of transcription	-3.10

Table 8 Description and related GO terms of the five transcripts with the highest fold-change assigned to biological processes pathways in the RR *vs* RRxBt enrichment analysis.

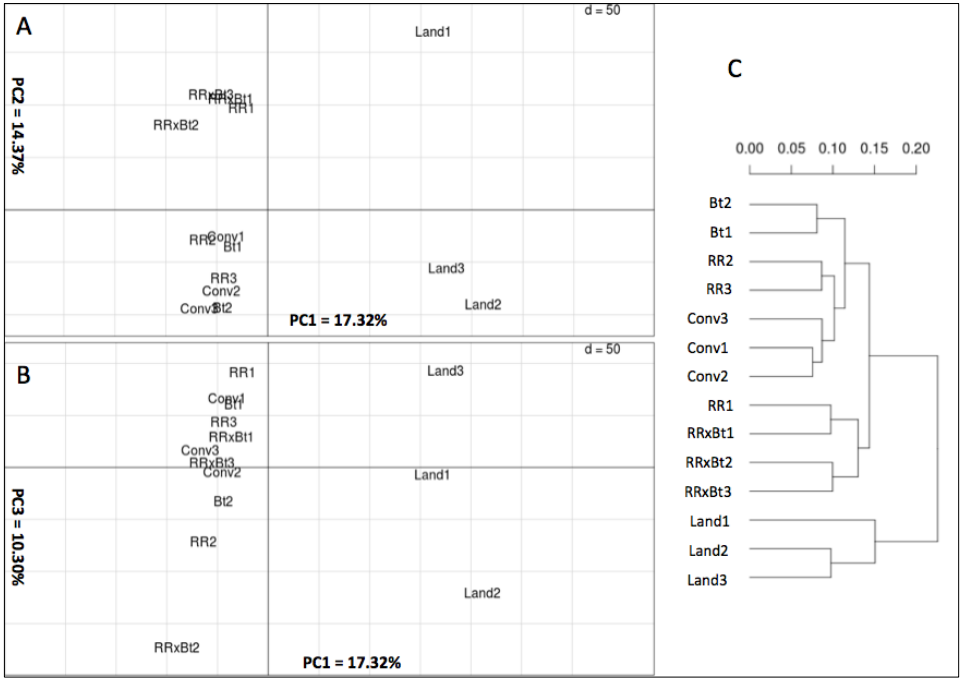
Ensembl Plants ID	Description	Related GO Terms	Log2FC
GRMZM5G862488_T01	Uncharacterized protein (UniProt:A0A096UDU8)	GO:0019358 nicotinate nucleotide salvage GO:0019363 pyridine nucleotide biosynthetic process GO:0004516 nicotinate phosphoribosyltransferase act.	-3.41
GRMZM2G563728_T01	Uncharacterized protein (UniProt:A0A096U2M3)	GO:0003779 actin binding GO:0015629 GO:0030036 actin cytoskeleton/actin	-2.80
GRMZM5G867798_T01	Putative leucine-rich repeat receptor-like protein kinase family prot. (UniProt:K7TR92)	GO:0004674 protein serine/threonine kinase activity GO:0006468 protein amino acid phosphorylation GO:0005524 GO:0005515 ATP binding/ protein binding GO:0004672 protein kinase activity	-2.47
GRMZM2G451443_T01	Uncharacterized protein (UniProt:K7U084)	GO:0008565 protein transporter activity GO:0006468 protein amino acid phosphorylation GO:0006886 intracellular protein transport GO:0004674 protein serine/threonine kinase activity GO:0005524 GO:0005488 binding/ATP binding	-2.44
GRMZM2G331105_T01	Uncharacterized protein (UniProt:K7VCA4)	GO:0005737 cytoplasm GO:0004812 aminoacyl-tRNA ligase activity GO:0000166 nucleotide binding GO:0006418 tRNA aminoacylation for protein translation GO:0006412 translation GO:0005524 GO:0005488 ATP binding/binding	-2.20

3.4. Landraces as a comparator of natural variability

A landrace variety was included in this study in order to consider the extent of gene expression variation related to different maize genetic backgrounds, as well as to possibly disclose differences in GM lines that might fit within the variation observed in non-modified materials (AGAPITO-TENFEN et al. 2014). As the main point of the Landrace samples is to address the natural variation of transcript expression, this section will not cover differences in specific metabolic pathways or any enrichment analysis of gene sets. In addition, the accumulated data obtained in this kind of studies will help to estimate the possible effects of landraces contamination by transgene(s).

Four comparisons were performed in order to address the aforementioned concerning: Conv *vs* Land, Bt *vs* Land, RR *vs* Land and RRxBt *vs* Land. The PCA plots and dendrogram for all the samples is shown in Figure 10. The first plot (PC1 x PC2) shows that Landrace samples are completely separated from the other ones, which accounted for the majority of the variation in the dataset (17.32%). Similarly, the second plot (PC1 x PC3) showed the same separation of Landrace samples from the other ones, explaining the same percentage of the variation. In addition, the dendrogram in Figure 10C shows a clear differentiation of the Landraces samples from the other high-breeding varieties. The differential expression analysis showed a high number of up and down regulated transcripts in the Landrace variety for all comparisons, as well as exclusive ones (Table 9).

Figure 10. Principal Component Analysis and dendrogram of Landrace variety and high-breeding varieties. A) PCA plot of PC1 x PC2. B) PCA plot of PC1 x PC3. C) Dendrogram of Landrace, stacked GM, single GM and conventional varieties. The dendrogram was construct with the differentially expressed transcript, using Jensen-Shanon divergence (JSD) (GROSSE et al. 2002).



Source: Author

Table 9. Number of up and down regulated differentially expressed (DE) transcrips for the comparisons between the Landrace and the high-breeding GM maize varieties.

Comparison	Category	N° of DE transcripts	Total	Exclusive
Conv *vs* Land	Up-regulated	1629	3298	Land = 213
	Down-regulated	1669		Conv = 33
Bt *vs* Land	Up-regulated	1145	2304	Land = 228
	Down-regulated	1159		Bt = 38
RR *vs* Land	Up-regulated	1625	2771	Land = 189
	Down-regulated	1146		RR = 25
RRxBt *vs* Land	Up-regulated	2225	3660	Land = 226
	Down-regulated	1435		RRxBt = 26

Our previous proteomics study using the same varieties showed similar PCA results, where 15.6% of the variation (PC2 x PC3) was explained by the Landrace variety (AGAPITO-TENFEN et al. 2014). In addition, our miRNA analysis with the same samples showed a separate clustering of Landrace samples, which accounted for 30.75% of the variation in the dataset (data not published). A landrace variety was also included in a comparative analysis of potato tuber proteomes of GM potato varieties (LEHESRANTA et al. 2005). Different varieties and landraces showed variation in the proteomic profile, which indicates extensive genotypic variation. Most of the proteins detected exhibited significant quantitative and qualitative differences between one or more GM varieties and landraces.

It is important to emphasize that the use of non-GM varieties that are genetically distant from the GM event under investigation is not a requirement of international guidelines addressing the issue for comparative assessments of the environmental and health risk analysis of GM plants (AHTEG, 2010). International guidelines demand that for the risk assessment of stacked GM events, both GM parental plants, which have been risk assessed previously, should be used as the comparators. In this case, the addition of the near-isogenic non-GM line is not mandatory, but it could be included if it can provide valuable information (EFSA, 2011).

However, the inclusion of a landrace in transcriptome studies will provide additional data of the distinct genetic backgrounds for the crop species. For outcrossing species, such as maize, transcription comparisons will allow in the near future to address the genetic and phenotypic effects of landraces contamination by transgenes. In

addition, landraces are the major components of germplasm banks (FAO, 2010) and major source of alleles that provides adaptation to specific conditions. Moreover, since a variety is a temporarily allelic association, any contamination could cause a major effect of that association, because recombination with very distinct alleles that comes from the high-breeding GM varieties can disrupt those allelic combinations. Furthermore, the long-term natural and artificial processes of selection of many allelic associations to a specific environment that had a contribution of their maintainers can be demolished. Thus, maintainers' rights and the loss of the allelic associations richness also shall be part of the biosafety analysis before the release of any GM variety.

4. CONCLUSIONS

Overall, our results show that several biological processes are differentially regulated in both single GM *vs* near-isogenic non-GM comparisons and stacked GM *vs* adequate comparators. In addition, the present study confirmed the relevance of investigating the adequate comparators for risk assessment analysis based on the use of omics technologies. This conclusion is based on the demonstration of differentially regulated transcripts in the single GM (Bt- and epsps-expressing traits) compared to the near-isogenic non-GM variety, which were involved in several metabolic processes, such as redox pathways (oxidoreductase activity and monooxygenase pathway), protein glycosylation and phosphorylation, transferase activity, among others. Moreover, differentially regulated transcripts were also observed in the comparison of the stacked GM event and its adequate comparators (single GM and near-isogenic non-GM lines), which were assigned to differentially modulated pathways, especially those related to regulation of transcription, such as DNA binding, protein serine/threonine kinase activity, protein kinase activity and ATP binding.

The findings in this study, together with our previous study (AGAPITO-TENFEN et al. 2014), indicate that the genome changes in stacked GM maize may influence the overall gene expression in ways that may have relevance for safety assessments. In addition, these findings also demonstrate that molecular profiling could be applied as a useful tool in ways to increase confidence in risk assessments if the profiles are properly designed to address relevant risks and are applied at the correct stage of the assessment (HEINEMANN et al. 2011). This is the first report on comparative transcriptomics analysis of stacked versus single event transgenic crops. Although the changes in transcript expression may not represent a biosafety issue *per se*, the detection of such changes should be carefully taken in account in risk assessments analysis in order to address the biological meaning of these changes.

Finally, the Landrace Pixurum 5 transcriptome analysis revealed that a variety outside of GM genetic background and with higher genetic variability would be useful to future biosafety studies. In addition to the specific studies of transgene contamination effects, landrace could be used to further address adaptation issues. Moreover, in the absence of the adequate comparator, as research independent groups are facing, a well-characterized landrace would be the single option in their studies.

5. REFERENCES

AD HOC TECHNICAL EXPERT GROUP (AHTEG): United Nations Environment Programme, CBD: Guidance document on risk assessment of living modified organisms, 2010.

AFZAL, A.J.; WOOD, A.J.; LIGHTFOOT, D.A. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact, v.21(5), 507-17, 2008.

AGAPITO-TENFEN, S.Z.; GUERRA, M.P.; WIKMARK, O.G.; NODARI, R.O. Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. Proteome Sci, v.11, 46, 2013.

AGAPITO-TENFEN, S.Z.; VILPERTE, V.; BENEVENUTO, R.F.; ROVER, C.M.; TRAAVIK, T.I.; NODARI, R.O. Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. BMC Plant Biology, v.14, 346, 2014.

AIRD, D.; ROSS, M.G.; CHEN, W.S.; DANIELSSON, M.; FENNELL, T.; RUSS, C. et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol, v.12(2), R18, 2011.

ARMSTRONG, D. J. Cytokinins: Chemistry, Activity and Function, eds. MOK, D.W.S.; MOK, M.C. CRC, Boca Raton, FL. pp. 139–154, 1994.

ASAI, T.; TENA, G.; PLOTNIKOVA, J.; WILLMANN, M.R.; CHIU, W.; GOMEZ-GOMEZ, L. et al. MAP kinase signalling cascade in Arabidopsis innate immunity. Nature, v.415, 977-983, 2002.

BARROS, E.; LEZAR, S.; ANTTONEN, M.J.; VAN DIJK, J.P.; RÖHLIG, R.M.; KOK, E.J. et al. Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnol J, v.8, 436–451, 2010.

BARTSCH, M.; GOBBATO, E.; BEDNAREK, P.; DEBEY, S.; SCHULTZE, J.L.; BAUTOR, J. et al. Salicylic acid-independent ENHANCED DISEASE SUSCEPTIBILITY1 signaling in Arabidopsis immunity and cell death is regulated by the monooxygenase FMO1 and

the Nudix hydrolase NUDT7. Plant Cell, v.18(4), 1038-51, 2006.

BECRAFT, P.W. Receptor kinases in plant development. Trends Plant Sci, v.3, 384–388, 1998.

CANCI, I.J.; BRASSIANI, I.A. Anchieta História, Memória e Experiência: Uma caminhada construída pelo povo. São Miguel do Oeste: Mclee; 2004.

CHINCHILLA, D.; ZIPFEL, C.; ROBATZEK, S.; KEMMERLING, B.; NURNBERGER, T. et al. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defense. Nature, v.448, 497–500, 2007.

CODEX ALIMENTARIUS COMMISSION. Guideline for the conduct of food safety assessment of food derived from recombinant DNA plants. Rome: CAC/GL 45–2003; 2003.

COLL, A.; NADAL, A.; COLLADO, R.; CAPELLADES, G.; KUBISTA, M.; MESSEGUER, J. et al. Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. Plant Mol Biol, v.73, 349–362, 2010.

COLL, A.; NADAL, A.; COLLADO, R.; CAPELLADES, G.; MESSEGUER, J.; MELE, E. et al. Gene expression profiles of MON810 and comparable non-GM maize varieties cultured in the field are more similar than are those of conventional lines. Transgenic Res, v.18, 801–808, 2009.

COLL, A.; NADAL, A.; PALAUDELMA, M.; MESSEGUER, J.; MELE, E.; PUIGDOMENECH, P. et al. Lack of repeatable differential expression patterns between MON810 and comparable commercial varieties of maize. Plant Mol Biol, v.68, 105–117, 2008.

Comissão Técnica Nacional de Biossegurança (CTNBIO). Parecer Técnico n° 1596/2008 - Liberação Comercial de Milho Geneticamente Modificado Tolerante ao Glifosato, Milho Roundup Ready 2, Evento NK603 - Processo n° 01200.002293/2004- 16. [http://www.ctnbio.gov.br/index.php/content/view/12341.html]

Comissão Técnica Nacional de Biossegurança (CTNBIO). Parecer

Técnico n° 2052/2009 - Liberação Comercial de Milho Resistente a Insetos, Milho MON 89034 - Processo n° 01200.003326/2008-61. [http://www.ctnbio.gov.br/index.php/ content/view/14098.html]

Comissão Técnica Nacional de Biossegurança (CTNBIO). Parecer Técnico n° 2725/2010 - Liberação Comercial de Milho Geneticamente Modificado Resistente a Insetos e Tolerante a Herbicidas, Milho MON 89034 x NK 603 - Processo n° 01200.003952/ 2009-38. [http://www.ctnbio.gov.br/index.php/content/view/15704.html]

CONVENTION ON BIOLOGICAL DIVERSITY. Cartagena Protocol on Biosafety to the Convention on Biological Diversity: text and annexes. Montreal: Secretariat of the Convention on Biological Diversity, 30p., 2000.

DAVIES, H. A role for "omics" technologies in food safety assessment. Food Control, v.21, 1601–1610, 2012.

DEL POZO, O.; PEDLEY, K.F.; MARTIN, G.B. MAPKKKalpha is a positive regulator of cell death associated with both plant immunity and disease. EMBO J. v.23, 3072–82, 2004.

DE SCHRIJVER, A.; DEVOS, Y.; VAN DEN BULCKE, M.; CADOT, P.; DE LOOSED, M.; REHEUL, D. et al. Risk assessment of GM stacked events obtained from crosses between GM events. Trends Food Sci Tech, v.18, 101–109, 2007.

DE SMET, I.; VOSS, U.; JÜRGENS, G.; BEECKMAN, T. Receptor-like kinases shape the plant. Nat Cell Biol, v.11(10), 1166-73, 2009.

DONG, Z.; CHEN, Y. Transcriptomics: advances and approaches. Sci China Life Sci, v.56(10), 960-7, 2013.

DU, Z.; ZHOU, X.; LING, Y.; ZHANG, Z.; SU, Z. agriGO: a GO analysis toolkit for the agricultural community. Nucl Acids Res, v.38, 64-70, 2010.

EFSA Panel on Genetically Modified Organisms (GMO). Guidance on selection of comparators for the risk assessment of genetically modified plants and derived food and feed. EFSA Journal, 9(5), 2149, 2011.

EULGEM, T.; RUSHTON, P.J.; ROBATZEK, S.; SOMSSICH, I.E. The WRKY superfamily of plant transcription factors. Trends Plant Sci, v.5(5), 199-206, 2000.

EULGEM, T.; SOMSSICH, I.E. Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol, v.10(4), 366-71, 2007.

European Food Safety Authority (EFSA). Guidance Document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants containing stacked transformation events. EFSA J, v.512, 1–5, 2007.

ESWARAMOORTHY, S.; BONANNO, J.B.; BURLEY, S.K.; SWAMINATHAN, S. Mechanism of action of a flavin- containing monooxygenase. Proc. Natl. Acad. Sci. USA, v.103, 9832–9837, 2006.

FAO. The Second Report on The State Of The World's Plant Genetic Resources For Food And Agriculture. Rome, 2010.

FASS, D. Disulfide bonding in protein biophysics. Annual Review of Biophysics, v.41, 63–79, 2012.

FRIEDRICHSEN, D.M.; JOAZEIRO, C.A.; LI, J.; HUNTER, T.; CHORY, J. Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. Plant Physiol, v.123(4), 1247-56, 2000.

GRAVEL, P.; GOLAZ, O.; WALZER, C., HOCHSTRASSER, D.F.; TURLER, H.; BALANT, L.P. Analysis of glycoproteins separated by two- dimensional gel electrophoresis using lectin blotting revealed by chemiluminescence," Analytical Biochemistry, v.221(1), 66–71, 1994.

GREGERSEN, P.L.; BRINCH-PEDERSEN, H.; HOLM, P.B. A microarray- based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. Transgenic Res, v.14, 887–905, 2005.

GROSSE, I.; BERNAOLA-GALVÁN, P.; CARPENA, P.; ROMÁN-ROLDÁN, R.; OLIVER, J.; STANLEY, H.E. Analysis of symbolic sequences using the Jensen-Shannon divergence. Phys Rev E Stat

Nonlin Soft Matter Phys, v.65(4), 2002.

GUO, Z.J.; CHEN, X.J.; WU, X.L.; LING, J.Q.; XU, P. Overexpression of the AP2/EREBP transcription factor OPBP1 enhances disease resistance and salt tolerance in tobacco. Plant Mol Biol, v.55(4), 607-18, 2004.

HARDISTY, J. F.; BANAS, D. A.; GOPINATH, C.; HALL, W. C.; HARD, G. C.; TAKAHASHI, M. Spontaneous renal tumors in two rats from a thirteen week rodent feeding study with grain from molecular stacked trait lepidopteran and coleopteran resistant (DP-ØØ4114-3) maize. Food Chem Toxicol, v.53, 428-31, 2013.

HEINEMANN, J.A.; KURENBACH, B.; QUIST, D. Molecular profiling - a tool for addressing emerging gaps in the comparative risk assessment of GMOs. Environ Int, 37, 1285–1293, 2011.

HEMATY, K.; HOFTE, H. Novel receptor kinases involved in growth regulation. Curr. Opin. Plant Biol, v.11, 321–328, 2008.

HÖSS, S.; REIFF, N.; OTTERMANNS, R.; PAGEL-WIEDER, S.; DOHRMANN, A.B.; TEBBEE, C.C. et al. Risk assessment of the cultivation of a stacked Bt-maize variety (MON89034 × MON88017) for nematode communities. Soil Biology and Biochemistry, v.91, 109–118, 2015.

HUANG, W.; SHERMAN, B.T.; LEMPICKI, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res, v.37(1), 1-13, 2009.

JAMES, C. Global Status of Commercialized Biotech/GM Crops: 2014. ISAAA Brief No. 49. Ithaca: ISAAA Press; 2014.

JONES, J.D.; DANGL, J.L. The plant immune system. Nature, v.444, 323–29, 2006.

KARIN, M. Signal Transduction from the Cell Surface to the Nucleus through the Phosphorylation of Transcription Factors, Curr Opin Cell Biol, v.6, 415–424, 1994.

KAWAKATSU, T.; KAWAHARA, Y.; ITOH, T.; TAKAIWA, F. A

whole-genome analysis of a transgenic rice seed-based edible vaccine against cedar pollen allergy. DNA Res, 20(6), 623-31, 2013.

KERSEY, P. J.; ALLEN, J. E.; CHRISTENSEN, M. et al. Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic acids research, v.42, 546-552, 2014.

KHOKHLATCHEV, A.V.; CANAGARAJAH, B.; WILSBACHER, J.; ROBINSON, M.; ATKINSON, M. et al. Phosphorylation of the MAP kinase ERK2 promotes its homodimerization and nuclear translocation. Cell, v.93, 605–15, 1998.

KIM, D.; PERTEA, G.; TRAPNELL, C.; PIMENTEL, H.; KELLEY, R.; SALZBERG, S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology, v.14, 36, 2013.

KIZIS, D.; LUMBRERAS, V.; PAGÈS, M. Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. FEBS Lett, 498(2-3), 187-9, 2001.

KRYSAN, P.J.; JESTER, P.J.; GOTTWALD, J.R.; SUSSMAN, M.R. An Arabidopsis mitogen-activated protein kinase kinase kinase gene family encodes essential positive regulators of cytokinesis. Plant Cell, v.14(5), 1109-20, 2002,

LAMBIRTH, K.C.; WHALEY, A.M.; BLAKLEY, I.C.; SCHLUETER, J.A.; BOST, K.L.; LORAINE, A.E. et al. A Comparison of transgenic and wild type soybean seeds: analysis of transcriptome profiles using RNA-Seq. BMC Biotechnology, v.15, 89, 2015.

LEHESRANTA, S.J.; DAVIES, H.V.; SHEPHERD, L.V.T.; NUNAN, N.; MCNICOL, J.W.; AURIOLA, S. et al. Comparison of Tuber Proteomes of Potato Varieties, Landraces, and Genetically Modified Lines. Plant Physiol, v.138, 1690–1699, 2005.

LOCKHART, D.J.; WINZELER, E.A. Genomics, gene expression and DNA arrays. Nature, v.405, 827–836, 2000.

MEYERS, B.C.; GALBRAITH, D.W.; NELSON, T.; AGRAWAL, V. Methods for transcriptional profiling in plants. Be fruitful and replicate.

Plant Physiol, v.135, 637–652, 2004.

MONTERO, M.; COLL, A.; NADAL, A.; MESSEGUER, J.; PLA, M. Only half the transcriptomic differences between resistant genetically modified and conventional rice are associated with the transgene. Plant Biotechnol J, v.9(6), 693-702, 2011.

MUNDY, J.; CHUA, N.H. Abscisic acid and water-stress induce the expression of a novel rice gene. EMBO J, 7(8), 2279-86, 1988.

OKAMURO, J.K.; CASTER, B.; VILLARROEL, R.; VAN MONTAGU, M.; JOFUKU, K.D. The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. PNAS, v.94, 7076– 7081, 1997.

ONDA, Y. Oxidative Protein-Folding Systems in Plant Cells. International Journal of Cell Biology, v.2013, 1-15, 2013.

QUACKENBUSH, J. Microarray data normalization and transformation. Nat Genet, v.32, 496-501, 2002.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

RODRIGUEZ, M.C.; PETERSEN, M.; MUNDY, J. Mitogen-activated protein kinase signaling in plants. Annu Rev Plant Biol, v.61, 621-49, 2010.

ROMEIS, T.; PIEDRAS, P.; ZHANG, S.; KLESSIG, D.F.; HIRT, H.; JONES, J.D. Rapid Avr9- and Cf-9-dependent activation of MAP kinases in tobacco cell cultures and leaves: convergence of resistance gene, elicitor, wound, and salicylate responses. Plant Cell, v.11, 273–87, 1999.

ROTH, Z.; YEHEZKEL, G.; KHALAILA, I. Identification and Quantification of Protein Glycosylation. International Journal of Carbohydrate Chemistry, v.2012, 1-10, 2012.

SCHUPPENER, M.; MÜHLHAUSE, J.; MÜLLER, A. K.; RAUSCHEN, S. Environmental risk assessment for the small

tortoiseshell Aglais urticae and a stacked Bt-maize with combined resistances against Lepidoptera and Chrysomelidae in central European agrarian landscapes. Mol Ecol, v.21, 4646-4662, 2012.

SHARONI, A.M.; NURUZZAMAN, M.; SATOH, K.; SHIMIZU, T.; KONDOH, H.; SASAYA, T. et al. Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. Plant Cell Physiol, v.52(2), 344-60, 2011.

SPÖK, A.; ECKERSTORFER, M.; HEISSENBERGER, A.; GAUGITSCH, H. Risk Assessment of "stacked events". 2007. Available: http://www.bmg.gv.at/cms/home/attachments/7/7/0/CH1052/CMS1180523433481/cms1200659635294_internetversion_2_07stacked_events.pdf

SUBOTA, I.Y.; ARZIEV, A.S.; SENZHENKO, L.P.; TARASENKO, V.I.; KONSTANTINOV, Y.M. Inhibitor analysis of protein phosphorylation/dephosphorylation in maize mitochondria in relation to redox conditions. Russian Journal of Plant Physiology, V.54(3), 343-349, 2007.

SUPEK, F.; BOŠNJAK, M.; ŠKUNCA, N.; ŠMUC T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. PLoS One, v.6(7), e21800, 2011.

TANG, M.; LIU, X.; DENG, H.; SHEN, S. Over-expression of JcDREB, a putative AP2/EREBP domain-containing transcription factor gene in woody biodiesel plant Jatropha curcas, enhances salt and freezing tolerance in transgenic Arabidopsis thaliana. Plant Sci, v.181(6), 623-31, 2011.

TAVERNIERS, I.; PAPAZOVA, N.; BERTHEAU, Y.; DE LOOSE, M.; HOLST-JENSEN, A. Gene stacking in transgenic plants: towards compliance between definitions, terminology, and detection within the EU regulatory framework. Environ Biosafety Res, v.7, 197-218, 2008.

TRAPNELL, C.; ROBERTS, A.; GOFF, L.; PERTEA, G.; KIM, D.; KELLEY, D.R. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc, v.7(3), 562-78, 2012.

TRIPATHI, P.; RABARA, R.C.; RUSHTON, P.J. A systems biology perspective on the role of WRKY transcription factors in drought responses in plants. Planta, v.239(2), 255-66, 2014.

VALDÉS, A.; IBÁÑEZ, C.; SIMÓ, C.; GARCÍA-CAÑAS, V. Recent transcriptomics advances and emerging applications in food science. Trends in Analytical Chemistry, v.52, 142–154, 2013.

VAN DER FITS, L.; MEMELINK, J. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondarymetabolism. Science, v.289, 295–297, 2000.

VEACH, Y.K.; MARTIN, R.C.; MOK, D.W.; MALBECK, J.; VANKOVA, R.; MOK, M.C. O-glucosylation of cis-zeatin in maize. Characterization of genes, enzymes, and endogenous cytokinins. Plant Physiol, v.131(3), 1374-80, 2003.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet, v.10, 57–63, 2009.

WHITMARSH, A.J.; DAVIS, R.J. Regulation of Tran- scription Factor Function by Phosphorylation, Cell Mol Life Sci, v.57, 1172–1183, 2000.

WRZACZEK, M.; HIRT, H. Plant MAP kinase pathways: how many and what for? Biol Cell, v.93(1-2), 81-7, 2001.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell, 6(2), 251-64, 1994.

ZHAO, Y.; CHRISTENSEN, S.K.; FANKHAUSER, C.; CASHMAN, J.R.; COHEN, J.D.; WEIGEL, D. et al. A role for flavin monooxygenase-like enzymes in auxin biosynthesis. Science, v.291, 306–309, 2001.

ZHENG, W.; MA, L.; ZHAO, J.; LI, Z.; SUN, F.; LU, X. Comparative transcriptome analysis of two rice varieties in response to rice stripe virus and small brown planthoppers during early interaction. PLoS One, v.8(12), e82126, 2013.

ZHOU, J.; TANG, X.; MARTIN, G.B. The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis- element of pathogenesis-related genes. EMBO J, v.16, 3207–3218, 1997.

ZHOU, Y.; XIA, H.; LI, X.J.; HU, R.; CHEN, Y.; LI, X.B. Overexpression of a cotton gene that encodes a putative transcription factor of AP2/EREBP family in Arabidopsis affects growth and development of transgenic plants. PLoS One, 8(10), e78635, 2013.

ZHU, J.Y.; SAE-SEAW, J.; WANG, Z.Y. Brassinosteroid signalling. Development, v.140(8), 1615-20, 2013.

ZOLLA, L.; RINALDUCCI, S.; ANTONIOLI, P.; RIGHETTI, P.G. Proteomics as a complementary tool for identifying unintended side effects occuring in transgenic maize seeds as a result of genetic modifications. J Proteome Res, v.7, 1850–1861, 2008.

6. CAPÍTULO II

MIRNOME PROFILING OF BT- AND EPSPS-EXPRESSING TRANSGENIC MAIZE

ABSTRACT

Background: Recent developments in molecular biology include the emerging technologies of omics profiling – transcriptomics, proteomics, metabolomics and, recently, mirnomics - which have been used as non-targeted approaches to detect possible unintended effects arising from the development of new types of GMOs. Synergic and antagonistic effects of the introduction of transgenes in the overall gene expression on the GM crop are still under debate of its requirement in risk assessment analysis. In addition, the study of small RNAs and their role in regulating gene expression are not even mentioned. Here we report the first results of a miRNA profiling comparative analysis of a stacked commercial maize hybrid containing insecticidal (CRY1A.105 and CRY2AB2 proteins), and herbicide tolerant (EPSPS protein) traits, in comparison to the single GM and near-isogenic hybrids with the same genetic background.

Results: Our results demonstrate that 13 endogenous conserved miRNAs were differentially regulated in some of the pair wise comparisons analyzed. miRNAs differentially regulated in the Bt-expressing event, in comparison with the near-isogenic non-GM line, showed to target endogenous transcriptions factors involved in several biological processes, such as leaf development, shoot maturation, flowering, mechanism of stress resistance, hormone signaling and RNA folding, processing and translation. In addition, the miRNAs differentially regulated in the stacked GM event (CRY and EPSPS proteins), in comparison to the single GM (EPSPS protein) and near-isogenic non-GM line, showed to target transcription factors mainly involved in ABA signal transduction pathways. Moreover, novel miRNAs were also detected in our miRNA dataset, which might represent a biosafety concern, since they could interact with the endogenous conserved miRNAs and regulate several metabolic processes in a completely different manner.

Conclusions: The results of this study indicated that specific miRNA

expressing varies between stacked GM, single GM and near-isogenic non-GM lines, which might impact the expression of endogenous genes and hence several major metabolic pathways. In addition, the use of RNA-Seq as a molecular profiling technique showed to be useful to detect such changes and, therefore, could be taken as a useful tool in risk assessment of GM crops. In face of our findings, and due to the well-known relationship of miRNAs with gene expression regulation, these studies should not be neglected in risk assessment analysis, since they could provide valuable information on possible unintended effects of genetic modifications.

Key words: miRNome, RNA-Seq, genetically modified organisms, stacked GMO, risk assessment, molecular profiling.

1. BACKGROUND

The increasing market of "stacked" or "pyramided" events, defined as GM events containing two or more traits combined by traditional breeding (TAVERNIERS et al. 2008), has been a trend in the last decade, with 51 million of the world hectarage being grown with these events in 2014 (JAMES, 2014). Regulatory practice within the European Union (EU) consider stacked events as new GM organisms, and additional information on the stability of transgene insertions, expression levels and potential antagonistic or synergistic interactions should be provided prior to marketing (DE SCHRIJVER et al. 2007; EFSA, 2007; AHTEG, 2010). The expression level of transgenes or endogenous genes in a stacked GM plant may not be identical to the parental single GM line due to trans-regulation. These changes are likely to occur if the parental single GM lines share homology in the transgene or regulatory sequences (e.g. promoter sequences) (CBD, 2011).

Recent developments in plant biotechnology include the emerging technologies of omics profiling – transcriptomics, proteomics and metabolomics, which have been used as non-targeted approaches to detect unintended effects and to fill the biosafety gap emerging from the development of new types of GMOs (HEINEMANN et al. 2011). Data on omics profiling of single GM events are available (AGAPITO-TENFEN et al. 2013; BARROS et al. 2010; COLL et al. 2010; ZOLLA et al. 2008), whilst data on stacked GM events are almost absent (AGAPITO-TENFEN et al. 2014). Moreover, when assessing the environmental and health risks of both single and stacked GM plants, adequate comparators should be used in order to isolate possible effects arising from the transgene from those of natural variations sources (AHTEG, 2010).

Routine genetic stability analyses performed to assess the safety of GM plants are often associated to protein quantification (TRTIKOVA et al. 2015; NGUYEN; JEHLE, 2007) and RNA expression (ZHAO et al. 2011; LA PAZ et al. 2010). In addition, studies have also detected transgene rearrangements, such as sequence deletions (HOLCK et al. 2002; HERNÁNDEZ et al. 2003; ROSATI et al. 2008) and nucleotide addition of undesired fragments into the transgene sequence (WINDELS et al. 2001). Nonetheless, recent attention has been drawn to small RNAs and their role in regulating gene expression, especially to the expression profile of endogenous micro RNAs (miRNAs). miRNAs constitute a major family of gene expression regulatory elements, with lengths ranging from 18 to 26 nt, that play crucial roles in post-

transcriptional gene silencing by either mRNA cleavage or translational inhibition (RAMESH et al. 2014). In animals, the majority of miRNA are processed from long hairpin transcripts through successive actions of enzymes from the RNA III family, DROSHA and DICER, while in plants only the DICER enzyme is responsible for miRNA processing (CARTHEW; SONTHEIMER, 2009). Most plants have more than 100 miRNA genes (called MIR genes) (NOZAWA et al. 2012), which are located almost exclusively in intergenic regions of the genome (REINHART et al. 2002).

miRNA are known to control a wide range biological processes in eukaryotic organism, such as embryogenesis (NODINE; BARTEL, 2010), development and cellular behavior (KROL et al. 2010), and response to biotic (NAVARRO et al. 2006) and abiotic stress (SUNKAR; ZHU, 2004). In addition, transcription factors (TF) are among the most common plant miRNAs regulatory targets. These TF are involved in various biological processes, such as leaf development, shoot maturation, phase change and flowering (SHIKATA et al. 2009; WU et al. 2009), drought stress through ABA signaling (LI et al. 2008), gene expression responses to auxin (MALLORY et al. 2005), and also folding, processing and translation of the RNA (FUJII; SMALL, 2011).

To further investigate potential alteration in miRNA regulatory network in GM plants, we used high-throughput technology to survey the differences in the composition and expression profiles of miRNAs among singles and stacked GM events and control plants under highly controlled conditions and to provide insight into the formulation of specific guidelines for the risk assessment of stacked events. We hypothesized that the combination of two transgenes could differentially modulate endogenous miRNA expression, which might have an effect on the regulation of gene expression in the plant metabolism and physiology. To test this hypotheses, we have used a GM stacked maize genotype containing *cry1A.105/cry2Ab2* and *epsps* cassettes, expressing both insect resistance and herbicide tolerance as unlinked traits, as well as genotypes of each single transgene alone, being all maize hybrids with the same genetic background. The seed set of stacked and single GM maize events, as well as the conventional near-isogenic counterpart, developed in the same genetic background, and a landrace variety, enables the isolation of potential effects derived from stacking two transgenes.

2. METHODOLOGY

2.1. Plant material and growth chamber conditions

Five maize varieties were used in this study (Table 1). Two of them are non-GM maize seeds, the hybrid AG8025 (named here as 'conventional') from Sementes Agroceres and the open pollinated variety Pixurum 5 (named here as 'landrace'). These are named in this study as 'Conv' and 'Land', respectively. Pixurum 5 has been developed and maintained by small farmers in South Brazil for several years (CANCI; BRASSIANI, 2004). The other three varieties are GM and have the same genetic background as the conventional variety since they are produced from the same endogamic parental lines. These are: AG8025RR2 (unique identifier MON-ØØ6Ø3-6 from Monsanto Company, glyphosate herbicide tolerance, Sementes Agroceres); AG8025PRO (unique identifier MON-89Ø34-3 from Monsanto Company, resistance to lepidopteran species, Sementes Agroceres) and AG8025PRO2 (unique identifier MON-89Ø34-3 × MON-ØØ6Ø3-6 from Monsanto Company, stacked event resistant to lepidopteran species and glyphosate-based herbicides, Sementes Agroceres). These are named in this study as RR, Bt and RRxBt, respectively. The AG8025 variety is the hybrid progeny of the single cross between maternal endogamous lines "A" with the paternal endogamous line "B". Thus, the used hybrid variety seeds have high genetic similarity (all seeds should be AB genotype in absence of self pollination). All these five commercial varieties were produced by the aforementioned company/farmers and are commonly found in the market in Brazil.

The cultivation of MON-ØØ6Ø3-6, MON-89Ø34-3, and MON-89Ø34-3 × MON-ØØ6Ø3-6 has been approved in Brazil in 2008, 2009 and 2010 respectively (CTNBIO, 2008; CTNBIO, 2009; CTNBIO, 2010). The stacked hybrid MON-89Ø34-3 × MON-ØØ6Ø3-6 expresses two insecticidal proteins (Cry1A.105 and Cry2Ab2 proteins derived from *Bacillus thuringiensis*, under the regulation of P-35S and FMV promoters, respectively), which are active against certain lepidopteran insect species, and two identical EPSPS proteins (under the regulation of P-Ract1 and P-35S promoters) providing tolerance to the herbicide glyphosate. The novel traits of each parent line have been combined through traditional plant breeding to produce this new hybrid. The experimental approach currently applied for the comparative assessment requires the use of conventional counterpart and the single-event counterparts, all with genetic background as close as possible to the GM

plant, as control (EFSA, 2007; CBD, 2013; CODEX ALIMENTARIUS, 2003).

After the confirmation by PCR of the transgenic events in both single and stacked GM seeds and the absence in the 'Conv' and 'Land' ones (data not shown), the seeds from all the five varieties were grown side by side in growth chambers (EletrolabTM model 202/3) set to 16 h light period and 25 °C (± 2 °C). Seedlings were germinated and grown in Plantmax HT substrate (Buschle & Lepper S.A.) and watered daily. No pesticide or fertilizer was applied. Around 50 plants were grown in climate chambers out of which 30 plants were randomly sampled per maize variety (genotype). The collected samples were separated in three groups of ten plants. The ten plants of each group were pooled and were considered one biological replicate. Maize leaves were collected at V4 stage (20 days after seedling). Leaf pieces were cut out, weighed and placed in 3.8 ml cryogenic tubes before immersion in liquid nitrogen. The samples were kept at −80 °C until miRNA isolation.

Table 1. Transgenic and non-transgenic commercial maize varieties used in this study.

Commercial name	GM event	Transgenes	Sample size	Labeled in this study
AG8025RR2	MON-ØØ6Ø3-6	epsps/epsps	30	RR
AG8025PRO	MON-89Ø34-3	cry1A.105/cry2Ab2	30	Bt
AG8025PRO2	MON-89Ø34-3 x MON-ØØ6Ø3-6	cry1A.105/cry2Ab2 x epsps/epsps	30	RRxBt
AG8025	None	None	30	Conv
Pixurum 5	None	None	30	Land

Note: The unique identifier used is from the Biosafety Clearing House (BCH)

2.2. miRNA isolation

Total RNA was isolated from approximately 100 mg of frozen leaf tissue using the miRNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. In brief, samples were homogenized with guanidine-isothiocyanate lyses buffer and further purified using silica-membrane. During purification, in-column DNA digestion was performed using RNAse-free DNAse I supplied by Qiagen to eliminate any remaining DNA. The isolated miRNA was quantified using NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, USA) and resolved in MOPS 1% denaturing gel.

2.3. miRNA deep sequencing

RNA samples (1μg) were sent to FASTERIS SA (Geneva, Switzerland) for library construction (15 cDNA libraries) and sequencing. The libraries sequencing was conducted using the TruSeq SBS Kit v3-HS (Illumina®) in an Illumina HiSeq 2500, with number of cycles of 1x50+7 (*single-end*) in one lane of the HiSeq Flow Cell v3 (Illumina®). Basecalling was performed using the pipelines HiSeq Control Software 2.2.38, RTA 1.18.61.0, CASAVA-1.8.2.

2.4. Library analysis of small RNAs

All low quality reads with FASTq values below 13 were removed, and 5' and 3' adapter, as well as index sequences, were trimmed using the Genome Analyzer Pipeline (Illumina) at Fasteris SA. In addition, quality control was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and no bases with Phred quality score below 30 were found; thus no additional trimming was necessary. Reads outside the 18 – 26 nt range were excluded from the analysis. sRNAs belonging to rRNAs, tRNAs, snRNAs and snoRNAs, as well as chloroplastidial and mitochondrial sequences derived from *Zea mays* and deposited in the Ensembl (KERSEY et al. 2014) and NCBI GenBank databases were identified through mapping using the Bowtie2 v.2.2.4 software (LANGMEAD; SALZBERG, 2012).

2.5. Identification of conserved and novel miRNAs

In order to determine conserved maize miRNA, the sRNA sequences were mapped to the *Zea mays* genome (B73, RefGen_v3, release 25) deposited in the Ensembl Plants database (KERSEY et al. 2014) using the Bowtie2 v.2.2.4 software (LANGMEAD; SALZBERG, 2012), with a maximum of two mismatches, where gaps count as mismatches.

The prediction of novel miRNA was performed using the miR-PREFeR pipeline (Lei; Sun, 2014) with the following parameters: 1) Maximum length of 250 nt for a miRNA precursor; 2) Reads depth cutoff of 200; 3) Maximum gap length of 50 nt between two contigs to form a candidate region; 4) Minimum and maximum length of the mature sequence of 21 and 24 nt, respectively; 5) No requirement of the star sequence to be expressed; and 6) Allow the mature star duplex to have only 2nt 3' overhangs. All annotated mRNA and miRNA were excluded from the analysis by supplying a GFF3 file which list all existing annotations on genomic sequences of the *Z. mays* genome (B73 RefGen_v3, release 25). In addition, the web-based tool MiPred (JIANG et al. 2007) was used to classify the predicted novel miRNA as real, pseudo or not a miRNA precursor. To be considered a real precursor, the miRNA sequence has to have a Minimum Fold Energy (MFE) < -20 kcal/mol and a *p*-value < 0.05. As for the pseudo precursor, at least one of the conditions has to be true.

2.6. Differential expression of miRNA

Read counts were retrieved using SeqMonk v0.29.0 (http://www.bioinformatics.babraham.ac.uk), where probes were design around known miRNA sequences in the reference genome. Only the reads that exactly overlapped the probes were considered for the read counting. The statistical analyses were performed through R language and environment (R Core Team, 2015), using the DESeq2 R package v1.8.1 (LOVE et al. 2014) to normalize the read counts and perform the differential expression analysis.

2.7. Validation of miRNAs by RT-qPCR

In order to validate the differentially expressed miRNAs, a series of RT-qPCR were performed in miRNA isolated from the same samples used for the sRNA libraries. We designed primers to amplify all

differentially expressed miRNA found, these are zma-MIR162, zma-MIR167c, zma-MIR167e, zma-MIR167j, zma-MIR169f, zma-MIR169m, zma-MIR399a, zma-MIR399h, zma-MIR399e, zma-MIR399i, zma-MIR399j, zma-MIR529 and zma-MIR827.

The choice of the endogenous reference genes and the selection of the two best genes were based on our previous work (AGAPITO-TENFEN et al. 2014). The two most suitable endogenous reference genes out of four candidates (ubiquitin carrier protein, folylpolyglutamate synthase, leunig and cullin) were selected as internal standards. The folylpolyglutamate synthase and leunig genes were used to normalize conserved miRNA data due to their best stability value (SV for best combination of two genes 0.062, data not shown). The primers used for the endogenous genes are the same ones used in our previous work, but only the forward primer. The amplification efficiency was obtained from relative standard curves provided for each primer and calculated according to Pfaffl equations (Pfaffl, 2001).

Poly(A) Tailing and cDNA synthesis was performed using the NCode™ VILO™ miRNA cDNA Synthesis Kit (Invitrogen, California, USA) according the manufactures recommendations. For each of the biological replicates, three independent cDNA syntheses were performed in order to assure a good quantity of cDNA for the reactions. Subsequently, the triplicates were mixed and quantified by NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, USA). The RT-qPCRs were performed using the EXPRESS SYBR® GreenER™ miRNA qRT-PCR kit (Invitrogen) according the manufactures recommendations. Briefly, reactions were carried out in triplicates in a volume of 20µL containing 10µL of EXPRESS SYBR® GreenER™ qPCR SuperMix Universal, 10µM miRNA-specific forward primer (or endogenous genes primers), 10µM Universal qPCR Primer and 50ng of cDNA. RT-qPCRs and cDNA synthesis were performed using a StepOne™ Real-Time PCR System (Applied Biosystems, Singapore, Singapore). The normalized relative quantity (NRQ) was calculated to the Pfaffl equations (Pfaffl, 2001).

Real-time relative quantification data were plotted and manually analyzed using Microsoft Excel (Microsoft, Redmond, WA). Normalized gene expression data was obtained using the Pfaffl method for efficiency correction (Pfaffl, 2001). Cq average from each technical replicate was calculated for each biological replicate and used to make a statistical comparison of the genotypes/treatment based on the standard deviation. Due to non-normal distribution, the fold change data were log10 transformed. The fold change means obtained for single versus

stacked GM event were compared using T-test at P <0.05 (R program software) (R Core Team, 2015). Information on real-time data for this study has followed guidelines from the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (BUSTIN et al, 2009).

2.8. Prediction of miRNA targets and pathway enrichment analysis

The prediction of gene targets for the differentially expressed miRNAs was performed using the psRNAtarget online tool (DAI; ZHAO, 2011). This tool uses a 0 – 5 scale to indicate the complementarity between the miRNA and its target, where the lowest values represent a higher complementarity between sequences. The following parameters were used: 1) Maximum expectation: 3.0; 2) Length for complementarity scoring (hspsize): 20; 3) Target accessibility - allowed maximum energy to unpair the target site (UPE): 25; 4) Flanking length around target site for target accessibility analysis: 17bp upstream and 13bp downstream; and 5) Range of central mismatch leading to translational inhibition: 9 – 11 nt.

The miRNA target genes previously predicted were submitted to Single Enrichment Analysis (SEA) using the online tool agriGO v1.2 (DU et al. 2010), with the following parameters: 1) Selected species: *Zea mays* ssp V5a; 2) Statistical test method: Hypergeometric; 3) Multi-test adjustment method: Hochberg (FDR); 4) Significance level of 0.05; 5) Minimum number of 5 mapping entries; and 6) Gene ontology type: Plant GO Slim. Following, the online tool REVIGO (SUPEK et al. 2011) was used to remove the redundant Gene Ontology (GO) terms. Only significant GO terms (False Discovery Rate (FDR) values < 0.05) were used, with the following parameters: 1) Allowed similarity: medium (0.7); 2) Database with GO term sizes: *Zea mays*; e 3) Semantic similarity measure: SimRel.

3. RESULTS AND DISCUSSION

3.1. RNA Library Sequencing and alignment with reference genome

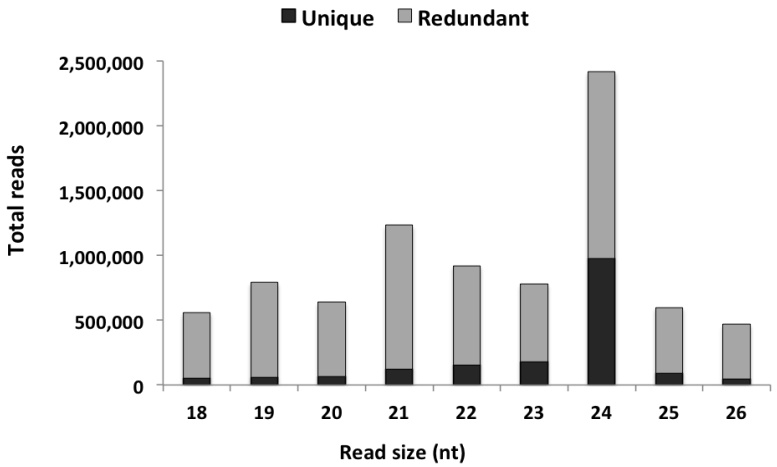
To identify conserved maize miRNAs, sRNA libraries were constructed from leaves and sequenced using Illumina high-throughput technology. A flowchart of the miRNA analysis pipeline is available in the Additional file 1. After removing low quality sequences, those without inserts, or those with adapter contaminants or lengths outside of the 18–26 nt range, a total mean of 8,406,360 reads were obtained (Table 2). The number of reads with different lengths in the redundant and non-redundant sRNA datasets is shown in Figure 1. One of the features used to distinguish miRNAs from other endogenous sRNAs is the size profile, which, for miRNAs, the mature miRNAs have often 21 to 25 nt (GUZMÁN et al. 2013). The majority of the sRNAs in our libraries showed a size of 24 nt, accounting for 28,7% of the total reads, followed by 21 nt (14,6%) and 22 nt (10,9%). This distribution pattern is in accordance with previous studies in other plant species (GUZMÁN et al. 2013; GUZMÁN et al. 2012; ZHAO et al. 2010; FAHLGREN et al. 2007). In maize, the most abundant mature miRNA size is 21 nt (LIU et al. 2014). The distribution of sRNAs in a given species and, therefore, the miRNA size, can be a result from tissue and physiological conditions (ZHU et al. 2008).

Table 2. Number of sequenced reads in the 15 sRNAs libraries for three different classes of size.

Type	Number of reads**	Percentage (%)
Total reads*	16,404,845	100
< 18	4,443,872	27
18 - 26	8,406,360	51
> 26	3,554,613	22

Note: '*'= Reads with Phred score above 30 and lengths of 1 to 44 nt; '**' = Values consist of an average of the 15 sRNA libraries.

Figure 1. Deep sequencing statistics of the sRNA libraries. Number of unique and redundant reads for each size from the 18 - 26 nt range.



Source: Author

On average, approximately 22% of the reads matched other types of non-coding sRNAs, such as rRNAs, tRNAs, snRNAs or snoRNAs and 19.44% matched organellar RNA (chloroplastidial and mitochondrial) (Table 3). The majority of reads matched rRNAs, which are the most abundant transcripts in total RNA, comprising the majority of the molecules in a sample (O'NEIL et al. 2013). Although miRNAs were extracted using a miRNA specific kit, this does not prevent rRNA to be present in the extracted samples. In addition, the library preparation protocol used for this study did not include any rRNA depletion in order to remove this class of RNA prior the sequencing; thus, the highly amount of rRNA in the samples was expected and did not interfere in the final results, once they were identified and excluded from the analysis. Moreover, the highly amount of reads matching cpRNA sequences was also expected, since we are using maize leaves in this study.

Furthermore, we have built a reference sequence with the cassette sequence of the MON89034 and NK603 events, which was used for mapping. Transgenic sequences were detected in Bt, RR and RRxBt varieties, where 0.04% of the reads matched transgenic RNAs, with the RRxBt showing a higher number of mapped reads. This result is expected since the RRxBt variety is a stacked event, containing both Bt and RR cassettes. These reads were mainly mapped to coding regions

(i.e. transgene sequences) of the transgenic reference sequence and no miRNA-like structures were identified using the miR-PREFeR pipeline. Therefore, these reads are most likely to be fragments of transgenic mRNA. However, these miRNA reads being of biosafety concern, and further studies should be performed in order to ensure their origin. As also expected, we verified that no transgenic RNA was detected in the Conventional and Landrace samples. Thus, this was proof that the non-GM samples were not contaminated during any experimental procedure.

Table 3. Number of reads aligned to each sequence type for each of the five varieties used in the study.

Sequence type	RR		Bt		RRxBt		Conv		Land	
	Total reads	%	Total reads	%	Total reads	%	Total reads	%	Total reads	%
snoRNA	1,365	0.02	797	0.01	1,454	0.02	1,247	0.02	1,432	0.02
snRNA	7,922	0.08	2,989	0.04	3,514	0.04	7,937	0.10	10,230	0.12
tRNA	244,114	2.10	83,806	1.01	152,588	1.80	178,399	2.18	226,603	2.68
rRNA	1,710,288	19.70	2,003,376	23.93	1,568,675	18.33	1,528,139	19.06	1,534,186	18.59
cpRNA	1724,918	18.75	1,100,942	13.32	1,531,116	17.99	1,526,939	18.95	1,969,977	23.69
mtRNA	53,568	0.69	48,856	0.59	69,729	0.81	64,551	0.81	52,463	0.64
Transgenic RNA	476	0.04	4,173	0.05	4,908	0.06	-	-	-	-
miRNA	410,396	5.12	462,865	5.58	445,291	5.22	384,181	4.80	441,261	5.31
Other sRNAs	4,708,056	53.52	4,606,799	55.46	4,755,858	55.73	4,346,399	54.08	4,049,017	48.95

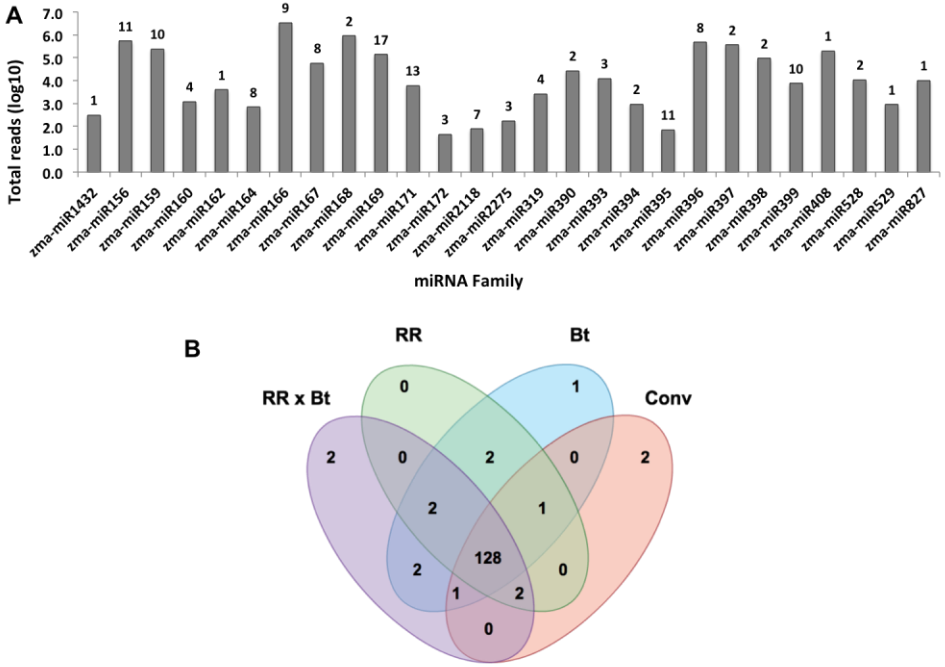
There are 172 miRNA from maize deposited in miRbase; however, only 156 miRNA are annotated for B73 RefGen_v3 genome, which was used as the reference in this study. On average, 428,799 reads matched 146 known maize miRNAs. From the 29 miRNA families previously annotated in maize, 27 miRNA families were identified in our study, with an average of approximately 5 miRNA members per family. The largest family was zma-miR169 with 17 members, followed by zma-miR171 (13 members) and zma-miR156 and zma-miR395 (11 members). Of the remaining miRNA families, 18 contained 2 to 10 members and 5 families were represented by a single member (Figure 2A). With respect to the abundance of each miRNA family, the frequencies varied from 45 reads (zma-miR172) to 3,341,816 reads (zma-miR166) when analyzing the sum of the 15 sRNA libraries, which indicates a high variance among different miRNA families. Variation from members of the same families was also observed. For instance, the abundance of zma-miR169 varied from 103 to 23,651 reads, with the same occurring to some other miRNA families, such as zma-miR167 (243 to 33,034 reads), zma-miR159 (315 to 50,010 reads) and 156 (337 to 95,108 reads). These results show that different members have variable expression levels within one miRNA family. In addition, the high abundance may be a reflection of the role of these miRNA families in different biological process (GUZMÁN et al. 2013).

Among the identified miRNAs, 128 were found to be shared by Conv, Bt, RR and RRxBt libraries, accounting for ~ 82% of the annotated miRNA in the maize reference genome (Figure 2B). A small number of miRNA were identified exclusively in one variety: two in the Conv, two in the RRxBt and one in the Bt samples. However, these exclusive miRNA were expressed in very low abundance, making it difficult to conclude whether they are truly exclusive or not. We have not included the Landrace variety in the Veen diagram due to data visualization. In addition, the miRNA expression in the Landrace variety showed to be highly different from the other varieties. Therefore, these differences will be discussed further in this manuscript.

Although mostly of the identified miRNAs mapped to the correspondent mature sequence in the pre-miRNA, some of them showed a higher abundance of sequences in other regions (Figure 3). The first example is where the corresponding miRNA* is more abundant than the annotated miRNA (Figure 3A). The miRNA* is thought to degrade rapidly when the mature miRNA is selectively incorporated into effector complexes for target recognition (KIM, 2005). However, it is possible that miRNA and miRNA* are selected and could silence

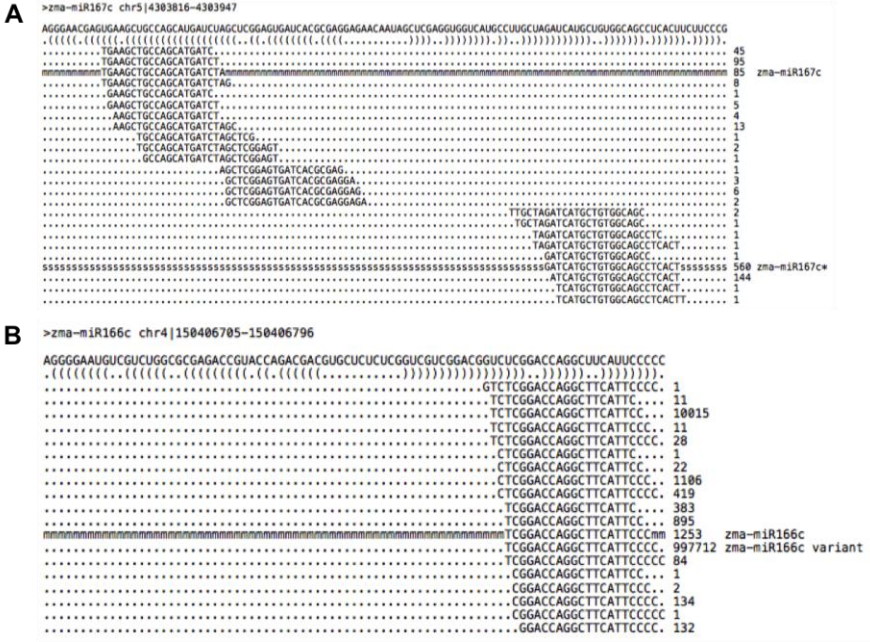
different targets (ZHANG et al. 2011). Yi et al. (2013) have also found similar results when analyzing rice plants, and suggested that the miRNA* might be the genuine product of the pre-miRNA or that both the miRNA and miRNA* are functional in regulating gene expression. The second example is where the most abundant sRNA is not the annotated miRNA or the miRNA*, but one of its variants (Figure 3B). Other plant studies have also identified variants within 1-2 nt range from annotated miRNAs (YI et al. 2013; LI et al. 2011; ZHU et al. 2008). Yi et al. (2013) suggest two possible explanations for these findings: i) these variants are simply sequencing errors, where the detected sRNAs are degradation products from unprocessed precursors; or ii) these sRNAs could be authentic miRNAs, and, therefore, they should be substituted for the annotated miRNAs in the databases. In any case, these findings should be thoroughly investigated in order to understand their true biological meaning.

Figure 2. Alignment of the reads with the reference genome. A) Distribution of the identified conserved miRNA along maize miRNA families. Numbers above the bars represent the number of miRNA identified for the respective miRNA family. B) Number of conserved miRNA identified in each variety.



Source: Author

Figure 3. Distribution of sRNAs along conserved miRNAs precursors. A) Example where the corresponding miRNA* is more abundant than the annotated miRNA. B) Example where the most abundant sRNA is not the annotated miRNA or the miRNA*, but one of its variants.



Source: Author

3.2. miRNA expression profiles in GM and non-GM varieties

Based on the normalized read count (performed using DESeq R package) for each identified miRNA, differential expression analysis was performed and 13 known maize miRNAs were found to show statistically significant changes (based on *p-adjusted* < 0.05) for different pair wise comparisons (Table 4). These differentially expressed miRNAs were quantified using SYBR real-time RT-PCR to corroborate the expression profiles obtained from Illumina sequencing. However, since the forward miRNAs primers were designed based on the full mature miRNA sequences and the reverse primer was the universal reverse primer for miRNA, some miRNA belonging to the same family could not be distinguished due its homology to the mature miRNA sequence. Therefore, information about the primers sequence and miRNA amplified by each primer can be seen in the Additional file 2. The expression profiles generated for most of the selected miRNAs were the same as those determined by Illumina sequencing, indicating that the sequencing data produced in this study were reliable and could be subjected to further analysis. The comparison of Illumina and real-time RT-PCR quantification results is shown in Table 5.

We have used the logarithm base 2 of the fold-change (Log2FC) as a measure of expression changes. Log2FC is the most widely used alternative transformation of the ratio, once it has the advantage of producing a continuous spectrum of values and treating up- and down-regulated genes in a similar fashion. Therefore, a gene up-regulated by a factor of 2 has a Log2FC of 1, a gene down-regulated by a factor of 2 has a Log2FC of −1, and a gene expressed at a constant level (with a FC of 1) has a Log2FC equal to zero (QUACKENBUSH, 2002).

Table 4. Log2Fold-Change, *p*- and *p-adjusted* values of the differentially expressed miRNAs.

miRNA	Comparison	log2Fold-Change	*p-adj* value
zma-MIR162	Conv *vs* Bt	0.5330	0.0275
zma-MIR167c	Conv *vs* Bt	0.5654	0.0266
zma-MIR167e	Conv *vs* Bt	0.5357	0.0275
zma-MIR167j	Conv *vs* Bt	0.5696	0.0266
zma-MIR169f	Conv *vs* Bt	-0.7750	0.0266
zma-MIR169m	Conv *vs* Bt	0.6216	0.0275
zma-MIR399a	Conv *vs* Bt	-0.7217	0.0275
zma-MIR399e	RR *vs* RRxBt	-0.8987	0.0044
zma-MIR399h	Conv *vs* Bt	-0.6764	0.0275
zma-MIR399i	Conv *vs* RRxBt	-0.6493	0.0311
	RR *vs* RRxBt	-0.9620	0.0007
zma-MIR399j	Conv *vs* RRxBt	-0.6330	0.0232
	RR *vs* RRxBt	-0.7307	0.0310
zma-MIR529	Conv *vs* Bt	0.7354	0.0275
zma-MIR827	Conv *vs* RRxBt	-0.5376	0.0006
	RR *vs* RRxBt	-0.4982	0.0310

Table 5. Comparison of Illumina and real-time RT-PCR quantification results.

miRNA	Comparison	FoldChange RT-qPCR	FoldChange RNA-seq
zma-MIR167ej	Conv vs Bt	1.0771	1.4669
zma-MIR169f	Conv vs Bt	0.4825	0.5844
zma-MIR169m	Conv vs Bt	1.3343	1.5385
zma-MIR399ah	Conv vs Bt	0.8194	0.6161
zma-MIR399eij	Conv vs Stacked	0.3504	0.6412
zma-MIR399eij	RR vs Stacked	1.0034	0.5508
zma-MIR827	Conv vs Stacked	0.6703	0.6889
zma-MIR827	RR vs Stacked	0.5004	0.7080

3.3. Target prediction and functional analysis

One of the challenges in elucidating the biological functions of miRNAs is to identify their regulatory targets. We have predicted targets for the 13 differentially regulated miRNAs using psRNAtarget and found 118 potential endogenous transcript targets. These included splicing variants and gene expression inhibition by either mRNA cleavage or translational inhibition. The miRNAs with the most targets were zma-miR529-5p (47) and zma-miR169f-5p (25). A singular enrichment analysis was performed for the 13 selected miRNA using AgriGO (Figure 4). The analysis showed 43 significant GO terms (FDR < 0.05), among these terms, binding (GO:0005488) had the higher number of assigned transcripts within the Molecular Function category, with 55 transcripts assigned to it. Other significant GO terms in this category were nucleic acid binding (GO:0003676), DNA binding (GO:0003677), transcription regulator activity (GO:0030528) and transcription factor activity (GO:0003700). Additionally, a high percentage of the targets were assigned to Cellular Processes (GO:0005623, GO:0044464, GO:0005622) and also Biological Processes, such as metabolic process (GO:0008152), cellular process (GO:0009987), primary metabolic process (GO:0044238), nitrogen compound (GO:0006807) and gene expression (GO:0010467) (Figure 4).

A great amount of putative targets of maize conserved miRNA identified were transcriptions factors (TFs). Among these TFs, we found squamosa promoter binding protein (SBP)-like (SPL) genes, which are targets of the MIR529. These same TFs are targeted by the same family in other species (*Aquilegia coerulea* – PUZEI; KRAMER, 2009), but are also targets of other miRNA families, such as MIR156 in *A. thaliana* (SHIKATA et al. 2009; WU et al. 2009) and *Vrisea carinata* (Guzmán et al. 2013). These TFs are known to affect a range of developmental processes such as leaf development, shoot maturation, phase change and flowering in plants (SHIKATA et al. 2009; WU et al. 2009). The zma-MIR529, which was found to target SBP-like genes, was found to be up regulated in the Bt samples when compared to the near-isogenic non-GM variety.

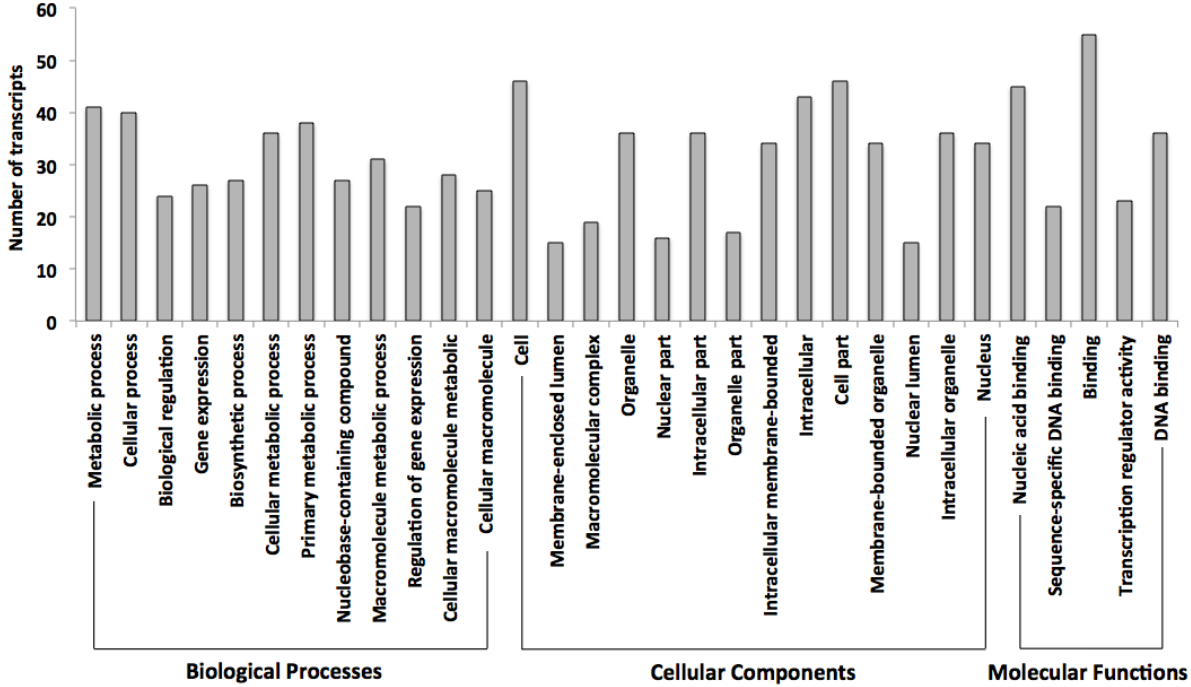
Another relevant TF identified as miRNA target is the nuclear transcription factor Y gene. Nuclear factor Y (NF-Y) is a ubiquitous transcription factor known to bind specifically to a highly conserved CCAAT sequence in the promoters of their target genes (FRONTINI et al. 2004). In addition, studies in *A. thaliana* showed that the down

regulation of MIR169a by drought stress contributes to the high level induction of a NF-Y (*NFYA5*) by drought and ABA. Moreover, the authors concluded that NFYA5 was important in controlling stomatal aperture and drought resistance and it is regulated by drought stress not only transcriptionally, but also posttranscriptionally via a miRNA pathway (LI et al. 2008). Furthermore, two miRNAs belonging to the MIR169 family were differentially expressed in the Bt samples when compared with the Conv samples: zma-MIR169f was down regulated and zma-MIR169m was up regulated. bZIP transcription factor was also found to be a miRNA target, specifically for the zma-MIR827, which was down-regulated in RRxBt samples when compared to Conv and RR samples. bZIP TFs play important roles in ABA signal transduction pathways (KANG et al. 2002).

We have also identified auxin response factor (ARF)-related protein, which is a plant-specific family of DNA binding proteins and are targets of MIR160 and MIR167 families (YANG et al. 2006; WU et al. 2006; GUZMÁN et al. 2013). ARF proteins bind to auxin response promoter elements and mediate gene expression responses to the plant hormone auxin (MALLORY et al. 2005). Three members of the MIR167 were down regulated in the Bt samples when compared to the Conv samples: zma-MIR167c, zma-MIR167e and zma-MIR167j. Another target identified belongs to pentatricopeptide repeat genes (PPR), which are targets of the MIR529 family in our study, but are also targeted by MIR156 and MIR396 families (GUZMÁN et al. 2013). PPR proteins form sequence-specific associations with RNA, which could affect folding, processing and translation of the RNA, thus regulating the expression of transcripts (FUJII; SMALL, 2011).

Overall, the results showed that 13 miRNAs were differentially modulated in different pair wise comparisons, and these miRNAs can target a range of endogenous transcripts, mostly transcription factors. These TFs are shown to participate in several biological processes, such as leaf development, shoot maturation, flowering, stomatal aperture, mechanism of abiotic resistance and hormone signaling.

Figure 4. Significantly enriched pathways for the 13 differentially expressed conserved miRNA targets. Differentially regulated pathways (FDR < 0.05) found in the agriGO online tool were submitted to removal of redundant GO terms using REVIGO online tool.



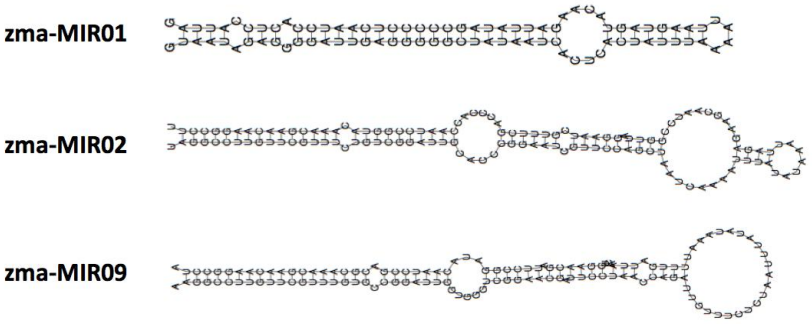
Source: Author

3.4. Prediction of novel miRNAs

In order to identify possible novel miRNAs in our samples, we have used miR-PREFeR pipeline, together with MiPred online tool to predict their pre-miRNA structure. A total of 20 putative novel miRNA have been identified in one or more of the varieties (Table 6). The novel miRNAs were temporarily named following the zma-MIR-number format, e.g., zma-MIR01, before being submitted to obtain an official designation. The predicted structure of three of the pre-miRNA is shown in Figure 5. Out of the 20 novel miRNAs, two were exclusively detected in Conv samples (zma-MIR03 and zma-MIR14), one in the RR samples (zma-MIR11) and one in the Bt samples (zma-MIR13). Although some of the putative novel miRNAs share homology in the mature sequence (zma-MIR03 / zma-MIR14 and zma-MIR04 / zma-MIR08), their pre-miRNAs are not the same and they are located in different locus in the genome (see sequences with a '*' and '**' in Table 6).

Several low-expressed miRNAs have been discovered in plant species using RNA-Seq, which indicates that, besides the well-known conserved miRNA, each species has its own specific miRNAs, which could play a range of biological processes (JIA et al. 2014; GUZMÁN et al. 2013), and they could vary according to biotic and abiotic situations (SHENG et al. 2015; NAVARRO et al. 2006). Jia et al. (2014) suggest that the interaction of novel miRNAs and conserved miRNAs might regulate metabolic processes in the plant more broadly and accurately than either of them alone. In light of this, the identification of novel miRNAs being expressed in GM plants might represent a biosafety concern, since they could interact in unique ways of what it is previously known.

Figure 5. Prediction of novel maize miRNA. Examples of the structures of three predicted novel maize miRNAs. The prediction was performed using the PREFeR pipeline, which uses RNAfold (from the ViennaRNA package) algorithm to calculate and draw the pre-miRNA hairpin structures.



Source: Author

A prediction of the novel miRNA targets showed 420 potential endogenous transcript targets, including splicing variants and gene expression inhibition by either mRNA cleavage or translational inhibition. These 420 transcripts were submitted to singular enrichment analysis using AgriGO. Only two GO terms (FDR < 0.05) were significantly enriched in the miRNA targets of the predicted novel miRNAs, enzyme regulator activity (GO:0030234) and endomembrane system (GO:0012505). Three of the transcripts assigned to the enzyme regulatory activity GO term of Serine/threonine protein phosphatase regulatory subunit Bbeta and two from RabGAP/TBC domain-containing protein. Serine/threonine protein phosphatases have been identified as miRNA targets in previous plant studies (GUZMÁN et al. 2012) and it seems to play major roles in resistance response to pathogen, drought tolerance mechanisms and cold stress (PAÍS et al. 2009; XU et al. 2007). The RabGAP/TBC genes were also identified as miRNA targets and are associated with signal transduction and also involved in resistance response to pathogens (FENG et al. 2015; PALMIERI et al. 2012).

These putative novel maize miRNA have not been deposited in miRbase yet, since no further confirmation of their existence has been conducted. Although we have seen that Illumina sequencing and real time RT-qPCR results were in accordance for the conserved miRNA identified in this study, quantification by RT-qPCR of the putative novel miRNA would bring a more reliable confirmation of the results.

Table 6. Location in the genome, miRNA sequence and number of mapped reads for the 20 putative novel miRNAs identified in maize varieties. miRNA sequences followed by '*' and '**' show identical mature sequences, but their pre-miRNA have different sequences.

miRNA ID	Location on genome	miRNA sequence	Reads matching mature sequence				
			RR	Bt	RRxBt	Conv	Land
zma-MIR01	chr01|95866804:95866892	AUGGAGUGGAUUGAGGGGGCU	36	51	43	50	34
zma-MIR02	chr01|224709986:224710116	AUCCGGUACAAACGAACAAGGCCU	168	179	162	130	97
zma-MIR03	chr10|5127325:5127461	AGAGUGGACAGUUGACGCCGGCCC*	0	0	0	152	0
zma-MIR04	chr10|12355028:12355155	AAUACAUGUGGAUUGAGCUCAAUA**	42	57	43	45	45
zma-MIR05	chr10|71614450:71614580	AUCCGACAGAAACGAACAAGGCCU	615	0	0	607	0
zma-MIR06	chr10|120859137:120859262	UAUUCGAGAACGGAUGUAGUACAU	546	672	655	525	138
zma-MIR07	chr10|145006838:145006964	AUUAGGGUAGAACCGAACAAGCCU	50	55	56	65	59
zma-MIR08	chr02|30270820:30270916	AAUACAUGUGGAUUGAGCUCAAUA**	42	57	43	45	45
zma-MIR09	chr02|144495862:144495992	AUCCGACGCAAACGAACAAGGCCU	78	109	102	68	99
zma-MIR10	chr02|203809131:203809256	AGGGUAUUGAUAGGACUAUAAUCC	352	351	323	373	151
zma-MIR11	chr02|229799695:229799756	GGGGAUGUAGUUCAGAUGGUAGAA	5893	0	0	0	0
zma-MIR12	chr03|162601521:162601594	UGUUUGGGAUUAUAAUCUGCC	47	71	48	56	41
zma-MIR13	chr03|178085743:178085863	AAAUACUGUAGAAGCCGCAGCCGC	0	2937	0	0	0
zma-MIR14	chr04|100693925:100694040	AGAGUGGACAGUUGACGCCGGCCC*	0	0	0	152	0
zma-MIR15	chr05|144731639:144731774	ACGAGAGAGGACGUCAGGGGACGA	11	29	27	27	15

Continuation Table 6.

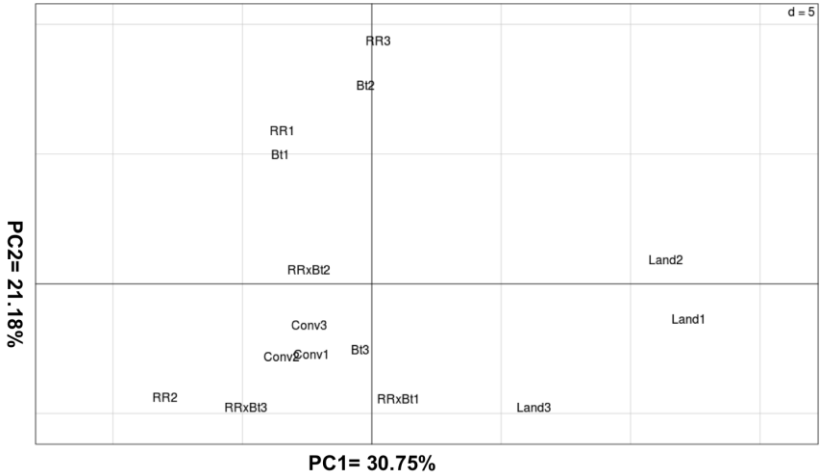
miRNA ID	Location on genome	miRNA sequence	Reads matching mature sequence				
			RR	Bt	RRxBt	Conv	Land
zma-MIR16	chr05|174841383:174841471 chr05|174842606:174842694 chr05|174912223:174912311 chr05|174938499:174938587	CUGAGCAAAAAAACACGACUAAG	21	26	23	16	10
zma-MIR17	chr07|104640325:104640456	AUUCCGGAACAAACGAACACACCC	18	17	18	28	14
zma-MIR18	chr07|123915132:123915213	UUUGAGAUUCGUAGCUUUUAC	77	79	95	77	90
zma-MIR19	chr08|61110522:61110652	AUCUGACACAAACGAACAAGGCCU	82	92	94	77	57
zma-MIR20	chr08|171925082:171925167	ACGGAUCAAAUCUAUGGUGAGAUU	53	78	59	58	23

3.5. Landrace shows different miRNA expression patterns

A Landrace variety was included in the miRNA differential analysis in order to have a different maize genetic background in the dataset. As the main point of the Landrace samples is to address the natural variation of miRNA expression, this section will not cover differences in specific metabolic pathways or enrichment analysis of gene targets. Four comparisons were performed in order to evaluate possible miRNA differential expression between varieties: Conv *vs* Land, Bt *vs* Land, RR *vs* Land and RRxBt *vs* Land. From the 146 conserved miRNAs identified in our dataset, 26 of them were differentially expressed in one or more comparisons. Those 26 miRNAs represent 12 different miRNA families, with the majority of them belong to the zma-MIR169 family (seven miRNAs), followed by zma-MIR167 and zma-MIR399, both with three miRNAs. For the Conv *vs* Land comparison, 7 miRNAs were up regulated and 13 were down regulated in the Landrace variety; Bt *vs* Land showed 6 up regulated and 6 down regulated miRNAs in the Landrace variety; in the RR *vs* Land and RRxBt *vs* Land comparisons, 6 miRNAs showed a up regulation and 4 showed a down regulation in the Land variety in both cases (Additional file 3).

We have performed a dimensionality reduction strategy, the PCA analysis, with all the samples in order to explore the relationship between these varieties. The PCA plot (PC1 x PC2) shows that Landrace samples are completely separated from the high-breeding varieties, which accounted for the majority of the variation in the dataset (30.75%) (Figure 6). Our previous proteomics study using the same transgenic varieties showed similar PCA results, where 15.6% of the variation (PC2 x PC3) was explained by the Landrace variety (AGAPITO-TENFEN et al. 2014). In addition, our transcriptomic analysis with the same varieties have also shown a separation of the Landrace samples from the high-breeding ones, accounting for 17.32% of the variation in the dataset (not published). A landrace variety was also included in a comparative analysis of potato tuber proteomes of GM potato varieties (LEHESRANTA et al. 2005). Different varieties and landraces showed variation in the protein profile, which indicates extensive genotypic variation. Most of the proteins detected exhibited significant quantitative and qualitative differences between one or more GM varieties and landraces.

Figure 6. Principal Component Analysis of Landrace variety and high-breeding varieties. Normalized read counts (performed by DESeq2 package) were used for PCA analysis.



Source: Author

Although the use of non-GM varieties that are genetically distant from the GM event under investigation is not a requirement of international guidelines addressing the issue for comparative assessments of the GM crops (AHTEG, 2010), we have included a landrace variety in this study in order to provide valuable information of miRNA expression variation related to different maize genetic backgrounds, as well as to possibly disclose differences in GM lines that might fit within the variation observed in non-modified materials (AGAPITO-TENFEN et al. 2014). In addition, the data accumulated so far could be used in future studies that will address the effects of the presence of transgenes in the Landrace Pixurum 5.

4. CONCLUSIONS

In conclusion, our results show that 13 endogenous conserved miRNA were differentially modulated in some of the pair wise comparisons analyzed. Most of these miRNA showed to be differentially expressed in the Bt event (expressing insecticidal CRY proteins) in comparison to the near-isogenic non-GM line, which showed to target endogenous transcription factors responsible for several biological processes, such as leaf development, shoot maturation, phase change and flowering, control of stomatal aperture, mechanism of stress resistance, hormone signaling and RNA folding, processing and translation. In addition, differentially expressed miRNAs were also found in the stacked GM event (expressing both insecticidal CRY and herbicide tolerant EPSPS proteins) in comparison to the single GM (expressing herbicide tolerant EPSPS protein) and near-isogenic non-GM line. These miRNAs showed to target transcription factors involved in ABA signal transduction pathways. Novel miRNAs were also detected in our miRNA dataset. The identification of these novel miRNAs might represent a biosafety concern, since they could interact with the endogenous conserved miRNAs and regulate several metabolic processes in a completely different manner. The novel miRNAs showed to target endogenous transcripts involved in resistance response to pathogen, mechanisms of abiotic stress and signal transduction.

To the best of our knowledge, this study is the first one to report changes in miRNA expression in commercial GM crops and their role in regulating gene expression. These findings demonstrate that novel molecular profiling approaches could be applied as a useful tool in ways to increase confidence in risk assessments (HEINEMANN et al. 2011). Although the use of profiling techniques are not a consensus of whether they add a valuable information for risk assessment or not (PAUWELS et al. 2015), such changes in miRNA expression should be taken in account in risk assessments analysis in order to address the biological meaning of these changes. Moreover, the Landrace Pixurum 5 miRNA analysis revealed that a variety outside of GM genetic background and with higher genetic variability would be useful to future biosafety studies, especially those aiming to address adaptation issues. Finally, in the absence of the adequate comparator, as research independent groups are facing, a well-characterized landrace would be the single option in their studies.

5. REFERENCES

AD HOC TECHNICAL EXPERT GROUP (AHTEG): United Nations Environment Programme, CBD: Guidance document on risk assessment of living modified organisms, 2010.

AGAPITO-TENFEN, S.Z.; GUERRA, M.P.; WIKMARK, O.G.; NODARI, R.O. Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. Proteome Sci, v.11, 46, 2013.

AGAPITO-TENFEN, S.Z.; VILPERTE, V.; BENEVENUTO, R.F.; ROVER, C.M.; TRAAVIK, T.I.; NODARI, R.O. Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. BMC Plant Biology, v.14, 346, 2014.

BARROS, E.; LEZAR, S.; ANTTONEN, M.J.; VAN DIJK, J.P.; RÖHLIG, R.M.; KOK, E.J. et al. Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnol J, v.8, 436–451, 2010.

BUSTIN, S.A.; BENES, V.; GARSON, J.A.; HELLEMANS, J.; HUGGETT, J.; KUBISTA, M. et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem, v.55, 611–622, 2009.

CANCI, I.J.; BRASSIANI, I.A. Anchieta História, Memória e Experiência: Uma caminhada construída pelo povo. São Miguel do Oeste: Mclee; 2004.

CARTHEW, R. W.; SONTHEIMER, E. J. Origins and mechanisms of miRNAs and siRNAs. Cell, v.136, 642–655, 2009.

CODEX ALIMENTARIUS COMMISSION. Guideline for the conduct of food safety assessment of food derived from recombinant DNA plants. Rome: CAC/GL 45–2003; 2003.

CONVENTION ON BIOLOGICAL DIVERSITY. Guidance On Risk Assessment Of Living Modified Organisms. Hyderabad, India, 1-5 October, 2012.

COLL, A.; NADAL, A.; COLLADO, R.; CAPELLADES, G.; KUBISTA, M.; MESSEGUER, J. et al. Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. Plant Mol Biol, v.73, 349–362, 2010.

Comissão Técnica Nacional de Biossegurança (CTNBIO). Parecer Técnico n° 1596/2008 - Liberação Comercial de Milho Geneticamente Modificado Tolerante ao Glifosato, Milho Roundup Ready 2, Evento NK603 - Processo n° 01200.002293/2004- 16. [http://www.ctnbio.gov.br/index.php/content/view/12341.html]

Comissão Técnica Nacional de Biossegurança (CTNBIO). Parecer Técnico n° 2052/2009 - Liberação Comercial de Milho Resistente a Insetos, Milho MON 89034 - Processo n° 01200.003326/2008-61. [http://www.ctnbio.gov.br/index.php/ content/view/14098.html]

Comissão Técnica Nacional de Biossegurança (CTNBIO). Parecer Técnico n° 2725/2010 - Liberação Comercial de Milho Geneticamente Modificado Resistente a Insetos e Tolerante a Herbicidas, Milho MON 89034 x NK 603 - Processo n° 01200.003952/ 2009-38. [http://www.ctnbio.gov.br/index.php/content/view/15704.html]

CONVENTION ON BIOLOGICAL DIVERSITY. Cartagena Protocol on Biosafety to the Convention on Biological Diversity: text and annexes. Montreal: Secretariat of the Convention on Biological Diversity, 30p., 2000.

DAI, X.; ZHAO, P.X. psRNATarget: A Plant Small RNA Target Analysis Server, Nucleic Acids Research, 2011.

DE SCHRIJVER, A.; DEVOS, Y.; VAN DEN BULCKE, M.; CADOT, P.; DE LOOSED, M.; REHEUL, D. et al. Risk assessment of GM stacked events obtained from crosses between GM events. Trends Food Sci Tech, v.18, 101–109, 2007.

DU, Z.; ZHOU, X.; LING, Y.; ZHANG, Z.; SU, Z. agriGO: a GO analysis toolkit for the agricultural community. Nucl Acids Res, v.38, 64-70, 2010.

European Food Safety Authority (EFSA). Guidance Document of the

Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants containing stacked transformation events. EFSA J, v.512, 1–5, 2007.

European Food Safety Authority (EFSA). Guidance Document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants containing stacked transformation events. EFSA J, v.512, 1–5, 2007.

FAHLGREN, N.; HOWELL, M.D.; KASSCHAU, K.D.; CHAPMAN, E.J.; SULLIVAN, C.M.; CUMBIE, J.S. et al. Carrington, High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE, v.2, e219, 2007.

FENG, H.; WANG, B.; ZHANG, Q.; FU, Y.; HUANG, L.; WANG, X. et al. Exploration of microRNAs and their targets engaging in the resistance interaction between wheat and stripe rust. Front Plant Sci. v.6, 469, 2015.

FRONTINI, M.; IMBRIANO, C.; MANNI, I.; MANTOVANI, R. Cell cycle regulation of NF-YC nuclear localization. Cell Cycle, v.3, 217–222, 2004.

FUJII, S.; SMALL, I. The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol, v.191, 37–47, 2011.

GUZMÁN, F.; ALMERÃO, M.P.; KORBES, A.P.; CHRISTOFF, A.P.; ZANELLA, C.M.; BERED, F. et al. Identification of potential miRNAs and their targets in Vriesea carinata (Poales, Bromeliaceae). Plant Sci, v.210, 214-23, 2013.

GUZMÁN, F.; ALMERÃO, M.P.; KORBES, A.P.; LOSS-MORAIS, G.; MARGIS, R. Identifica- tion of microRNAs from Eugenia uniflora by high-throughput sequencing and bioinformatics analysis. PLoS ONE, v.7, e49811, 2012.

HEINEMANN, J.A.; KURENBACH, B.; QUIST, D. Molecular profiling - a tool for addressing emerging gaps in the comparative risk assessment of GMOs. Environ Int, 37, 1285–1293, 2011.

HERNÁNDEZ, M.; PLA, M.; ESTEVE, T.; PRAT, S.;

PUIGDOMENECH, P.; FERRANDO, A. A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard based on the 3'-transgene integration sequence. Transgenic Res, v.12, 179–89, 2003.

HOLCK, A.; VA, M.; DIDIERJEAN, L.; RUDI, K. 5'-Nuclease PCR for quantitative event-specific detection of the genetically modified Mon810 MaisGard maize. Eur Food Res Technol, v.214, 449–53, 2002.

JAMES, C. Global Status of Commercialized Biotech/GM Crops: 2014. ISAAA Brief No. 49. Ithaca: ISAAA Press; 2014.

JIANG, P.; WU, H.; WANG, W.; MA, W.; SUN, X; LU, Z. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res, v.35, W339-W344, 2007.

JIA, L.; ZHANG, D.; QI, X.; MA, B.; XIANG, Z.; HE, N. Identification of the Conserved and Novel miRNAs in Mulberry by High-Throughput Sequencing. PLoS One, v.9(8), e104409, 2014.

KANG, J.Y.; CHOI, H.I.; IM, M.Y.; KIM S.Y. Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell, v.14, 343–357, 2002.

KERSEY, P. J.; ALLEN, J. E.; CHRISTENSEN, M. et al. Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic acids research, v.42, p.546-552, 2014.

KIM, V.N. Small RNAs: classification, biogenesis, and function. Mol Cells, v.19, 1–15, 2005.

KROL, J.; LOEDIGE, I., FILIPOWICZ, W. The widespread regulation of microRNA biogenesis, function and decay. Nature Rev. Genet, v.11, 597–610, 2010.

LANGMEAD, B.; SALZBERG, S. Fast gapped-read alignment with Bowtie 2. Nature Methods, v.9, p.357-359, 2012.

LA PAZ, J.L.; VICIENT, C.; PUIGDOMÈNECH, P.; PLA, M. Characterization of polyadenylated cryIA(b) transcripts in maize

MON810 commercial varieties. Anal Bioanal Chem, v.396, 2125–33, 2010.

LEI, J.; SUN, Y. miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data. Bioinformatics, v.30(19), 2837-92014, 2014.

LI, T.; LI, H.; ZHANG, Y.X.; LIU, J.Y. Identification and analysis of seven H2O2- responsive miRNAs and 32 new miRNAs in the seedlings of rice (Oryza sativa L. ssp. indica). Nucleic Acids Res, v.39, 2821–2833, 2011.

LI, W.; OONO, Y.; ZHU, J.; HE, X.; WU, J.; IIDA, K. et al. The arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. Plant Cell, v.20, 2238–2251, 2008.

LIU, H.; QIN, C.; CHEN, Z.; ZUO, T.; YANG, X.; ZHOU, H. Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. BMC Genomics, v.15, 25, 2014.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv, 2014.

MALLORY, A.C.; BARTEL, D.P.; BARTEL B. MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. Plant Cell, v.17(5), 1360-75, 2005.

NAVARRO, L.; DUNOYER, P.; JAY, F.; ARNOLD, B.; DHARMASIRI, N.; ESTELLE, M. et al. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science, v.312, 436–439, 2006.

NGUYEN, H.T.; JEHLE, J.A. Quantitative analysis of the seasonal and tissue specific expression of Cry1Ab in transgenic maize MON 810. J Plant Dis Protect, v.114, 82–7, 2007.

NODINE, M.D.; BARTEL, D.P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryo- genesis.

Genes Dev, v.24, 2678–2692, 2010.

NOZAWA, M.; MIURA, S.; NEI, M. Origins and evolution of microRNA genes in plant species. Genome Biol Evol, v.4, 230–239, 2012.

O'NEIL, D.; GLOWATZ, H.; SCHLUMPBERGER, M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. Curr Protoc Mol Biol, Chapter 4:Unit 4.19, 2013.

PAÍS, S.M.; GONZÁLEZ, M.A.; TÉLLEZ-IÑÓN, M.T.; CAPIATI, D.A. Characterization of potato (Solanum tuberosum) and tomato (Solanum lycopersicum) protein phosphatases type 2A catalytic subunits and their involvement in stress responses. Planta, v.230, 13-25, 2009.

PALMIERI, M.C.; PERAZZOLLI, M.; MATAFORA, V.; MORETTO, M.; BACHI, A.; PERTOT, I. Proteomic analysis of grapevine resistance induced by Trichoderma harzianum T39 reveals specific defence pathways activated against downy mildew. J Exp Bot, v.63, 6237–6251, 2012.

PAUWELS, K.; DE KEERSMAECKER, S.C.J.; DE SCHRIJVER, A.; DU JARDIN, P.; ROOSENS, N.H.C.; HERMANA, P. Next-generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: Added value or not? Trends in Food Science & Technology, v.45(2), 319–326, 2015.

PUZEY, J.R.; KRAMER, E.M. Identification of conserved Aquilegia coerulea microRNAs and their targets. Gene, v.448(1), 46-56, 2009.

QUACKENBUSH, J. Microarray data normalization and transformation. Nat Genet, v.32, 496-501, 2002.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

RAMESH, S.V.; RATNAPARKHE, M.B.; KUMAWAT, G.; GUPTA, G.K.; HUSAIN, S.M. Plant miRNAome and antiviral resistance: a retrospective view and prospective challenges. Virus Genes, v.48(1), 1-14, 2014.

REINHART, B.J.; WEINSTEIN, E.G.; RHOADES, M.W.; BARTEL, B.; BARTEL, D.P. MicroRNAs in plants. Genes Dev, v.16, 1616–1626, 2002.

ROSATI, A.; BOGANI, P.; SANTARLASCI, A.; BUIATTI, M. Characterisation of 3' transgene insertion site and derived mRNAs in MON810 YieldGard maize. Plant Mol Biol, v.67, 271–81, 2008.

SHENG, L.; CHAI, W.; GONG, X.; ZHOU, L.; CAI, R.; LI, X. et al. Identification and Characterization of Novel Maize Mirnas Involved in Different Genetic Background Int J Biol Sci, v.11(7), 781-793, 2015.

SHIKATA, M.; KOYAMA, T.; MITSUDA, N.; OHME-TAKAGI, M. Arabidopsis SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. Plant Cell Physiol, v.50, 2133–2145, 2009.

SUNKAR, R.; ZHU, J. Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell, v.16, 2001–2019, 2004.

SUPEK, F.; BOŠNJAK, M.; ŠKUNCA, N.; ŠMUC T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. PLoS One, v.6(7), e21800, 2011.

TAVERNIERS, I.; PAPAZOVA, N.; BERTHEAU, Y.; DE LOOSE, M.; HOLST-JENSEN, A. Gene stacking in transgenic plants: towards compliance between definitions, terminology, and detection within the EU regulatory framework. Environ Biosafety Res, v.7, 197-218, 2008.

TRTIKOVA, M.; WIKMARK, O.; ZEMP, N.; WIDMER, A.; HILBECK A. Transgene Expression and Bt Protein Content in Transgenic Bt Maize (MON810) under Optimal and Stressful Environmental Conditions. PLoS ONE, v.10(4), e0123011, 2015.

WINDELS, P.; TAVERNIERS, I.; DEPICKER, A.; VAN BOCKSTAELE, E.; DE LOOSE, M. Characterisation of the Roundup Ready soybean insert. Eur Food Res and Technol, v.213, 107–12, 2001.

WU, G.; PARK, M.Y.; CONWAY, S.R.; WANG, J-W.; WEIGEL, D.; POETHIG, R.S. The sequential action of miR156 and miR172 regulates

developmental timing in Arabidopsis. Cell, v.138, 750–759, 2009.

WU, M-F.; TIAN, Q.; REED, J.W. Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. Development (Cambr., Engl.), v.133, 4211–4218, 2006.

YANG, J.H.; HAN, S.J.; YOON, E.K.; LEE, W.S. Evidence of an auxin signal pathway, microRNA167-ARF8-GH3, and its response to exogenous auxin in cultured rice cells. Nucleic Acids Res, v.34, 1892–1899, 2006.

YI, R.; ZHU, Z.; HU, J.; QIAN, Q.; DAI, J.; DING, Y. Identification and expression analysis of microRNAs at the grain filling stage in rice (Oryza sativa L.) via deep sequencing. PloS one, v.8(3), e57863, 2013.

ZHANG, X.M.; ZHAO, H.W.; GAO, S.; WANG, W.C.; KATIYAR-AGARWAL, S.; HUANG, H.D. et al. Arabidopsis Argonaute 2 regulates innate immunity via miRNA393(*)-mediated silencing of a Golgi-localized SNARE gene, MEMB12. Mol Cell, v.42, 356–366, 2011.

ZHAO, C-Z.; XIA, H.; FRAZIER, T.P.; YAO, Y-Y.; BI, Y-P.; LI, A-Q. Deep sequencing identifies novel and conserved microRNAs in peanuts (Arachis hypogaea L.). BMC Plant Biol, v.10(3), 2010.

ZHAO, Y.; GUO, L.; WANG, H.; HUANG, D. Integration and Expression Stability of Transgenes in Hybriding Transmission of Transgenic Rice Plants Produced by Particle Bombardment. Molecular Plant Breeding, v.2(8), 48-59, 2011.

ZHU, Q-H.; SPRIGGS, A.; MATTHEW, L.; FAN, L.; KENNEDY, G.; GUBLER, F. et al. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res, v.18, 1456–1465, 2008.

ZOLLA, L.; RINALDUCCI, S.; ANTONIOLI, P.; RIGHETTI, P.G. Proteomics as a complementary tool for identifying unintended side effects occuring in transgenic maize seeds as a result of genetic modifications. J Proteome Res, v.7, 1850–1861, 2008.

7. CONCLUSÕES

Em conclusão, foi possível observar neste trabalho que o perfil de expressão de transcritos endógenos de amostras de milho contendo eventos simples (expressando proteínas inseticidas CRY ou enzimas EPSPS tolerantes a herbicidas à base de glifosato) apresentou diferenças significativas em relação à linha isogênica convencional. A Análise de Componentes Principais (PCA) mostrou que a presença dos transgenes foi o fator que mais afetou a divergências entre as amostras. Além disso, diversos processos metabólicos apresentaram diferenças em sua regulação, como vias de redox, glicosilação e fosforilação de proteínas, atividade de transferase, entre outros. As amostras contendo eventos estaqueados também diferiram, em seu perfil transcriptômico, das amostras contendo eventos simples e da linha isogênica convencional. Estas amostras agruparam separadamente na análise de PCA, sendo que a presença dos dois transgenes nas amostras foi o fator que mais contribuiu para a sua separação das amostras de evento simples e isolinha convencional. Os processos metabólicos que mais apresentaram alterações nos eventos estaqueados estão relacionados com a regulação da transcrição, como por exemplo, DNA *binding*, atividade de kinase de serinas/treoninas e ATP *binding*.

Os perfis de miRNAs também apresentaram diferenças de expressão de miRNAs específicos para algumas das comparações realizadas. As maiores diferenças foram encontradas para as amostras do evento simples contendo proteínas CRY quando comparadas com a isolinha convencional. Além disso, também foram observadas diferenças na expressão de miRNAs presente nas amostras do evento estaqueado em relação ao evento simples contendo as proteínas EPSPS e também à isolinha convencional. Estes miRNAs são capazes de regular a expressão de diversos transcritos endógenos, dos quais fatores de transcrição foram os alvos de miRNA mais encontrados neste estudo. Estes fatores de transcrição estão envolvidos em diversos processos metabólicos da planta, como por exemplo, desenvolvimento foliar, florescimento, resposta à estresse abiótico e transdução de sinais. Outros alvos de miRNA encontrados mostraram estar envolvidos com fatores de resposta de auxinas e com o processamento, arranjo e tradução de RNA. Também foram detectados 20 novos miRNA putativos nas variedades de milho estudadas, os quais ainda não estão depositados em bancos de dados de miRNAs. Transcritos envolvidos em vias de atividade enzimáticas mostraram ser os principais alvos destes miRNAs, desempenhando papéis importantes em respostas de resistência à

patógenos mecanismos de tolerância à estresse abiótico (seca e frio).

Outro resultado importante deste trabalho foi a inclusão de uma variedade crioula nas análises. O perfis transcriptômicos e de miRNA apresentaram grandes diferenças nas amostras da variedade crioula em relação às amostras que passaram por melhoramento genético. Em ambas abordagens, a análise de PCA mostrou um agrupamento das amostras crioulas, as quais se separaram dos outros genótipos, sendo que a presenças das variedades crioulas foi o fator que mais contribuiu para esta separação. A adição de amostras crioulas em estudos de *profiling* se mostrou de extrema importância para analisar a variação nas expressões de transcritos e miRNAs relacionada à diferentes backgrounds genéticos. A adição e uma variedade com um background genético diferente e com alta variabilidade genética pode ser útil para futuros estudos de biossegurança, principalmente em estudos visando a avaliação de respostas à estresse biótico e abiótico.

Baseado em nosso conhecimento, este é o primeiro estudo sobre os perfis de miRNAs em plantas geneticamente modificadas e o primeiro estudo sobre a análise comparativa de eventos transgênicos estaqueados. Apesar de as diferenças de expressão de miRNAs e transcritos endógenos não apresentar um risco direto para a segurança destes eventos transgênicos, estas alterações devem ser cuidadosamente levadas em consideração na elaboração de novos *guidelines* de análise de riscos de OGMs.

8. CONSIDERAÇÕES FINAIS

A análise de risco é o primeiro passo para identificação de possíveis efeitos adversos e mudanças não intencionais de OGMs. A principal abordagem utilizada nestas análises é a equivalência substancial, onde características dos organismos modificados são comparadas com as linhagens parentais, envolvendo estudo de desempenho agronômico, fenotípico e de composição centesimal dos alimentos. Além disso, estudo sobre a caracterização molecular do inserto transgênico também são exigidos, como expressão das proteínas e mRNA transgênicos, localização no genoma e sequenciamento. Entretanto, estas análises possuem uma capacidade limitada na identificação de possíveis riscos e subestimam os possíveis efeitos em longo prazo que podem surgir devido à modificações genéticas. É neste cenário que as abordagens *profiling* se fazem necessárias, pois são capazes de gerar uma visão mais ampla e holística da análise risco, por meio da geração de perfis de expressão de proteínas, metabólitos, transcritos, miRNAs, entre outros.

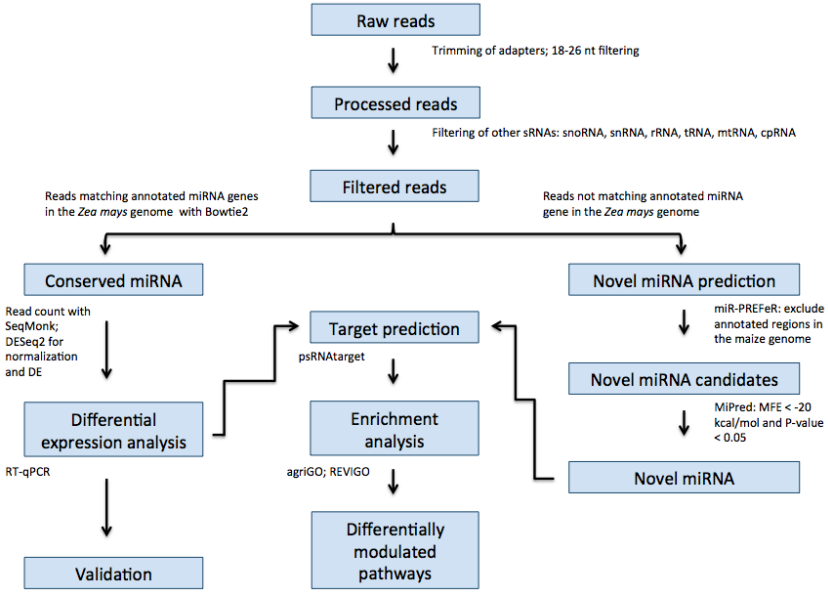
Apesar de não existir um consenso sobre a necessidade e os valores das abordagens *profiling* nos estudos de risco de OGMs, os resultados obtidos neste trabalho demonstraram que o uso destas abordagens se mostrou extremamente útil para a geração de conhecimento sobre a regulação da expressão de transcritos e miRNAs endógenos em variedades de milho geneticamente modificado, assim como em uma variedade crioula. Além disso, as abordagens testadas neste trabalho foram capazes de responder as hipóteses inicialmente levantadas de uma forma sensível e robusta.

Investigações futuras devem ser realizadas visando elucidar as implicações biológicas das alterações encontradas, principalmente àqueles ligados a impactos ambientais e na saúde humana e animal. Espera-se também que esse trabalho auxilie na elaboração de abordagens mais adequadas para a identificação de possíveis efeitos adversos na regulação da expressão de genes endógenos de eventos transgênicos estaqueados.

APÊNDICES

Capítulo II

Additional file 1. Pipeline used for the analysis of miRNA data.



Additional file 2. List of primers used for real-time RT-PCR quantification of differentially expressed conserved maize miRNA.

Primer name	Sequence	Conserved miRNA amplified
MIR162-3p*	TCGATAAACCTCTGCATCCA	zma-MIR162-3p
MIR167c-3p	GCTGTGGCAGCCTCACTA	zma-MIR167c-3p
MIR167ej-5p	AGCTGCCAGCATGATCTG	zma-MIR167e-5p zma-MIR167j-5p
MIR169f-5p	GCCAAGGATGACTTGCCTA	zma-MIR169f-5p
MIR169m-3p	GGCATCCATTCTTGGCTAA	zma-MIR169m-3p
MIR399ah-3p	CCAAAGGAGAATTGCCCT	zma-MIR399a-3p zma-MIR399h-3p
MIR399eij-3p	CCAAAGGAGAGTTGCCCT	zma-MIR399e-3p zma-MIR399i-3p zma-MIR399j-3p
MIR529-5p*	GCAGAAGAGAGAGAGTACAGCCT	zma-MIR529-5p
MIR827-3p	CTTAGATGACCATCAGCAAACA	zma-MIR827-3p
MIR827-5p*	CTTTGTTGGTGGTCATTTAACC	zma-MIR827-5p

Note: '*' = Those primers were excluded from the analysis due to multiple peaks in the melting curve.

Additional file 3. Log2Fold-Change and *p-adjusted* values of the differentially expressed miRNAs in the Landrace variety compared to high-breeding varieties.

	miRNA	**Comparison**	**log2 Fold-Change**	***p-adj* value**
Up-regulated	zma-MIR397a	Bt vs Land	1.6198	0.0002
	zma-MIR397b	Bt vs Land	1.3408	0.0016
	zma-MIR319c	Bt vs Land	1.0607	0.0119
	zma-MIR408b	Bt vs Land	0.8931	0.0172
	zma-MIR396f	Bt vs Land	0.7807	0.0036
	zma-MIR396e	Bt vs Land	0.7283	0.0075
	zma-MIR397a	Conv vs Land	1.4495	0.0011
	zma-MIR397b	Conv vs Land	1.0619	0.0139
	zma-MIR396f	Conv vs Land	0.8843	0.0000
	zma-MIR396e	Conv vs Land	0.8585	0.0000
	zma-MIR171k	Conv vs Land	0.8088	0.0038
	zma-MIR162	Conv vs Land	0.7859	0.0008
	zma-MIR167c	Conv vs Land	0.6672	0.0027
	zma-MIR171c	RR vs Land	1.5324	0.0024
	zma-MIR397a	RR vs Land	1.4303	0.0013
	zma-MIR397b	RR vs Land	1.1109	0.0141
	zma-MIR396f	RR vs Land	0.9236	0.0017
	zma-MIR396e	RR vs Land	0.9170	0.0013
	zma-MIR319c	RR vs Land	0.8740	0.0491
	zma-MIR397a	RRxBt vs Land	1.5723	0.0001
	zma-MIR397b	RRxBt vs Land	1.4375	0.0001
	zma-MIR171c	RRxBt vs Land	1.0693	0.0205
	zma-MIR408b	RRxBt vs Land	0.8819	0.0283
	zma-MIR396e	RRxBt vs Land	0.8454	0.0001
	zma-MIR396f	RRxBt vs Land	0.8123	0.0006

Continuation Additional file 3.

	miRNA	Comparison	log2 Fold-Change	*p-adj* value
Down-regulated	zma-MIR169h	Bt vs Land	-0.7329	0.0036
	zma-MIR169e	Bt vs Land	-0.7335	0.0036
	zma-MIR167f	Bt vs Land	-0.7426	0.0136
	zma-MIR169d	Bt vs Land	-0.7509	0.0010
	zma-MIR167j	Bt vs Land	-0.8421	0.0036
	zma-MIR169m	Bt vs Land	-1.0021	0.0154
	zma-MIR166m	Conv vs Land	-0.3125	0.0407
	zma-MIR168b	Conv vs Land	-0.5400	0.0299
	zma-MIR168a	Conv vs Land	-0.5652	0.0190
	zma-MIR156k	Conv vs Land	-0.6278	0.0000
	zma-MIR169j	Conv vs Land	-0.6634	0.0107
	zma-MIR169d	Conv vs Land	-0.8725	0.0000
	zma-MIR169i	Conv vs Land	-0.8843	0.0000
	zma-MIR169k	Conv vs Land	-0.9179	0.0002
	zma-MIR399c	Conv vs Land	-0.9309	0.0358
	zma-MIR169h	Conv vs Land	-1.1046	0.0000
	zma-MIR169e	Conv vs Land	-1.1051	0.0000
	zma-MIR399h	Conv vs Land	-1.1616	0.0064
	zma-MIR399a	Conv vs Land	-1.3275	0.0000
	zma-MIR169d	RR vs Land	-0.8030	0.0013
	zma-MIR169h	RR vs Land	-0.8162	0.0141
	zma-MIR169e	RR vs Land	-0.8168	0.0141
	zma-MIR169m	RR vs Land	-1.0837	0.0141
	zma-MIR169k	RRxBt vs Land	-0.6470	0.0437
	zma-MIR169h	RRxBt vs Land	-0.7256	0.0080
	zma-MIR169e	RRxBt vs Land	-0.7261	0.0080